Modeling the Household Purchasing Process Using a Panel Data Tobit Model

Diansheng Dong, Todd M. Schmit, and Harry M. Kaiser Cornell University

Chanjin Chung Oklahoma State University

Department of Applied Economics and Management College of Agriculture and Life Sciences Cornell University Ithaca, New York 14853-7801

Modeling the Household Purchasing Process Using a Panel Data Tobit Model

Diansheng Dong*
Chanjin Chung
Todd M. Schmit
Harry M. Kaiser

*Department of Applied Economics and Management
Cornell University, Ithaca, NY 14850
Phone: (607) 255-2985
Fax: (607) 254-4335
E-mail:dd66@cornell.edu

September 2002

Modeling the Household Purchasing Process Using a Panel Data Tobit Model

Abstract

A panel data Tobit model is developed to examine the household purchase process for a

frequently purchased commodity. The proposed model accounts not only for censoring or

sample selectivity, but also the temporal dependence of the purchasing process using household

panel data. The flexible error structure in the model accounts for both state dependence and

household heterogeneity. Empirical findings show that purchase habits of milk persist across

households over time, and most of them come from the household heterogeneity in preferences.

Results also show that advertising increases the purchase quantity and purchase frequency

simultaneously.

Key words: household purchase, fluid milk, advertising, panel data, Tobit model, probability
        simulation

Modeling the Household Purchasing Process Using a Panel Data Tobit Model

Introduction

This paper investigates the factors that influence household purchases on a frequently purchased commodity using household panel data. Most studies using panel data have ignored household heterogeneity in preferences over commodities and state dependence caused by the casual links between past and present purchase behavior. This simplification is mainly due to the considerable computational burden in model estimation in which evaluations of multidimensional integrals are involved. However, both heterogeneity and state dependence can be a serious source of misspecification and ignoring them will, in general, yield inconsistent parameter estimates.

In this study, we develop the panel data Tobit model originated by Hajivassiliou (1994), who used it to study the external debt crises of developing countries. Simulation probability is used to mitigate the computational burden of using maximum-likelihood estimation. Using household panel data, we address the following two important empirical issues more explicitly than has been addressed in the literature to date:[1]

(1) Does observed purchase habit-persistence on a frequently purchased commodity come from household heterogeneity or state dependence?

(2) Do purchase quantity and purchase frequency move in the same or opposite direction over time given an economic disturbance?

---

[1] Keane (1997) surveyed the studies and generalized a model to separate household heterogeneity and state dependence in a study of consumer's brand choices. This study only considered brand switching within a utility maximizing framework.

1

The answer to (1) is particularly important for market policy initiatives designed to enhance sales. For example, if existing household-purchase habits arise primarily from state dependence, then an increase in household purchase induced by a price promotion will be carried over to future purchases. In this case, price promotion would be an effective policy to increase long-run sales. On the other hand, if habit-persistence comes from heterogeneity in preferences, a marketing strategy oriented in altering household preferences, e.g., from advertising, would be recommended. The answer to (2) is also important in evaluating the effectiveness of demand enhancing strategies. Of particular importance is whether these strategies result in an increase in both purchase quantity and frequency of purchase, or an increase in purchase quantity and a decrease in frequency of purchase due to household stockpiling.

The research background addressing household heterogeneity and state dependence, as well as the estimation issues, is given in section II. Section III layouts the derivation of the econometric model surrounding the specification to control for censoring, heterogeneity and state dependence, followed by model prediction in section IV. An empirical application studying weekly household fluid milk purchases is given in section V. We close with some conclusions and directions for future research in section VI.

Background

Over the past two decades, the increased availability of electronic scanner panel data on household purchasing behavior has allowed researchers to investigate more fully the factors that influence household purchase decisions. Household scanner panel data

contains detailed demographic information on a selected household panel and their

purchase records over a certain period of time. Panel data offers us the possibility of

studying the household-level purchase process in a dynamic way. Examples can be found

in Keane (1997), Erdem and Keane, and Erdem, Keane and Sun.

The availability of multiple time-series observations per household allows one to

control for the presence of state dependence as well as the permanent unobserved

heterogeneity across households. State dependence is the temporal linkage of purchasing

arising from purchase carryover, learning behavior, and other factors. This is a common

phenomenon in aggregate time-series models. However, at the household level, it

complicates the study because of the non-negativity restriction on household purchases.[2]

Further, the temporal linkage of purchasing in panel data models, unlike in aggregate

models, arises not only from state dependence, but also from unobservable household

heterogeneity (Hajivasiliou, 1994). Heterogeneity across households persists over time. It

may be caused by different preferences, endowments, or attributes (Keane, 1997). This

type of temporal dependence in the household purchase process can be a source of serious

misspecification. Ignoring it tends to produce inconsistent parameter estimates.

Another important issue regarding the use of panel data is how to control for bias

caused by censored data or sample selection.[3] Failure to account for censoring or sample

selection will lead to inconsistent estimation of the behavioral parameters of interest since

these are compounded with parameters that determine the probability of entry into the

---

[2] This dynamic version of Tobit model has gained much attention recently (e. g., Zeger and Brookmeyer; Lee; and Wei).
[3] According to Amemiya, Type 1 and Type 2 Tobit models are the censored regression and sample selection models, respectively.

sample. The controlling of censoring or sample selectivity has been well addressed for cross-sectional data. However, censoring or sample selectivity is an equally acute problem in panel data. In the case of temporal independence, by pooling the data, censoring or sample selection in panel data can be accounted for as in the cross-sectional case. For example, if no links among present purchases and previous purchases are assumed, the non-negative purchase selection can be modeled by the traditional censored-Tobit model or its variations.

