Panel Data Double-Hurdle Model: An Application to Dairy Advertising

Diansheng Dong*
Chanjin Chung

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

AAEA Annual Meeting
August 5-8, 2001-Chicago

All statements in this article are the conclusions of the authors; ACNielsen does not support or confirm any of the conclusions that the authors base on ACNielsen information.

Panel Data Double-Hurdle Model: An Application to Dairy Advertising

Zero purchase outcomes in household survey data are usually interpreted as the result of short-run consumption behavior (infrequency of purchase), consumers' sensitivity to commodity prices (corner solutions), or social, psychological or ethical distinction (double-hurdles). For household panel data, appropriate temporal aggregation may eliminate the infrequency-of-purchase problem. However, the problems of possible corner solutions and double-hurdles will persist.

In this study, we extend the double-hurdle model used in cross-sectional data to a panel data structure. Following Cragg and Pudney, this extended model envisions that households must overcome two hurdles before realizing a positive purchase: (1) entering the market (becoming a potential purchaser), and (2) making the purchase. While accounting for sample selection bias through the double-hurdle structure in a conventional manner, this study's distinctive characteristic is our accounting for temporal linkage (dependence) among household purchases using panel data. The temporal dependence arises from state dependence caused by purchase carryover, learning behavior, and other factors, and from household heterogeneity in preferences over different commodities. Most studies on household purchases using panel data have ignored state dependence and household heterogeneity mainly because of the considerable computational burden of evaluating multidimensional integrals. However, both heterogeneity and state dependence can be a serious source of misspecification and ignoring them will, in general, yield inconsistent parameter estimates.

The research background on household purchasing using panel data, as well as the estimation issues, is given in the following section. Next follows the derivation of the econometric model, and model prediction. We then present an empirical application studying household fluid milk purchases, and close with conclusions and directions for future research.

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 the purchase records of consumers over a certain period of time. Panel data provides multiple time-series observations for each household, which 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 use of panel data to study household commodity purchases raises, in general, two issues. The first has to do with the temporal linkage (dependence) of purchasing arising from state dependence caused by the 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.[1] 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 (Hajivassiliou, 1994). Heterogeneity across households persists over time. It may be caused by different preferences, endowments, or attributes (Keane, 1997). Ignoring this temporal dependence in the household purchase process tends to produce inconsistent parameter estimates.

Another important issue in the use of panel data is how to control for sample selection. Sample selection arises from either household self-selection, or data analysts' sample selection decisions. Failure to account for 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 sample. The controlling of sample selectivity has been well addressed for cross-sectional data. However, sample selectivity is an equally acute problem in panel data.[2] In the case of temporal independence, by pooling the data, censoring or sample selection in panel data can be accounted for as with 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 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. To avoid the multidimensional integration, Kyriazidou proposed a Heckman-type two-step approach for obtaining consistent estimates for the panel data sample selection model. Similar research can also be found in Wooldridge and Vella and Verbeek. However, two-step procedures are generally inefficient (Newey). The recent discussion and development of probability simulation methods make maximum-likelihood estimation feasible for use in panel data 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 traditional 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.

The Tobit-type censored model interprets the household zero-purchase outcomes as being the result of strictly economic decisions, i.e., goods are not purchased when they are too expensive (corner solutions). However, not all zero expenditures reflect corner solutions or rationed behavior. Non-purchase could be the result of short-run consumption behavior (infrequency of purchase), or a social, psychological or ethical distinction unconnected to price and income (Deaton and Irish; Jones). As examples, vegetarians do not shun meat because it is expensive and, many non-smokers would not smoke even if tobacco were free (Atkinson, Gomulka, and Stern; Garcia and Labeaga). This suggests that zero expenditure may be best modeled by means of discrete variables altering the nature of individual preferences (Kan and Kao). For household panel data, appropriate temporal aggregation may eliminate the infrequency-of-purchase problem, but the problems of possible corner solutions and double-hurdles will persist.

In this study, we propose a panel data double-hurdle model and use it to study household milk purchases over time. This double-hurdle model is an extension of the panel data Tobit model used by Hajivassiliou in studying the external debt crises of developing countries. In addition to the panel data Tobit model, a discrete equation is defined to determine the participation decisions. For each given time unit, we observe whether the household purchases, and if it does, the amount. The model provides information on what variables influence the consumer's discrete decision of whether or not to participate in the market within a particular shopping period and, if the consumer participates, the continuous decision of how much to purchase.

