DISCRETE/CONTINUOUS MODELS OF CONSUMER DEMAND WITH BINDING NON-NEGATIVITY CONSTRAINTS

by

Jeongwen Chiang and Lung-Fei Lee

Discussion Paper No. 261, October 1990

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, MN 55455

DISCRETE/CONTINUOUS MODELS OF CONSUMER DEMAND WITH BINDING NON-NEGATIVITY CONSTRAINTS*

by

Jeongwen Chiang

William E. Simon Graduate School
of Business Administration
University of Rochester

and

Lung-Fei Lee

Department of Economics
University of Minnesota

October 1990

ABSTRACT

Consumers in this model are assumed to maximize utility with respect to two goods, one of which is inessential and available in several alternative forms. Decisions rules for the consumers are derived for two cases: choosing one alternative or forgoing the consumption of the inessential good. Choice probabilities, the demand equations and the zero-consumption probability are derived and reflect the interrelatedness of the discrete/continuous choices and the binding non-negativity constraint. We show that traditional discrete/continuous choice models yield estimates of demand which are biased and choice probabilities, which in some circumstances, are not biased. The model is then applied to coffee data.

Correspondence Address:

William E. Simon Graduate School of Business Administration
University of Rochester, Rochester, NY 14627

* This is a revised version of the paper, "A Model of Simultaneous Brand and Purchase Quantity Decisions - With and Without A Purchase Option", presented at the 6-th World Congress of the Econometric Society, Barcelona, Spain, 1990. We thank Arnold Zellner, Peter Rossi and other seminar participants for valuable comments at the University of Chicago. The revision benefitted greatly from the comments of two anonymous referees.

1. Introduction

Consumers who are considering the purchase of a good, available in a variety of substitutable forms, make three related decisions which are often recorded in disaggregate data: a decision to consume or not to consume the good, a selection of one or more alternatives as well as the corresponding quantities in the event of consumption. Observations in which consumers forgo the product have been referred to as 'zero-consumption' points while consumption data has been described as involving 'mixed discrete/continuous choices'. Although econometric models of consumer demand already acknowledge the interrelatedness of discrete and continuous choices, they ignore the possible ties of these choices to the consumption decision by omitting zero-consumption observations. Some have justified the exclusion of zeros by arguing that they most likely reflect a high level of time disaggregation and thus diminish in importance for long investigative periods (Deaton and Irish (1984)). Frequently, though, the essence of a demand analysis involves consumers' responses to short-lived events: brand and/or product promotions are common examples. In these cases, zeros are legitimate observations which are likely derived from the same stimuli as discrete/continuous choices.

Models exist which handle either zero-consumption (Wales and Woodland (1983), Lee and Pitt (1986, 1987)) or discrete/continuous choices (King (1980), Dubin and McFadden (1984) and Hanemann (1984)) but not both. We propose an econometric consumer demand model in which all three decisions are simultaneously considered and so that their interdependence can be examined. We focus on three aspects of the model: first, the implementation of the three decisions in a consumer's utility maximization problem while maintaining computational tractability of the estimating equations; second, a theoretical examination of the consequences of omitting zero-consumption observations, and finally, an illustration by way of empirical application.

The paper is organized as follows. Section 2 presents the model and the MLE method for the estimation. Choice probabilities and demands with self-selection with and without the zero-consumption option are discussed in Section 3. The two-stage estimation method is discussed in Section 4. The empirical application, which uses coffee purchase data, is in Section 5. The conclusion is found in Section 6.

2. Random Utility Model and Decision Rules

Following Deaton and Muellbauer (1980) and Hanemann (1984), the utility of a representative consumer is defined over the product of the quantity and perceived quality for each of two goods. The first good is available in k alternative forms, which are perfect substitutes, and is assumed to be inessential to the consumer. The second good is a composite good and is assumed to be essential. The utility function is assumed to be strictly quasi-concave in its arguments and has the form:[1]

$$(2.1) \qquad U\left(\sum_{j=1}^{k} \psi_j x_j, \psi_{k+1} x_{k+1}\right)$$

where x_j denotes the consumption of alternative j, $j=1,...,k$, and x_{k+1} is the consumption of the composite good. Let ψ_j, $j=1,...,k+1$ denote the subjective evaluation of good j by the representative consumer. Even though these evaluations are deterministic from the consumer's point of view, they have unobservable components and are therefore regarded in what follows as random variables. Let the joint distribution of the ψ_j be denoted $F_\psi(\psi_1, ..., \psi_{k+1})$.