Recently, much attention has been focused on dealing with the censoring or sample selection problem in panel data analysis along with the assumption of temporal dependence. The difficulty in estimating this model comes from the evaluation of multidimensional probability integration. The recent discussion and development of probability simulation methods make maximum-likelihood estimation feasible for use in panel data censoring or sample selection models. Hajivassiliou (1994) used the simulated maximum-likelihood in a panel data structure to study the external debt crises of developing countries. He simulated the likelihood contributions as well as the scores of the likelihood and its derivatives. To keep the conventional maximum likelihood style, one can use some well-behaved simulators to replace the multidimensional probability integrals in the log-likelihood function, and then numerically evaluate the gradients, or even Hessians, for the continuous simulated log-likelihood function.

In the next section, following Hajivassiliou (1994), we derive a panel data Tobit model, which accounts for both the temporal linkage and censoring across households. A

4

conventional likelihood function is eventually built that can be well approximated by simulated probability.

Econometric Model

Consider a panel of N households whose weekly purchases on the studied commodity are observed over T weeks. In this case, a data array for the ith household, y_i and x_i, is observed where y_i is a $T \times 1$ vector of observed weekly commodity purchases and x_i is a $T \times K$ matrix of exogenous market-related and household-specific variables. A censored-type Tobit model to account for censoring is assumed in this study as,

$$y_{it} = \begin{cases} x_{it}\,\beta + u_{it}, & \text{if } u_{it} > -x_{it}\,\beta \\ 0, & \text{otherwise} \end{cases} \quad i = 1, \ldots, N; \ t = 1, \ldots, T \tag{1}$$

where β is a $K \times 1$ vector of estimated parameters and u_{it} is an error term. The subscript t refers to time (week) so that y_{it} is the tth element in y_i. We assume u_{it} is jointly distributed normal over t with a mean vector of zero and household-specific variance-covariance matrix Ω_i.

The likelihood function for the ith household can be represented as

$$L_i = \int_{U(y_i)} \phi(u_i; \Omega_i)\, du_i, \ i = 1, \ldots, N \tag{2}$$

where ϕ is the probability density function (pdf) of multivariate normal and $U(y_i)$ is the probability integration range of u_i given observed y_i. Like y_i, u_i is a vector of $T \times 1$.

To facilitate the presentation, we can partition the T-week observations for the ith household into two mutually exclusive sets, one containing data associated with the T_{i0} non-purchase weeks and another containing data associated with the T_{i1} purchase weeks

5

where $T=T_{i0}+T_{i1}$. Accordingly, the ith household's error term variance-covariance matrix can be partitioned as:

$$\Omega_i = \begin{bmatrix} \Omega_{i00} & \Omega'_{i01} \\ \Omega_{i01} & \Omega_{i11} \end{bmatrix}, \tag{3}$$

where Ω_{i00} is a T_{i0} x T_{i0} submatrix associated with the non-purchase weeks, Ω_{i11} is a T_{i1} x T_{i1} submatrix associated with purchase weeks, and Ω_{i01} is a T_{i0} x T_{i1} submatrix of covariance across purchase and non-purchase weeks.

With this partitioning, the likelihood function for the ith household under a particular purchase pattern over T weeks can be simplified as

$$L_i\left(y_{i0}, y_{i1} \mid y_{i0}=0, y_{i1}>0\right) = \phi_1\left(u_{i1}\right) \int_{-\infty}^{-x_i\beta} \phi_{0/1}\left(u_{i0}\right) d u_{i0}, \ i=1,\ldots,N \tag{4}$$

where u_{i0} is the error term vector in (1) associated with the non-purchase weeks and u_{i1} is the error term vector associated with the purchase weeks. The multinormal pdf of u_{i1}, ϕ_1, has a mean vector zero and variance-covariance matrix Ω_{i11}, while the T_{i0}-fold integral, $\int_{-\infty}^{-x_{i0}\beta}$, is evaluated at the upper bound $-x_{i0}\beta$, where $-x_{i0}$ is a T_{i0} x K vector data associated with non-purchase weeks. The conditional pdf of u_{i0} given u_{i1}, $\phi_{0/1}$, is distributed multinormal with a mean vector $u^i_{0/1}$ and variance-covariance matrix $\Omega^i_{0/1}$,

where $u^i_{0/1} = \Omega_{i01}\Omega_{i11}^{-1}u_{i1}$, $\Omega^i_{0/1} = \Omega_{i00} - \Omega_{i01}\Omega_{i11}^{-1}\Omega'_{i01}$.