Econometric Model

Consider a panel of *N* households whose dairy product purchases are observed over *T* time periods. This yields a data array for the i^{th} household, y_i and x_i, where y_i is a *T X 1* vector of observed purchases and x_i is a *T X K* matrix of exogenous market-related, household-specific, price, and advertising variables. A censored-type model is assumed in this study as,

$$(1) \qquad \begin{aligned} y_{it}^{*} &= x_{it}\,\beta + u_{it}\,, \\ y_{it} &= \begin{cases} y_{it}^{*}, & \text{if } u_{it} > -\,x_{it}\,\beta \\ 0, & \text{otherwise} \end{cases} \qquad i = 1, \ldots, N\,;\ t = 1, \ldots, T \end{aligned}$$

where y_{it} is the i^{th} household's purchase of dairy product at time *t*. y_{it}^{*} is the latent variable of y_{it}, β is a *K X 1* vector of estimated parameters, and u_{it} is an error term. We assume u_{it} is jointly distributed normal over *t* with a mean vector of zero and variance-covariance matrix Ω_{i}. Under this model, $y_{it} = 0$ is the corner solution to the consumer's utility maximization problem. However, a zero-valued purchase outcome may include two possible cases: (1) the typical corner solution, driven by price being above the household's reservation price, or (2) a household's decision not to participate in the market. Accordingly, we adopt the double-hurdle model of consumer purchase behavior originally presented by Cragg, reviewed by Blundell and Meghir, and recently applied to a variety of household-based analyses of consumer demand (Jones; Blaylock and Blisard; Yen and Jones; Dong and Gould). Under this double-hurdle model, only market participants determine demand curve parameters.

Following Pudney, one can model the discrete participation decision using the familiar probit structure extended to the panel data:

$$(2) \qquad D_{it} = \begin{cases} 1, & \text{if } e_{it} > -\,z_{it}\,\gamma \\ 0, & \text{otherwise} \end{cases} \qquad i = 1, \ldots, N\,;\ t = 1, \ldots, T$$

where z_{it} is a vector of exogenous variables, γ is a vector of estimated coefficients, and e_{it} an error term. If household i is not a potential purchaser, then $D_i = 0$ for all t; otherwise $D_i = 1$ for at least one time period. We assume e_{it} is jointly distributed normal over t with a mean vector of zero and variance-covariance matrix Σ_i and independent of u_{it}.

The likelihood function for the i^{th} household for the above model can be represented as

$$\text{(3)} \qquad \mathrm{L_i} = (1 - D_i) \cdot prob(y_i = 0) + D_i \cdot prob(D_i = 1) \int_{\mathrm{U}(y_i)} \phi\left(u_i ; \Omega_\mathrm{i}\right) \mathrm{d}\, u_i \,, \; i = 1, \ldots, N$$

where ϕ is the probability density function (pdf) of multivariate normal and $\mathrm{U}(y_i)$ is the probability integration range of u_i given observed y_i. Like y_i, u_i is a vector of $T \; x \; 1$. $prob(D_i = 1)$ is the probability of at least one potential purchase week for household i. $prob(y_i = 0)$ is the probability of zero purchasing, i.e., the household does not purchase in any time period. $prob(y_i = 0)$ consists of two components: (i) not a participant (first hurdle) associated with $prob(D_i = 0)$ for all t; (ii) a participant, but decided not to purchase in any time period (second hurdle) associated with $(1 - prob(D_i = 0)) \cdot prob(y_i^* \leq 0)$. Under this model, the household has to overcome two hurdles before a positive purchase is observed for any time period. That is, the household must be (1) a potential purchaser and, (2) an actual purchaser. Given the multivariate normal distributions of e_{it} and u_{it}, we have the first hurdle:

$$\text{(4)} \qquad prob(D_i = 0) = \int_{\infty}^{-z_i\gamma} \phi\left(e_i ; \Sigma_\mathrm{i}\right) \mathrm{d}\, e_i \;, \; i = 1, \ldots, N \;\; \text{; and the second hurdle:}$$

$$\text{(5)} \qquad (1 - prob(D_i = 0)) \cdot prob(y_i^* \leq 0) = (1 - \int_{\infty}^{-z_i\gamma} \phi\left(e_i ; \Sigma_\mathrm{i}\right) \mathrm{d}\, e_i) \cdot \int_{\infty}^{-x_i\beta} \phi\left(u_i ; \Omega_\mathrm{i}\right) \mathrm{d}\, u_i \;, \; i = 1, \ldots, N \;,$$

where $Z_i\gamma = (Z_{i1}\gamma, Z_{i2}\gamma, \cdots, Z_{it}\gamma)$, $x_i\beta = (x_{i1}\beta, x_{i2}\beta, \cdots, x_{it}\beta)$ and $y_i^* = (y_{i1}^*, y_{i2}^*, \cdots, y_{it}^*)$.