The maximization of (2.1) is subject to the non-negativity constraints, $x_j \geq 0$, $j=1, ..., k+1$, and the budget constraint, $\Sigma_{j=1}^{k+1} p_j x_j \leq y$, where p_j, $j=1, ..., k+1$ denotes the price of good j and y denotes the budget of the representative consumer. Therefore, the Lagrangean function corresponding to the maximization problem is given by

$$(2.2) \qquad L = U\left(\sum_{j=1}^{k} \psi_j x_j, \psi_{k+1} x_{k+1}\right) + \lambda\left(y - \sum_{j=1}^{k+1} p_j x_j\right)$$

where λ is the multiplier.

---

[1] Hanemann (1984) provides alternative forms for the first argument. This analysis can be easily modified to accommodate these alternatives.

There are two mutually exclusive consumption outcomes: either (case 1) zero-consumption of the first good occurs and none of the k alternatives are selected, or (case 2) one of the competing alternatives is selected.[2] Let z_f and z_s denote, respectively, the first and second arguments of the utility function so that $z_f = \psi_f x_f$ and $z_s = \psi_{k+1} x_{k+1}$. Let z denote the vector (z_f, z_s).

2.1 Case 1: Zero-Consumption

Let z_0^* denote the optimizing consumption vector in the event of zero-consumption for the first good so that $z_0^* = (0, \psi_{k+1} x_{k+1}^*)$, where $x_{k+1}^* = y/p_{k+1}$. Differentiating (2.2) with respect to x_j, $j = 1, \ldots, k+1$, the Kuhn-Tucker conditions at z_0^* are

$$\frac{\partial U(z_0^*)}{\partial z_f} \psi_j - \lambda p_j \leq 0 \qquad j = 1, \ldots, k \tag{2.3a}$$

$$\frac{\partial U(z_0^*)}{\partial z_s} \psi_{k+1} - \lambda p_{k+1} = 0 \tag{2.3b}$$

These conditions can be rewritten by using the concept of virtual prices, developed in the quantity rationing literature (Rothbarth (1941), Neary and Roberts (1980)). Let ξ_f and ξ_s, the virtual prices of the two goods, be defined by

$$\xi_f(z_0^*) = \frac{\partial U(z^*)}{\partial z_f} / \lambda \tag{2.4a}$$

$$\xi_s(z_0^*) = \frac{\partial U(z^*)}{\partial z_s} / \lambda \tag{2.4b}$$

Substituting (2.4a) and (2.4b) into (2.3a) and (2.3b), respectively, gives

[2] The probability of selecting multiple alternatives is of measure zero because of the linear structure of the first argument and the assumption that ψ is stochastic.

(2.5a) $\xi_f(z_0^*) \le p_j/\psi_j \qquad j = 1,\dots,k$

(2.5b) $\xi_s(z_0^*) = p_{k+1}/\psi_{k+1}$

Since the ψ_j are deterministic from the point of view of the consumer, condition (2.5a) implies that if the minimum adjusted price, $\min\{p_j/\psi_j,\ j=1,\dots,k\}$, exceeds $\xi_f(z_0^*)$ then zero-consumption results. Hence, $\xi_f(z_0^*)$ is a reservation price.

2.2 Case 2: One Alternative Is Selected

Let h denote the selected alternative and let z_h^* denote the optimizing consumption vector when the first good is consumed so that $z_h^* = (\psi_h x_h^*,\ \psi_{k+1} x_{k+1}^*)$. The Kuhn-Tucker conditions at z_h^* are:

(2.6a) $\dfrac{\partial U(z_h^*)}{\partial z_f}\psi_h - \lambda p_h = 0$

(2.6b) $\dfrac{\partial U(z_h^*)}{\partial z_f}\psi_j - \lambda p_j \le 0 \qquad j = 1,\dots,h-1,h+1,\dots,k$

(2.6c) $\dfrac{\partial U(z_h^*)}{\partial z_s}\psi_{k+1} - \lambda p_{k+1} = 0$

where, as in case 1, (2.6) can be expressed as:

(2.7a) $\xi_f(z_h^*) = p_h/\psi_h$

(2.7b) $\quad \xi_f(z_h^*) \le p_j/\psi_j \qquad j=1,\ldots,h-1,h+1,\ldots,k$

(2.7c) $\quad \xi_s(z_h^*) = p_{k+1}/\psi_{k+1}$

Since these conditions characterize an interior solution, $\xi(z_h^*)$ equals the minimum adjusted price of the first good. The selected alternative must have the minimum adjusted price because any other selection would reduce the budget set and consequently the utility level of the consumer. Therefore, $p_h/\psi_h = \min\{p_j/\psi_j,\ j=1,\ldots,k\}$.