The likelihood function for N households can then be written as the product of (4) over all households, i.e.:

$$L = \prod_{i=1}^{N} \left[\phi_1 \left(u_{i1} \right) \int_{-\infty}^{-x_i \beta} \phi_{0/1} \left(u_{i0} \right) d u_{i0} \right] . \tag{5}$$

To obtain the maximum-likelihood estimates of (5), we need to evaluate the T_{i0}-fold integral. With an unrestricted Ω_i, the traditional numerical evaluation is computationally intractable when T_{i0} exceeds 3 or 4. One conventional approach is to restrict Ω_i to be household and time invariant, thus:

$$\Omega_i = E \left(u_i u_i^{'} \right) = \sigma^2 I_T , \text{ for all } i, \tag{6}$$

where σ^2 is an estimated variance parameter and I_T is a T-dimensional identity matrix. This structure yields the pooled cross-sectional Tobit model that ignores all temporal and spatial linkages and can be estimated by traditional maximum-likelihood procedures.

However, to account for household-specific heterogeneity and state dependence, one can assume u_{it} consists of two error-components:

$$u_{it} = \alpha_i + \varepsilon_{it} , \tag{7}$$

where α_i, uncorrelated with ε_{it}, is a household-specific normal random variable used to capture household heterogeneity. If state dependence can be ignored, one can assume ε_{it} as an *i.i.d.* normal random variable. In this model, the multidimensional integral can be written as a univariate integral of a product of cumulative normal distributions, which dramatically reduces the computational burden (Hajivassiliou, 1987). In general, state dependence is not negligible; however, it can be imposed by an autoregressive structure of ε_{it}.

In this study, we assume that ε_{it} follows a first-order autoregressive process; however, it is extendable to higher order autoregression. Specifically, for this one-factor plus AR (1) error structure, we assume:

$$\varepsilon_{it} = \rho \; \varepsilon_{it-1} + v_{it}; \quad |\rho| < 1 , \tag{8}$$

where ρ is the autocorrelation coefficient and $v_{it} \sim N\left(0, \sigma_0^2\right)$ for all i and t.

Additionally, $\alpha_i \sim N\left(0, \sigma_2^2\right)$ for all i, which persists over time. To warrant stationarity, we assume $\varepsilon_{it} \sim N\left(0, \sigma_1^2\right)$ and $\sigma_0^2 = \sigma_1^2(1 - \rho^2)$. Accordingly, the above error structure implies that Ω_i has the following form:

$$\Omega_i = \sigma_2^2 J_T + \sigma_1^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{T-3} & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{T-4} & \rho^{T-3} \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \cdots & \rho & 1 \end{bmatrix} \tag{9}$$

where J_T is a $T \times T$ matrix of ones. The Ω_i in (9) is invariant across households. To correct for possible heteroskedasticity, one may also specify σ_1^2 or σ_2^2 or both as a function of some continuous household specific variables such as income and household size (Maddala).

With Ω_i as given in (9), the likelihood function in (5) requires the evaluation of a T_{i0}-fold integral. Recall that T_{i0} varies across households. When T_{i0} exceeds 3 or 4, as aforementioned, the evaluation of these multi-dimensional integrals becomes unacceptable

in terms of low speed and accuracy. As an alternative we use a simulated probability method in evaluating these integrals.

Recently, several probability simulators have been introduced and investigated in literature (Hijivassiliou and McFadden; Geweke; Breslaw; Borsch-Supan and Hajivassiliou; Keane, 1994; Hajivassiliou, McFadden and Rudd; Geweke, Keane and Runkle). The smooth recursive conditioning simulator (GHK) proposed by Geweke; Hajivassiliou and McFadden; and Keane(1994) has been chosen for this study because this algorithm was the most reliable simulator among those examined by Hajivassiliou, McFadden and Rudd.

Model Predictions

This panel data Tobit model has the ability to predict both the purchases in a static as well as dynamic environment. Given time period t, the static expected purchases and purchase probabilities of model (1) can be derived as follows:

$$E(y_{it}) = \Phi(\theta_{it}) \cdot x_{it}\beta + \sqrt{\sigma_1^2 + \sigma_2^2} \cdot \phi(\theta_{it}) , \qquad (10)$$

$$Prob\ (y_{it} > 0) = \Phi(\theta_{it}) , \qquad (11)$$

$$E(y_{it} \mid y_{it} > 0) = x_{it}\beta + \sqrt{\sigma_1^2 + \sigma_2^2} \cdot \frac{\phi(\theta_{it})}{\Phi(\theta_{it})} \qquad (12)$$

where $\phi(\theta_{it})$ is the standard normal pdf evaluated at $\theta_{it} = \dfrac{x_{it}\beta}{\sqrt{\sigma_1^2 + \sigma_2^2}}$, and $\Phi(\theta_{it})$ is the

standard normal cdf with the support of (-∞, θ_{it}). Equation (10) is the unconditional

expected purchase of household *i* at time *t*, (11) is the expected probability of purchase, and (12) is the conditional expected purchase given a purchase occasion. It is clear that (10) is the product of (11) and (12). Therefore, the elasticity of the unconditional purchase can be decomposed into two components: the elasticity of conditional purchase and the elasticity of the positive purchase probability (McDonald and Moffitt).