To facilitate the presentation, we can partition the T time-period observations for the i^{th} participating household into two mutually exclusive sets, one containing data associated with the T_{i0} non-purchase time periods and the other containing data associated with the T_{i1} purchase time periods where T=T_{i0}+T_{i1}. Accordingly, the t^{th} household's error term variance-covariance matrix in the purchase equation can be partitioned into the following:

$$(6) \qquad \Omega_i = \begin{bmatrix} \Omega_{i00} & \Omega'_{i01} \\ \Omega_{i01} & \Omega_{i11} \end{bmatrix}$$

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

With this partitioning, the second part of the likelihood function in (3) for the i^{th} household ($i = 1, \ldots, N$) under a particular purchase pattern over T time periods can be simplified as

$$(7) \qquad \mathrm{L_i}(y_{i0}, y_{i1} \mid y_{i0} = 0, y_{i1} > 0) = [\int_{-\infty}^{-z_i\gamma} \int_{-z_i\gamma}^{+\infty} \varphi(e_{i0}, e_{i1}) de_{i0} de_{i1}\,]\phi_1(u_{i1}) \int_{-\infty}^{-x_i\beta} \phi_{0/1}(u_{i0}) du_{i0} \ ,$$

where u_{i0} is the error term vector in (1) associated with the non-purchase time periods and u_{i1} is the error term vector associated with the purchase time periods. Similarly, e_{i0} is the error term vector in (2) associated with the non-purchase time periods and e_{i1} is the error term vector associated with the purchase time periods. ϕ_1 and φ are the multinormal pdfs of u_{i1} and $e_{i = (}e_{i0}$, e_{i1}) with a mean vector zero and variance-covariance matrix Ω_{i11} and Σ_i respectively. $\phi_{0/1}$ is the conditional pdf of u_{i0} given u_{i1} and is distributed multinormal with a mean vector $u_{0/1}$ and variance-covariance matrix $\Omega_{0/1}$; where

$$(8) \qquad u_{0/1} = \Omega_{i01}\Omega_{i11}^{-1}u_{i1}, \quad \Omega_{0/1} = \Omega_{i00} - \Omega_{i01}\Omega_{i11}^{-1}\Omega'_{i01}$$

Then the likelihood function for *N* households can be written as,

$$(9) \qquad \mathrm{L} = \prod_{i=1}^{N} \left[(1-D_i) \cdot prob(y_i = 0) + D_i \cdot \mathrm{L_i} \left(y_{i0}, y_{i1} \,|\, y_{i0} = 0, y_{i1} > 0 \right) \right]$$

Note that when $prob(D_i = 0) = 0$, the double-hurdle model collapses to the Tobit-type model, that is, the first hurdle doesn't exist.

To obtain the maximum likelihood estimates of (9), one needs to evaluate $\int_{-\infty}^{-z_i\gamma} \phi \left(e_i ; \Sigma_\mathrm{i} \right) \mathrm{d}\, e_i$ and $\int_{-\infty}^{-x_i\beta} \phi \left(u_i ; \Omega_\mathrm{i} \right) \mathrm{d}\, u_i$, the T_i-fold integral, and $\int_{-\infty}^{-\mathrm{x_i}\beta} \phi_{0/1} \left(u_{i0} \right) \mathrm{d}\, u_{i0}$ the T_{i0}-fold integral, etc. With an unrestricted Σ_i and Ω_i, the traditional numerical evaluation is computationally intractable when *Ti* exceeds 3 or 4. One conventional approach is to restrict Σ_i and Ω_i to be household and time invariant, thus:

$$(10) \qquad \Sigma_i = E\left(e_{it}\, e_{it}^{'} \right) = \sigma_e^{\,2} I_T \text{ and } \quad \Omega_i = E\left(u_{it}\, u_{it}^{'} \right) = \sigma_u^{\,2} I_T$$

where $\sigma_e^{\,2}$ and $\sigma_u^{\,2}$ are the estimated variance parameters and I_T is a *T*-dimensional identity matrix. This structure yields a pooled cross-sectional double-hurdle model that ignores all intertemporal 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:

$$(11) \qquad u_{it} = \alpha_i + \varepsilon_{it},$$

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:

$$(12) \qquad \varepsilon_{it} = \rho \ \varepsilon_{it\text{-}1} + \nu_{it}; \quad |\rho| < 1 \ ,$$

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

$$(13) \qquad \Omega_i = \sigma_2^2 J_T + \sigma_1^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{T-3} & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{T-4} & \rho^{T-3} \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \cdots & \rho & 1 \end{bmatrix}$$

where J_T is a $T \times T$ matrix of one=s.[3] The term, Ω_i, in (13) 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).