Cases 1 and 2 imply a simple decision rule for the consumer: when all adjusted prices exceed the reservation price the first good is forgone; and, when one or more adjusted prices are exceeded by the reservation price, the alternative with the minimum adjusted price is selected.

2.3 Maximum Likelihood Estimation

Muellbauer (1974) has shown that demand for the first good, x_f, can be expressed

(2.8) $\quad x_f = Q_f(p_f/\psi_f,\ p_{k+1}/\psi_{k+1},\ y)/\psi_f$

where p_f/ψ_f is the adjusted price of the first good.

Then, for case 1,

(2.9) $\quad 0 = Q_f(\xi_f(z_0^*),\ p_{k+1}/\psi_{k+1},\ y)$

Hence, inverting (2.9), $\xi_f(z_0^*)$ can be expressed as a function of p_{k+1}/ψ_{k+1} and y.

Let $H(\psi_1,\ldots,\psi_k \mid \psi_{k+1})$ denote the conditional distribution for the ψ_j, $j=1,\ldots,k$ and let $f_{k+1}(\psi_{k+1})$ denote the marginal density function for ψ_{k+1}, both derived from $F_\psi(\psi_1, \ldots, \psi_{k+1})$. These distributions are assumed i.i.d. across consumers. Let $H^h(.)$ denote the first partial derivative of $H(\cdot)$ with respect to ψ_h. Using (2.5a) the density, in the zero-consumption case, for individual i, is

$$L_i(z_0^*) = \int H\left(\frac{p_1}{\xi_f},\ldots,\frac{p_k}{\xi_f} \mid \psi_{k+1}\right) f_{k+1}(\psi_{k+1})\, d\psi_{k+1} \tag{2.10}$$

For case 2, equation (2.8) becomes

$$x_h^* = Q_f(p_h/\psi_h, p_{k+1}/\psi_{k+1}, y)/\psi_h \tag{2.11}$$

Inverting (2.11), ψ_h can be expressed as a function of x_h^* and p_h in addition to p_{k+1}/ψ_{k+1} and y. Using (2.7a) and (2.7b), the density, when alternative h is consumed, is

$$L_i(z_h^*) = \int H^h\left(\frac{p_1}{p_h}\psi_h,\ldots,\psi_h,\ldots,\frac{p_k}{p_h}\psi_h \mid \psi_{k+1}\right) |J|\, f_{k+1}(\psi_{k+1})\, d\psi_{k+1} \tag{2.12}$$

where $|J|$ is the jacobian of the transformation from ψ_h to x_h^* in (2.11).

Therefore, the sample likelihood function, in which N_j, $j=1,\ldots,k$ denotes the set of consumers who consume alternative j, and N_{k+1} denotes the set of consumers who forgo the first good, is

$$L(\cdot) = \prod_{j, N_j} L(z_j^*) \prod_{N_{k+1}} L(z_0^*) \tag{2.13}$$

3. Choice Probability and Zero-Consumption Probability

3.1 Choice Probability and Sample Selection Bias

In this section we show a necessary and sufficient condition which ensures the functional forms of the choice probabilities without the zero-consumption option are identical to that of the conditional choice probabilities when the first good is assumed inessential. This implies that only in such cases can the estimating equations for the discrete choices derived without the zero-consumption option be immune from sample selection bias.

From the decision rules described in Section 2.2, alternative h is selected by the consumer whenever the following pair of conditions are satisfied:

$$(3.1) \qquad p_h/\psi_h = \min\{p_j/\psi_j, j\in M\} \quad \text{and} \quad \xi_f(z_0^*) > \min\{p_j/\psi_j, j\in M\}$$

where $M=\{1, ..., k\}$. Therefore, letting D be the discrete choice indicator and I a dichotomous indicator for which $I=0$ corresponds to zero-consumption and $I=1$ corresponds to consumption of the first good, the probability of such a selection is given by

$$(3.2) \qquad P(D=h, I=1) = P\left(p_h/\psi_h = \min\{p_j/\psi_j, j\in M\} \text{ and } \xi_f(z_0^*) > p_h/\psi_h\right)$$

Equation (3.2) can be expressed alternatively as:

$$(3.3) \qquad P(D=h, I=1) = P(\xi_f(z_0^*) > p_h/\psi_h \mid p_h/\psi_h = \min\{p_j/\psi_j, j\in M\}) \cdot P(p_h/\psi_h = \min\{p_j/\psi_j, j\in M\})$$