In order to take advantage of the dynamic nature of the model, we also derive the following sets of expected values:

$$E(y_{it} \mid y_{it-1} = 0) = Prob(y_{it} > 0 \mid y_{it-1} = 0) \cdot E(y_{it} \mid y_{it} > 0, y_{it-1} = 0), \tag{13}$$

where,

$$Prob(y_{it} > 0 \mid y_{it-1} = 0) = 1 - \frac{\Phi_2(-\theta_{it}, -\theta_{it-1}, \delta)}{\Phi(-\theta_{it-1})}, \tag{14}$$

and,

$$E(y_{it} \mid y_{it} > 0, y_{it-1} = 0) = x_{it}\beta +$$

$$\sqrt{\sigma_1^2 + \sigma_2^2} \; \frac{\phi(\theta_{it})\Phi(\frac{-\theta_{it-1} - \delta\theta_{it}}{\sqrt{1-\delta^2}}) + \delta\phi(\theta_{it-1})\Phi(\frac{\theta_{it} + \delta\theta_{it-1}}{\sqrt{1-\delta^2}})}{\Phi(-\theta_{it-1}) - \Phi_2(-\theta_{it}, -\theta_{it-1}, \delta)}, \tag{15}$$

where $\Phi_2(-\theta_{it}, -\theta_{it-1}, \delta)$ represents the standard joint (bivariate) normal cdf of y_{it} and y_{it-1}

with correlation coefficient $\delta = \dfrac{\sigma_1^2\rho + \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.

Equations (13)-(15) provide information on the tth period purchase given a non-purchase occasion at the t-1th period. Similarly, given a purchase occasion at t-1, we have,

$$E(y_{it} \mid y_{it-1} > 0) = Prob(y_{it} > 0 \mid y_{it-1} > 0) \cdot E(y_{it} \mid y_{it} > 0, y_{it-1} > 0), \tag{16}$$

where,

$$Prob\,(y_{it} > 0 \mid y_{t-1} > 0) = \frac{\Phi_2(\theta_{it}, \theta_{it-1}, \delta)}{\Phi(\theta_{it-1})}\,, \tag{17}$$

and,

$$E(y_{it} \mid y_{it} > 0, y_{it-1} > 0) = x_{it}\beta +$$

$$\sqrt{\sigma_1^2 + \sigma_2^2}\ \frac{\phi(\theta_{it})\Phi(\dfrac{\theta_{it-1} - \delta\theta_{it}}{\sqrt{1-\delta^2}}) + \delta\phi(\theta_{it-1})\Phi(\dfrac{\theta_{it} - \delta\theta_{it-1}}{\sqrt{1-\delta^2}})}{\Phi_2(\theta_{it}, \theta_{it-1}, \delta)}\,. \tag{18}$$

Both sets of equations (13)-(15) and (16)-(18) are determined by the correlation between current purchase (y_{it}) and last purchase (y_{it-1}). If there is no correlation between y_{it} and y_{it-1}, that is, σ_2^2 and ρ defined in (9) are both zeros, the two sets of equations will be the same and equal to the set of equations (10)-(12) respectively. Elasticities can then be calculated based on these expected values.

Empirical Model

In this empirical application, we follow a panel of U.S. households over a four-year period from 1996 through 1999. For each given time unit (one week), we observe whether the household buys fluid milk, and if it does, the amount. We are particularly interested in the estimation of σ_2^2 (which captures the household heterogeneity in preferences) and ρ (which captures the state dependence), as well as the impacts of price, income, advertising, and other demographic variables on household purchase decisions for fluid milk over time.

Data

Household data are drawn from ACNielsen Homescan Panel,[4] including household purchase information for fluid milk products and annual demographic information. The purchase data is purchase-occasion data collected by the households, who used hand-held scanners to record purchase information. This data includes date of purchase, UPC code, total expenditure, and quantities purchased. The final purchase data were reformulated to a weekly basis and combined with the household demographic information. The household data was merged with national weekly generic milk advertising expenditures obtained from Bozell, Inc. The data are over a 208-week period from January 1996 through December 1999, and include more than 30,000 households. The generic advertising expenditures are national-level expenditures and vary over time but not across households. The number of households in the panel varies from year to year, but only those households participating in all years (23,008) are included in the sample.[5] Given the large size of the panel, we select a 10% random sample of households for estimation purposes.

This application is concerned with weekly purchases of fluid milk for home consumption only. The household weekly purchase quantities and expenditures are defined as the sum of quantities and expenditures on all types of fluid milk such as whole, reduced fat, and skim milk purchased within that week. To control for computation within a reasonable time, we selected the 52 weeks of 1999 to estimate the panel data Tobit model specified in equations (1)-(6), including data from the last 39 weeks in 1998 to derive the nine-month advertising lags (as described below). The dependent variable in our model is

---

[4] Copyright 2000 by ACNielsen
[5] We ignore the possible selection bias caused by the sample attrition.

household milk purchase quantity. On average, 27 of the 52 weeks are purchase occasions

with a mean purchase of 0.66 gallons over all weeks and 1.25 gallons for purchase weeks.