Given e_{it} the same structure of u_{it}, above, we have,

$$(14) \qquad \Sigma_i = \sigma_a^2 J_T + \sigma_e^2 \begin{bmatrix} 1 & \rho_e & \rho_e^{2} & \rho_e^{3} & \cdots & \rho_e^{T-2} & \rho_e^{T-1} \\ \rho_e & 1 & \rho_e & \rho_e^{2} & \cdots & \rho_e^{T-3} & \rho_e^{T-2} \\ \rho_e^{2} & \rho_e & 1 & \rho_e & \cdots & \rho_e^{T-4} & \rho_e^{T-3} \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & \cdots & . & . \\ \rho_e^{T-1} & \rho_e^{T-2} & \rho_e^{T-3} & \rho_e^{T-4} & \cdots & \rho_e & 1 \end{bmatrix}$$

where σ_a^2 and σ_e^2 are the corresponding parts of σ_2^2 and σ_1^2 in (13), respectively. Since Σ_i is the variance-covariance matrix of e_{it}, the error term in the probit equation, σ_a^2 and σ_e^2 are not identified from each other. To solve this problem, one can assume $\sigma_a^2 + \sigma_e^2 = 1$ (Hajivassiliou and Ruud). In the empirical work of this study, we assume $\sigma_e^2 = 1$.

With Ω_i and Σ_i as given in (13) and (14), the likelihood function in (9) requires the evaluation of the T_i-fold and T_{i0}-fold integrals. Note that T_{i0} varies across households. When T_i exceeds 3 or 4, as aforementioned, the evaluation of these multidimensional 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 Hijivassiliou; Keane, 1994; Hijivassiliou, McFadden and Rudd; Geweke, Keane and Runkle). The smooth recursive conditioning simulator (GHK) proposed by Geweke, Hajivassiliou and McFadden, and Keane (1994) is chosen for this study because this algorithm was the most reliable simulator among those examined by Hajivassiliou, McFadden and Rudd.[4]

Prediction Ability of the Model

This double-hurdle model has the ability to predict both the static purchase and the dynamic purchase. Given a certain time period *t*, the static expected purchases and purchase probabilities of model (1) can be derived as follows:

$$(15) \qquad E(y_{it}) = [\Phi(\frac{z_{it}\gamma}{\sqrt{1+\sigma_e^2}})]^2 \cdot \{ \Phi(\frac{x_{it}\beta}{\sqrt{\sigma_1^2+\sigma_2^2}}) \cdot x_{it}\beta + \sqrt{\sigma_1^2+\sigma_2^2} \cdot \phi(\frac{x_{it}\beta}{\sqrt{\sigma_1^2+\sigma_2^2}})\},$$

$$(16) \qquad E(y_{it} \mid y_{it} > 0) = \Phi(\frac{z_{it}\gamma}{\sqrt{1+\sigma_e^2}}) \cdot \{ \ x_{it}\beta + \sqrt{\sigma_1^2+\sigma_2^2} \cdot \frac{\phi(\frac{x_{it}\beta}{\sqrt{\sigma_1^2+\sigma_2^2}})}{\Phi(\frac{x_{it}\beta}{\sqrt{\sigma_1^2+\sigma_2^2}})}\},$$

$$(17) \qquad Prob\,(y_{it} > 0) = \Phi(\frac{z_{it}\gamma}{\sqrt{1+\sigma_e^2}}) \cdot \Phi(\frac{x_{it}\beta}{\sqrt{\sigma_1^2+\sigma_2^2}}),$$

$$(18) \qquad Prob\,(D_{it} = 1) = \Phi(\frac{z_{it}\gamma}{\sqrt{1+\sigma_e^2}})$$

where ϕ(.) is the standard normal pdf and Φ(.) is the standard normal cdf. Equation (15) is the unconditional expected purchases of household *i* at time *t*, (16) is the conditional expected purchases given a purchase occasion, and (17) is the expected probability of purchase. It is clear that (15) is the product of (16) and (17). 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). Equation (18) is the probability of participation for household *i* at time *t*. Each household must overcome two hurdles to have a positive purchase for every time *t*, i.e., both (18) and the second factor in (17) must be non-zero. The second factor in (17) represents the purchase probability given participation.