Advertising and Price

Advertising is considered to be a demand shifter in the marketing literature. In this

analysis, it is based on total weekly national generic fluid milk advertising expenditures

(funded by dairy farmers and fluid milk processors) aggregated over all media types. To

capture the carry-over effect of advertising, advertising expenditures are lagged 39 weeks

(9 months) and a polynomial distributed lag model is adopted as follows (Clarke):

$$A^* = \sum_{i=0}^{L} \omega_i A_{t-i} , \qquad\qquad (19)$$

where A_{t-i} is the ith lag of advertising, L is the total lag length, and $\omega_i = 1 + (1 - \frac{1}{L})i - \frac{1}{L}i^2$ (i

= 0, 1, …, L) are the quadratic weights of the lag advertising. Three point restrictions are

imposed on ω_i: (i) the weight of current advertising is 1, i.e., $\omega_0 = 1$; (ii) the weight of the

39th lag is 0 ($\omega_{39} = 0$), that is, the effect of advertising ends at the 39th week; and (iii) $\omega_{-1} =$

0, that is, future advertising has no effect on today's market. A^*, the sum of weighted

advertising over the current and all the lags, is used as an explanatory variable in equation

(1). The coefficient of A^* then represents the long-run effect of advertising.

　　Prices are not observed directly in the panel data. An estimate of price can be

obtained by dividing reported expenditures by quantity for the purchase weeks. However,

Theil, Deaton (1987, 1988), and Cox and Wohlgenant have showed that this method of

calculating a composite commodity price reflects not only differences in market prices

faced by each household but also endogenously determined commodity quality.

Furthermore, no price information is available for those non-purchase weeks. A number of

alternative approaches can be used to obtain estimates of the missing prices. In this

analysis, we use a zero-order correction for the missing prices. For each household the

imputed prices for non-purchase weeks are set equal to the mean prices of the purchase

weeks for that household.

A number of annual household characteristics are also incorporated as explanatory

variables. Table 1 provides an overview of these household characteristics as well as the

advertising and price variables used in the analysis.

Empirical Findings

Parameter estimates were obtained by maximizing the likelihood function in (5) using the

GAUSS software system. Numerical gradients of (5) were used in the optimization

algorithm proposed by Berndt, Hall, Hall, and Hausman. The standard errors of the

estimated coefficients were obtained from the inverse of the negative numerically

evaluated Hessian matrix. We use 500 replicates to simulate the multinormal probability

in the likelihood function using the GHK procedure. The estimated coefficients are

presented in Table 2.

We can address the first question posed in section I by examining the estimates of

σ_2^2 and ρ. As expected, σ_2^2 and ρ are both significantly different from zero at the 0.01

significance level, implying that habits persist in household fluid milk purchasing. The

correlation coefficient between current purchase (y_{it}) and last purchase (y_{it-1}) is

$\delta = \dfrac{\sigma_1^2 \rho + \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = 0.2934$, implying that lagged purchase are positively related to current

purchase. The component of this correlation associated with serial state dependence is

$\dfrac{\sigma_1^2 \rho}{\sigma_1^2 + \sigma_2^2} = -0.0665$, implying that if household A purchased more than household B at

time t-1, then household A will purchase less than household B at time t, ceteris paribus,

but the difference is small.[6] This is likely caused by the fact that fluid milk is perishable.

However, the component of this correlation associated with the household heterogeneity is

0.3599 ($\dfrac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$), implying that if household A purchased more than household B at time

t-1, then household A will still purchase more than household B at time t, ceteris paribus.

This results from the difference in household preferences for milk: household A prefers to

drink more fluid milk than household B does. Since most of the temporal correlation is

due to household heterogeneity, the negative effect of state dependence on purchase is

overwhelmed by the positive effect of household heterogeneity, resulting in a positive

overall correlation coefficient. This has important implications for the effectiveness of

short-term price promotions versus long-term advertising programs aimed at increasing

total milk consumption over time. These results would indicate that long-term preference

changes would be more effective.

In addition, from Table 2, we also see that all of the other coefficients are

significant at the 0.05 significance level as well except for the proportion of teenage girls

---

[6] In contrast, Allenby and Lenk found a positive correlation coefficient associated with state dependence in the study of household's choice among different brands of ketchup using Logistic normal regression model.

and elderly people in the household and the education of household head.[7] Consumption

of milk by young children (under age 13) was higher than that of adults. Male teenage

children were also found to consume proportionally more milk relative to mature adults

(age18-64), albeit less than younger children, while no significant difference was found for

teenage girls relative to mature adults. As expected, the empirical results also showed that

African American and Hispanic households consumed less milk relative to whites. Not

surprisingly, we found that if mothers worked outside the household, less milk was

purchased. The effect of working-mother households was expected to be negative, a priori.

While working mothers would generate additional income (captured in the household

income effect), these households would likely have less time to prepare meals at home and

the children's diet could not be monitored closely to include milk with their meals at home;

both negatively influence household milk consumption. Furthermore, income, family size,

and family head age were all positively associated with household milk purchases. As

expected, price was found to have a significant and negative effect on household

purchases, while advertising was found to be significant and positive.

To better understand the economic effects and to interpret the dynamic results of

the model, we calculate elasticities of some key variables based on the expected values

derived earlier. The elasticities of the 30th week in 1999 evaluated at the household sample

mean with respect to equations (10)-(12), (13)-(15) and (16)-(18) are presented in Tables 3,

4, and 5, respectively.[8]

---

[7] The household head is defined as the female household head, if present, else the male household head.
[8] The elasticities are quite similar when evaluated at the first, tenth, twentieth, fortieth, and fiftieth week as well.

The long-run elasticity of generic milk advertising is 0.147(Table 3). That is, a 1% increase in generic advertising would increase household milk purchases by 0.147%, on average. The 0.147% increase in household purchase counts as 0.089% (60.8%) from the increase of household milk purchase probability and 0.058% (39.2%) from the increase of household conditional milk purchase. An increase in purchase probability implies an increase in purchase incidence or number of purchasers. Thus, of the total impact of advertising on household milk demand, the majority of the effect comes from purchase incidence rather than conditional purchase levels.

As expected, the price elasticity is negative and inelastic at -0.693. The income effects are relatively low, while household size has a much more prominent effect. As was evident with the elasticity estimates, male teenage children have a smaller effect than younger children. Compared to all the households, positive purchase households appeared less sensitive to price changes, given that the total price effect is composed of the purchase probability effect. Interestingly, the effects of all the variables in increasing unconditional purchase quantities through the increase in the conditional purchase quantities are weighted less than through the increase in the probability of purchase.

Since the expected values in (10)-(12) are associated with one period (t) only, Table 3 is indeed the results of the conventional (static) Tobit model. Tables 4 and 5 examine the influences on household purchases in a dynamic way by reporting the calculated purchase relationship between two consecutive purchase periods. By analyzing Tables 4 and 5, we can also address the second question posed in section I.

The results in Table 4 and 5 are quite consistent with those in Table 3. However, additional information can be acquired by comparing the two tables. Even though the effects on the total (unconditional) purchases at t (the columns on the left) given a non-purchase occasion at t-1 are larger than those given a purchase occasion, the conditional purchases at time t (the columns in the middle) are virtually identical under the two situations for all the variables. This implies that if the household purchases at time t, the effects of advertising, price, and other demand factors on purchase quantity are about the same no matter whether it purchased or not at the previous time period. This also verifies the fact that fluid milk is perishable. However, the effects on purchase probability at time t (the columns on the right) are greater if there is no purchase at time t-1 than those if there is a purchase. For example, a 1% decrease in price would increase the current purchase probability by 0.535%, given a non-purchase occasion, and by 0.356%, given a purchase occasion, during the last period. In both cases, the purchase frequency would tend to increase. Intuitively, if household A did not purchase at time t-1 while household B did, then a decrease in the milk price will increase the purchase at time t by about the same amount for households A and B given that they will purchase at time t, ceteris paribus. However, the purchase frequency of household A is larger than that of household B, implying that household A would purchase more frequently than household B, ceteris paribus. Similar information can also be drawn for advertising and other variables. Therefore, the purchase quantity and purchase frequency move in the same direction given an economic disturbance as is evident from the same sign of the associated elasticities in the middle columns and the left columns, respectively, for both Tables 4 and 5.

Summary and Conclusions

In this study, we developed a panel data Tobit model to study the household purchasing process while accounting for both household heterogeneity in preferences and state dependence. The effects of advertising, price, as well as other household characteristic variables on the household's decision of whether to purchase milk and how much to purchase were investigated. The model is a dynamic extension of the conventional Tobit model of censored consumption. The proposed model is able to account not only for the censored nature of commodity purchases, but also for the dynamics of the purchase process. In this censored model, a flexible error structure is assumed to account for both state dependence and household-specific heterogeneity. Even though this study focused on the household purchases of fluid milk, the model can easily be applied to other commodities when a censored panel structure data is confronted.

In the empirical application, we found that the purchase habit of fluid milk persists across households over time, and most of the habit-persistence comes from household heterogeneity in preferences. This implies that advertising aimed at increasing long-run consumption of milk (and hence to change household's preferences) will be more effective than short-term price promotions. We also found that a disturbance in consumer demand would drive household purchases and purchase frequency change in the same direction, indicating that advertising could increase household purchase quantity and purchase frequency simultaneously. The results also demonstrate that advertising has a larger effect on increasing purchase frequency than on increasing conditional purchase quantity, and

19

advertising increases the purchase probability more given a non-purchase in the prior time period, than if a purchase occasion occurred. Given these findings, it appears that milk-advertising efforts should concentrate on attracting new purchasers or increasing purchase frequency rather than on increasing the purchase amount of current consumers.

Prices were found to be inelastic. The prices used in this study were derived from the observed expenditures and quantities and reflect differences in market prices faced by each household as well as endogenously determined commodity quality. Further research is needed to separate the exogenous and the endogenous parts of this kind of derived price from each other. If these are not separated, care must be taken when using conventional price theory to interpret the empirical results. For instance, an increase in income would allow the household to buy a higher price milk product without change in the amount of purchase. A conclusion that price has no effect on purchases seems inescapable in this example. Indeed this increase in price (derived from the quantity and expenditure) is caused by the household's endogenous choice of a higher quality product, not from the increase in the exogenous market price.