In order to take advantage of the dynamic nature of the model, we derived the following expected probabilities.

(19) $Prob\,(y_{it} > 0 \mid y_{it-1} = 0)$ and

(20) $Prob\,(y_{it} > 0 \mid y_{it-1} > 0)$.[5]

Equation (19) represents the purchase probability given a non-purchase occasion during the last time period and (20) represents the purchase probability given a purchase occasion during the last period. A larger number for (19) or (20) implies a higher purchase incidence, or a shorter purchase time, while a smaller number for (19) or (20) implies a lower purchase incidence, or a longer purchase time.

Both (19) and (20) are determined by the correlation between current purchase (y_{it}) and last purchase ($y_{it\text{-}1}$). If there is no correlation between y_{it} and $y_{it\text{-}1}$, (19) and (20) are the same and equal to (17), the probability of current purchasing. This implies that the last purchase has no impact on the current purchase.

Empirical Model of Milk Purchases

In this empirical application, we follow a panel of upstate New York households over a four-year period from 1996 through 1999. For each given time unit, we observe whether the household buys fluid milk, and if it does, the amount. We focus particularly on the hurdle equation to see if the non-economic barrier exists in milk purchases. We are also 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 the ACNielsen Homescan Panel[6] for upstate New York (excluding New York City) households, including household purchase information for fluid milk products and annual demographic information. The purchase data is purchase-occasion data collected by 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. To eliminate the possible infrequency-of-purchase problem, the weekly purchases are then aggregated to monthly purchases. Monthly generic-fluid-milk advertising expenditures for upstate New York are obtained from Dairy Management, Inc., and the American Dairy Association and Dairy Council (ADADC). The two data sets of purchase and advertising are merged over a 48-months period from January 1996 through December 1999 for 1,320 households. Generic advertising expenditures vary over time, but not across households. The total number of observations in this sample is 63,360 (48 x 1320).

This application of the panel data double-hurdle model discussed above is concerned with monthly purchases of fluid milk for home consumption only. The monthly household 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 month. We selected the last 3 months of data for 1996 and the 36-month data from 1997 through 1999 to estimate the econometric model specified above, and used the data from the first 9 months of 1996 to derive the lag advertising variables (as described below). The dependent variables in our model are household fluid milk purchase quantities. Among the 1,320 households, 16 did not purchase any fluid milk in the whole time period. Among the purchase households, on average 30 of the 39 months are purchase occasions with a mean purchase of 3.32 gallons over all months and 3.67 gallons for purchase months.

Advertising and Price

Generic advertising used in this analysis includes monthly national and upstate New York milk advertising expenditures aggregated over all media types. The effect of advertising on consumers' behavior could last as long as 9 months (Clarke). In this analysis, the advertising expenditures are lagged 9 months and a polynomial lag model is adopted as following:

$$(21) \qquad A^{*} = \sum_{i=0}^{L} \omega_i A_{t-i} ,$$

where A_{t-i} is the i^{th} lag of advertising, L the total number of lags, and $\omega_i = 1 + (1 - \frac{1}{L})i - \frac{1}{L} i^2$, the quadratic weights of the lag advertising. Three point restrictions are imposed in ω_i: (i) the weight of current advertising is 1, that is, $\omega_0 = 1$ if $i = 0$; (ii) the weight of the 9th lag is 0, that is, the effect of advertising ends at the 9th month; 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^{*} represents the long-term effect of advertising.

Prices are not observed directly in the household scanner panel data. An estimate of price can be obtained by dividing reported expenditures by quantity for the purchase months. However, many studies (e.g., Theil; Cox and Wohlgenant; and Dong, Shonkwiler and Capps) have recognized 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 months. A number of alternative approaches can be used to obtain estimates of the missing prices. In this analysis, we assume a zero-order correction for the missing prices. For each household the imputed prices for non-purchase months are set equal to the mean price of the purchase months for that household. If the household did not purchase over the whole period, the monthly mean prices over all the households are used.

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 expenditure and price per gallon used in this analysis.