Table 1. Descriptive Statistics of Explanatory Variables

Variables	Unit	Mean	Std. Dev.
Household Variables			
Household income	$ 000	48.0	33.2
Household size	Number	2.36	1.25
Percentage of household under age 12	Number	8.24	17.4
Percentage of household female and age 13-17	Number	2.03	7.46
Percentage of household male age 13-17	Number	2.08	8.20
Percentage of household over age 65 and above	Number	21.6	38.1
Head of household has high school degree or above	1/0	0.29	0.45
Age of household head	Number	51.8	13.6
Black household	1/0	0.06	0.23
Hispanic household	1/0	0.04	0.20
Mother of household works	1/0	0.43	0.50
Purchase Related Variables			
Price	$/Gallon	2.98	1.06
Advertising	$000,000	2.40	1.80
Sum of weighted lag advertising ($A*$)	$000,000	701.9	49.5

Table 2. Estimated Dynamic Tobit Parameters

Variable	Estimate	Std. Error
Intercept	0.7500*	0.1959
Household Characteristics		
Log household income	0.0651*	0.0295
Inverse household size	-0.9719*	0.0897
Percentage of household under age 12	1.0528*	0.1275
Percentage of household female and age 13-17	0.0756	0.2416
Percentage of household male age 13-17	0.8775*	0.2201
Percentage of household over age 65 and above	0.0655	0.0788
Head of household has high school degree or above	0.0206	0.0429
Age of household head	0.0070*	0.0024
Black household	-0.6461*	0.0957
Hispanic household	-0.2639*	0.0983
Mother of household works	-0.1179*	0.0410
Purchase Characteristics		
Log price	-0.8319*	0.0058
Sum of weighted lag advertising ($A*$)	0.00024*	0.00005
Regression Coefficients		
Standard error 1 (σ_1)	1.1344*	0.0008
Standard error 2 (σ_2)	0.8507*	0.0068
Auto correlation coefficient (ρ)	-0.1039*	0.0020

"*" indicates significance at the 0.05 level or higher.

Table 3 Elasticities with respect to Equations (10)-(12)[*]

		Type of Expected Value		
		$E(y_t)$	$E(y_t \mid y_t > 0)$	$Prob(y_t > 0)$
Estimated Value[**] **(Actual Value)**		0.6608 (0.6637)	1.1997 (1.2477)	0.5508 (0.5319)
Elasticity	Advertising	0.1471	0.0577	0.0894
	Price	-0.6934	-0.2719	-0.4215
	Income	0.0543	0.0213	0.0330
	Household size	0.4483	0.1768	0.2725
	Age of household head	0.3026	0.1187	0.1839
	% of household under age 12	0.0723	0.0284	0.0439
	% of hh. male and age 13-17	0.0152	0.0060	0.0093

* The t-test based on the standard errors derived from the Delta Method (Rao) showed that all the elasticities in Tables 3 are significant at the 0.05 level or higher.
**The estimated values are obtained from equations (10)-(12). The actual values are obtained from the actual data. For example, the actual value of $E(y_t \mid y_t > 0)$ is the mean purchase at time t over those purchase households.

Table 4 Elasticities with respect to Equations (13)-(15)[*]

		Type of Expected Value		
		$E(y_t \mid y_{t-1} > 0)$	$E(y_t > 0 \mid y_{t-1} > 0)$	$Prob(y_t > 0 \mid y_{t-1} > 0)$
Estimated Value[**] **(Actual Value)**		0.8242 (0.8458)	1.2959 (1.3276)	0.6360 (0.6371)
Elasticity	Advertising	0.1362	0.0606	0.0756
	Price	-0.6420	-0.2858	-0.3562
	Income	0.0502	0.0224	0.0279
	Household size	0.4150	0.1847	0.2303
	Age of household head	0.2801	0.1247	0.1554
	% of household under age 12	0.0669	0.0298	0.0371
	% of hh. male and age 13-17	0.0141	0.0063	0.0078

*The t-test based on the standard errors derived from the Delta Method (Rao) showed that all the elasticities in Tables 4 are significant at the 0.05 level or higher.
**The estimated values are obtained from equations (13)-(15). The actual values are obtained from the actual data. For example, the actual value of $E(y_t \mid y_{t-1} > 0)$ is the mean purchase at time t over those households that purchased at time t-1.

Table 5 Elasticities with respect to Equations (16)-(18)[*]

		Type of Expected Value		
		$E(y_t \mid y_{t-1} = 0)$	$E(y_t > 0 \mid y_{t-1} = 0)$	$Prob(y_t > 0 \mid y_{t-1} = 0)$
Estimated Value[**] **(Actual Value)**		0.5914 (0.4760)	1.3227 (1.1239)	0.4472 (0.4235)
Elasticity	Advertising	0.1744	0.0611	0.1134
	Price	-0.8224	-0.2879	-0.5345
	Income	0.0644	0.0225	0.0418
	Household size	0.5317	0.1861	0.3455
	Age of household head	0.3589	0.1256	0.2332
	% of household under age 12	0.0858	0.0300	0.0557
	% of hh. male and age 13-17	0.0181	0.0063	0.0118

*The t-test based on the standard errors derived from the Delta Method (Rao) showed that all the elasticities in Tables 5 are significant at the 0.05 level or higher.

**The estimated values are obtained from equations (16)-(18). The actual values are obtained from the actual data. For example, the actual value of $E(y_t \mid y_{t-1} = 0)$ is the mean purchase at time t over those households that didn't purchase at time t-1.

REFERENCES

Allenby, G.M. and P.J. Lenk. "Modeling Household Purchase Behavior with Logistic Normal Regression." *Journal of the American Statistical Association* 89(1994):1218-31.

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

Berndt, E., B. Hall, R. Hall and J. Hausman. "Estimation and Inference in Nonlinear Structural Models." *Annals of Economic and Social Measurement* 3(1974):653-65.

Borsch-Supan, A. and V. Hajivassiliou. "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models." *Journal of Econometrics* 58(1993):347-68.