Empirical Findings

We obtained the parameter estimates by maximizing the likelihood function in (9) using the GAUSS software system. Numerical gradients of (9) were used in the optimization algorithm proposed by Berndt, Hall, Hall, and Hausman. The standard errors of the estimated parameters 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.

Since the participation equation captures only the non-economic factors that influence the household decision to join the market, as stated earlier, the price and income are excluded from the participation equation. However, advertising is assumed to alter household preferences, and therefore to impact the participation equation. The estimated coefficients for the participation equation and for the purchase equation are presented in Table 2. As expected, generic advertising has a positive and statistically significant impact on participation. Other variables with positive and significant effects on participation include: percentage of teenagers in the household, percentage of persons over 65 years of age in the household, household size, education level of household head, households living in metropolitan areas, and middle-aged couples with no children. Significant household characteristics negatively associated with participation included: African American households and households where the mother is employed. The direction of impacts of all household characteristics was consistent with our a priori expectations.

Since the double-hurdle model nests the Tobit model, we perform a likelihood ratio test between the two. The test rejects the hypothesis that the first hurdle does not exist. This result implies that the double-hurdle model is a significant improvement over the Tobit model. In other words, some of the zero milk purchases are due to non-economic reasons. However, the expected probability of overcoming the first hurdle is 0.91, which implies that most of the zero purchases are due to the second hurdle, i.e., the economic reason.

In contrast to the participation equation, the purchase equation captures the economic factors that affect household purchases. In Table 2, we see that both household income and generic advertising have positive and significant effects on household milk purchase, as we had expected. We also found household size to be positively related to milk purchase, whereas milk price is negatively related. The percentages of teenage girls and elderly persons in the household have a positive impact on household milk purchases. Consistent with the findings of previous studies, single-person households purchase more, while middle-aged couples without children purchase less. Surprisingly, the percentage of children under 12-years-old in the household had a negative effect on household milk purchases. However, this variable was positively related to participation, as discussed above. This may indicate that the households with more children were likely to participate in the milk market, but given their participation, adults would consume more than children. Relative to white households, Hispanic households have higher, African Americans have lower, and Asian households have the same level of milk purchases. As with participation, the employment status of the female head of household is negatively related to milk purchase.

Habit persistence is found in both the purchase and participation equations from the statistically significant estimates of ρ's and σ's. In fact, the correlation coefficient between current purchase (y_{it}) and last purchase ($y_{it\text{-}1}$) is $\frac{\sigma_1^2 \rho + \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$ = 0.7619, and that between the current and last participations is $\frac{\rho_e + \sigma_a^2}{1 + \sigma_a^2}$ = 0.1168. This result means that lagged purchases and participation are positively related to current purchase and participation, respectively. However, more temporal dependence is found in the purchase equation than in the participation equation. This difference indicates that purchase relies more on previous behavior than does participation. Further, for the purchase equation, the component of temporal correlation associated with serial state dependence is 0.0716 ($\frac{\sigma_1^2 \rho}{\sigma_1^2 + \sigma_2^2}$), and the component of this correlation associated with the household heterogeneity is 0.6903 ($\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$). The positive values imply 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. We see that, in this purchase equation, most of the correlation comes from the household heterogeneity. This results from the difference in household preferences for milk: household A prefers to drink more fluid milk than household B does.

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 last month in the sample evaluated at the household sample mean with respect to (15)-(20) are presented in Table 3. The elasticities of the second and the twentieth month are also computed, with results quite close to the last month results.

The long-run elasticity of generic milk advertising is 0.149 (Table 3). In other words, a 1% increase in generic advertising would increase household milk purchases by 0.149%, on average. The 0.149% increase in household purchase counts as 0.058% (38.9%) from the increase of household milk purchase probability and 0.091% (61.1%) 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, about 40% of the effect comes from purchase incidence. The elasticity of advertising on participation is 0.0065, which contributes 11.2% to the elasticity of the positive purchase probability (0.058). This implies that advertising increases the purchase probability mostly (88.8%) by overcoming the second hurdle. This finding allows us to interpret the effects of advertising as follows. If milk is not in the household's preference function, advertising may convince them to include it (the first hurdle). Also, if milk is already in the household's preference function, advertising may increase the weight the household places on it (the second hurdle).

As expected, the price elasticity is negative and inelastic at -0.078. The income effects are relatively low, while household size has a much more prominent effect. 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 more than through the increase in the probability of purchase.

The last two columns in Table 3 indicate that the elasticities of current purchase probability vary depending upon whether a purchase occurred during the last period: the results were more elastic when there was no purchase occasion than when there was a purchase occasion. The positive value will increase the purchase probability that would increase purchase incidence, or reduce the inter-purchase time. For example, a 1% decrease in price would increase the current purchase probability by 0.1128%, given a non-purchase occasion, and by 0.0203%, given a purchase occasion, during the last period. In both cases, the inter-purchase time tends to shrink.

Summary and Conclusions

In this study, we developed a panel data double-hurdle model to estimate the effects on a household's decision of whether to purchase and how much to purchase of the advertised commodity, fluid milk. The model is a dynamic extension of Cragg's conventional double-hurdle 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 state dependence and household specific-heterogeneity. In addition, a discrete equation is specified to determine the participation decision.

In the empirical application, we found that generic dairy advertising could increase the probability of market participation; that is, advertising attracts new participants into the dairy market. Temporal dependence was found to be statistically significant in both purchase and participation equations. However, purchases are much more dependent on previous behavior than is participation. Generic advertising was also found to increase simultaneously the purchase quantity and purchase incidence. In addition, advertising increases the purchase probability more given non-purchase in the prior time period, than if a purchase occasion occurred, which is an intuitively appealing result.

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 in Equations (1) and (2)

Variables	Unit	Mean	Std. Dev.
Household income	$ 000	42.4	28.8
Household size	Number	2.50	1.41
Percentage of children under 12	Number	8.95	18.9
Percentage of girls aged 13-17	Number	2.06	7.95
Percentage of boys aged 13-17	Number	2.16	8.31
Percentage of persons aged 65 and above	Number	21.96	39.7
Head of household has high school degree or above	1/0	0.32	0.47
Age of head of household	Number	52.1	14.6
Black household	1/0	0.02	0.16
Hispanic household	1/0	0.02	0.14
Asian household	1/0	0.003	0.06
Mother of household works	1/0	0.43	0.50
Metropolitan	1/0	0.90	0.30
Middle-aged (35-64) couple without kids	1/0	0.27	0.45
Single-person household	1/0	0.24	0.43
Purchase-Related Variables			
Price	$/Gallon	2.36	0.71
Advertising	$000	110.1	28.6
Sum of weighted lag advertising (*A**)	$000,000	2.04	0.28

Table 2. Estimated Parameters

Variable	Participation Equation		Purchase Equation	
	Estimate	Std. Error	Estimate	Std. Error
Intercept	1.9462*	0.0459	4.6417*	0.0622
Household Characteristics				
Log household income	--	--	0.0337*	0.0059
Inverse household size	-0.1196*	0.0152	-2.3673*	0.2606
Percentage of children under 12	0.0563	0.0411	-0.3329*	0.1425
Percentage of girls aged 13-17	0.2339*	0.0533	0.3472*	0.0572
Percentage of boys aged 13-17	0.2645*	0.1212	-0.1119*	0.0528
Percentage of persons aged 65 and above	0.0443*	0.0131	0.1030*	0.0216
Head of household has high school degree or above	0.0346*	0.0114	-0.0381*	0.0142
Age of head of household	0.0017	0.0010	-0.0080*	0.0026
Black household	-0.1755*	0.0976	-0.2148*	0.1046
Hispanic household	0.0945	0.1095	0.0300*	0.0094
Asian household	-0.0753	0.0889	0.7622	0.5783
Mother of household works	-0.0845*	0.0185	-0.1826*	0.0338
Metropolitan	0.0384*	0.0059	-0.3192*	0.0710
Middle-aged (35-64) couple without kids	0.0703*	0.0100	-0.0070*	0.0031
Singl- person household	0.0144	0.0074	0.5851*	0.1668
Purchase Characteristics				
Log price	--	--	-0.3993*	0.0567
Sum of weighted lag advertising (A*)	0.0740*	0.0145	0.2417*	0.0239
Regression Coefficients				
Standard error 1 (σ_1)	--	--	1.7908*	0.0059
Standard error 2 (σ_2)	0.1452*	0.0305	2.6741*	0.0514
Auto correlation coefficient (ρ)	0.0962*	0.0255	0.2310*	0.0043