Breslaw. "Evaluation of Multivarite Normal Probability Integrals Using a Low Variance Simulator." *Review of Economics and Statistics* 4(1994):673-82.

Clarke, D. "Econometric Measurement of The Duration of Advertising Effect on Sales." *Journal of Marketing Research* 4(1976):345-57.

Cox, T.L., and M.K.Wohlgenant. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *American Journal of Agricultural Economics* 68(1986):908-19.

Deaton, A. "Estimation of Own- and Cross-Price Elasticities from Household Survey Data." *Journal of Economics* 36(1987):7-30.

--------. "Quality, Quantity, and Spatial Variation of Price." *American Economic Review* 78(1988):418-30.

Erdem, T. and M. Keane. "Decision Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science* 15(1996):1-20.

Erdem, T., M. Keane and B. Sun. "Missing Price and Coupon Availability Data in Scanner Panels: Correcting for the Self-Selection Bias in Choice Model Parameters." *Journal of Econometrics* 89(1999): 177-96.

Geweke, J. "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints." *Computer Science and Statistics*: Proceedings of the Twenty-Third Symposium on the Interface, American Statistical Association, Alexandria (1991):571-578.

Geweke, J.F., M.P., Keane and D.E. Runkle. "Statistical Inference in the Multinomial Multiperiod Probit Model." *Journal of Econometrics* 80(1997):125-65.

Hajivassiliou, V.A. "A Simulation Estimation Analysis of The External Debt Repayments Problems of LDC's: An Econometric Model Based on Panel Data." *Journal of Econometrics* 36(1987):205-30.

--------. "A Simulation Estimation Analysis of the External Debt Crises of Developing Countries." *Journal of Applied Econometrics* 9(1994):109-31.

Hajivassiliou, V. and D. McFadden. The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crisis. Cowles Foundation Discussion Paper No. 967, Yale University, 1990.

Hajivassiliou, V. and P.A. Ruud. Classical Estimation Methods for LDV Models Using Simulation. In *Handbook of Econometrics*, Vol. 4, eds., C.Engle and D. McFadden, pp. 2383-41. Amsterdam: North-Holland, 1994.

Hajivassiliou, V., D. McFadden and P. Ruud. "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results." *Journal of Econometrics* 72(1996):85-134.

Keane, M.P. "A Computationally Efficient Practical Simulation Estimator for Panel Data." *Econometrica* 62(1994):95-116.

Keane, M.P. "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior." *Journal of Business and Economic Statistics* 15(1997):310-27.

Lee, L. "Estimation of Dynamic and ARCH Tobit Models." *Journal of Econometrics* 92(1999):355-90.

Maddala, G.S. Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press, New York, 1983.

McDonald, J.F. and R.A. Moffitt. "The Use of Tobit Analysis." *Review of Economics and Statistics* 62(1980):318-21.

Rao, C.R. Linear Statistical Inference and Its Applications, New York: Wiley, 1973.

Theil, H. "Qualities, Prices, and Budget Enquiries." *Review of Economic Studies* 19(1952):129-47.

Wei, S.X. "A Bayesian Approach to Dynamic Tobit Models." *Econometric Reviews* 18(1999):417-39.

Zeger, S.L. and R. Brookmeyer. "Regression Analysis with Censored Autocorrelated Data." *Journal of the American Statistical Association* 81(1986):722-29.

OTHER A.E.M. RESEARCH BULLETINS

RB No	Title	Fee (if applicable)	Author(s)
2003-06	Estimation of a Censored AIDS Model: A Simulated Amemiya-Tobin Approach		Dong, D. and Kaiser, H.
2003-05	Price and Quality Effects of Generic Advertising: Salmon in Norway		Myrland, O., Dong., D. and Kaiser, H.
2003-04	Coupon Redemption and Its Effect on Household Cheese Purchases		Dong, D. and Kaiser, H.
2003-03	Dairy Farm Management Business Summary, New York State, 2002	($15.00)	Knoblauch, W., Putnam, L. and Karszes, J.
2003-02	Fruit Consumption, Dietary Guidelines, and Agricultural Production in New York State--Implications For Local Food Economies		Peters, C., Bills, N., Wilkins, J., and Smith, R.D.
2003-01	Future Structure of the Dairy Industry: Historical Trends, Projections and Issues		LaDue, E., Gloy, B. and Cuykendall, C.
2002-12	Prospects for the Market for Locally Grown Organic Food in the Northeast US	(12.00)	Conner, D.
2002-11	Dairy Farm Management Business Summary: New York State, 2001	($15.00)	Knoblauch, W. A., L. D. Putnam, and J. Karszes
2002-10	Needs of Agriculture Educators for Training, Resources, and Professional Development in Business Management and Marketing		C. A. Schlough and D. H. Streeter
2002-09	Financial Management Practices of New York Dairy Farms		Gloy, B. A., E. L. LaDue, and K. Youngblood
2002-08	Rural, Suburban and Urban Single Mothers' AFDC and FSP Participation and Labor Supply: Lessons for Welfare Reform		Ranney, C. K.
2002-07	Vegetable Consumption, Dietary Guidelines and Agricultural Production in New York State—Implications for Local Food Economies		Peters, C., N. Bills, J. Wilkins and R. D. Smith