“*” indicates significance at the 0.05 level or higher

Table 3 Elasticities

		Type of Expected Value					
		$Prob(D=1)$ Prob. of Particip.	$Prob(y_t > 0)$ Prob. of Purchase	$E(y_t \| y_t > 0)$ Cond. Purchase	$E(y_t)$ Uncond. Purchase	$Prob(y_t \| y_{t-1} > 0)$	$Prob(y_t \mid y_{t-1} = 0)$
Elasticity	Advertising	0.0065*	0.0582*	0.0905*	0.1487*	0.0418*	0.1397*
	Price	---	-0.0297*	-0.0483*	-0.0780*	-0.0203*	-0.1128*
	Income	---	0.0032	0.0051	0.0083	0.0022	0.0120
	Household size	0.0025*	0.1230*	0.1981*	0.3211*	0.0848*	0.4353*
	Age	0.0032*	-0.0350	-0.0588*	-0.0938	-0.0229	-0.1732

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

References

Atkinson, A.B., J. Gomulka, and N.H. Stern. "Spending on Alcohol: Evidence from the Family Expenditure Survey." *The Economic Journal* 100(1990): 808-27.

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

Blaylock, J.R., and W.N. Blisard. "Women and the Demand for Alcohol: Estimating Participation and Consumption." *Journal of Consumer Affairs*, 27(1993): 319-34.

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

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, J. "Evaluation of Multivariate Normal Probability Integrals Using a Low Variance Simulator." *Review of Economics and Statistics* (Nov., 1994): 673-82.

Clarke, D. "Econometric Measurement of the Duration of Advertising Effect on Sales." *Journal of Marketing Research* (November, 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.

Cragg, J.G. "Some Statistical Models for Limited Dependent Variables with Applications to the Demand for Durable Goods." *Econometrica* 39(1971): 829-44.

Deaton, A.S. and M. Irish. "Statistical Models for Zero Expenditures in Household Budgets." *Journal of Public Economics* 23(1984): 59-80.

Dong, D., and B.W. Gould. "Quality versus Quantity in Mexican Household Poultry and Pork Purchases." Agribusiness 16(2000): 333-55.

Dong, D., J.S. Shonkwhiler, and O. Capps. "Estimation of Demand Functions Using Cross-Sectional Household Data: The Problem Revisited." *American Journal of Agricultural Economics* 80(1998): 466-73.

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.

Garcia, J. and J.M. Labeaga. "Alternative Approaches to Modeling Zero Expenditures: An Application to Spanish Demand for Tobacco." *Oxford Bulletin of Economics and Statistics* 58(1996): 489-503.

Geweke, J.F. "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints", in *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.

Gould, B.W.and D. Dong. "The Decision of When to Buy a Frequently Purchased Good: A Muti-Period Probit Model." *Journal of Agricultural and Resource Economics* 25(2000): 636-52.

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. McFaddan. “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.

Heckman, J. “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models.” *Annals of Economic and Social Measurement* 5(1976): 475-92.

Jones, A. “A Double-hurdle Model of Cigarette Consumption.” *Journal of Applied Econometrics* 4(1989): 23-39.

Kan, K. and C. Kao. “A Maximum Simulated Likelihood Estimation of Consumer Demand Systems with Zero Expenditures: A Double-hurdle Model.” Working Paper, Department of Economics, Syracuse University, 1996.

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.

Kyriazidou, E. "Estimation of Panel Data Sample Selection Model." *Econometrica* 65(November, 1997): 1335-64.

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.

Newey, W. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36(1987): 231-50.

Pudney, S. "Modeling Individual Choice: The Econometrics of Corners, Kinks and Holes." New York: Basil Blackwell Ltd., 1989.

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.

Vella, F. and M. Verbeek. "Two-step Estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias." *Journal of Econometrics* 90(1999): 239-63.

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

Wooldridge, J.M. "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions." *Journal of Econometrics* 68(1995): 115-32.

Yen, S.T. and A. Jones. "Household Consumption of Cheese: An Inverse Hyperbolic Sine Double-Hurdle Model with Dependent Errors." *American Journal of Agricultural Economics* 79(1997): 246-51.

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

Endnotes:

[1] This dynamic version of Tobit model has gained much attention recently (e. g., Zeger and Brookmeyer, Lee, and Wei).

[2] In addition, panel data sets are commonly characterized by non-randomly missing observations due to sample attrition (Kyriazidou).

[3] This one-factor effect plus AR (1) error structure was also used by Hajivassiliou and Ruud; Hajivassiliou (1994); and Gould and Dong.

[4] A brief overview of the GHK simulation algorithm can be found in the Appendix of Gould and Dong.

[5] Results and derivations are available from authors upon request.

[6] Copyright 2000 by ACNielsen