Suppose that $\xi_f(z_0^*)$ is deterministic. Then, if the conditional distribution of p_h/ψ_h given that $p_h/\psi_h=\min\{p_j/\psi_j, j\in M\}$ is the same for all $h\in M$, then $P(\xi_f(z_0^*)>p_h/\psi_h \mid p_h/\psi_h=\min\{p_j/\psi_j, j\in M\})=C$, a constant for all h. Therefore, factoring this term when summing the choice probabilities given in (3.3) over the k alternatives gives the probability of consumption, or $P(I=1)$, as

$$(3.4) \qquad P(I=1) = \sum_{h=1}^{k} P(D=h, I=1)$$

$$= C \cdot \sum^{k}$$

Since $\Sigma_{h=1}^{k} P(p_h/\psi_h=\min\{p_j/\psi_j, j\in M\})=1$, we have that $P(I=1)=C$. Therefore,

$$(3.5) \qquad P(D=h \mid I=1) = \frac{P(D=h, I=1)}{P(I=1)}$$

$$= P(p_h/\psi_h = \min\{p_j/\psi_j, j\in M\})$$

which states that the conditional choice probability when zero-consumption is considered is identical to the choice probability without the zero-consumption possibility. This conclusion is not altered when $\xi_f(z_0^*)$ is stochastic and independent of the ψ_j, but, the sufficient condition is instead that the probability of $\xi_f(z_0^*)>p_h/\psi_h$ given $p_h/\psi_h=\min\{p_h/\psi_h, j\in M\}$ is a common constant for all $h\in M$. It can be easily shown that both versions of this condition are also necessary.

Two distributions frequently used in discrete choice analyses, the normal and generalized extreme value (GEV) distributions, are next examined to see whether they satisfy the necessary and sufficient condition. Let $\psi_j=\exp(w_j'\tau+\epsilon_j)$, $j\in M$, where $w_j'\tau$ is a product of relevant explanatory variables and their coefficients, and ϵ_j is a random term. Hanemann (1984) has shown that when the ϵ_j are jointly distributed with a normal or GEV distribution, then the choice probability is, respectively, a probit or a logit probability. Under these specifications and taking the logarithm of the conditions in (3.2), (3.2) becomes

$$(3.6) \qquad P(D=h, I=1) = P(\theta_h+\epsilon_h \geq \theta_j+\epsilon_j, j\in M \quad \text{and} \quad \theta_h+\epsilon_h > \kappa)$$

where $\kappa=-\ln\xi_r(z_0^*)$ and $\theta_j=w_j/\tau-\ln p_j$.

Let $H(e)=\exp(-G(e^{-e}))$ denote the joint GEV distribution for the e_j, $j\in M$, where $G(\cdot)$ is a homogeneous function of degree one, and let $H^h(\cdot)$ and $G^h(\cdot)$ denote the first partial derivatives of $H(\cdot)$ and $G(\cdot)$ with respect to the h^{th} argument. Then, the choice probability becomes:

$$
\begin{aligned}
(3.7)\qquad & P(\theta_h+e_h\geq\theta_j+e_j, j\in M \quad \text{and} \quad \theta_h+e_h>\kappa) \\
&= \int_{\kappa-\theta_h}^{\infty} H^h(\theta_h-\theta_1+t,\ldots,t,\ldots,\theta_h-\theta_k+t)dt \\
&= \int_{\kappa-\theta_h}^{\infty} e^{-t}G^h(e^{\theta})\exp[-e^{-t-\theta_h}G(e^{\theta})]dt \\
&= \frac{G^h(e^{\theta})e^{\theta_h}}{G(e^{\theta})}\{1-\exp[-e^{-\kappa}G(e^{\theta})]\}
\end{aligned}
$$

where $e^{\theta}=(\exp(\theta_1),\ldots,\exp(\theta_k))$.

Since $P(\theta_h+e_h\geq\theta_j+e_j,j\in M)=G^h(e^{\theta})\exp(\theta_h)/G(e^{\theta})$ and $P(\theta_h+e_h\geq\theta_j+e_j,j\in M$ and $\theta_h+e_h>\kappa)=P(\theta_h+e_h\geq\theta_j+e_j,j\in M)\cdot P(\theta_h+e_h>\kappa\,|\,\theta_h+e_h\geq\theta_j+e_j,j\in M)$, (3.7) implies that

$$(3.8)\qquad P(\theta_h+e_h>\kappa\,|\,\theta_h+e_h\geq\theta_j+e_j,j\in M)=1-\exp[-e^{-\kappa}G(e^{\theta})]$$

where the expression on the right hand side is a constant for all $h\in M$.[3] Therefore, with a GEV distribution for the e_j, both the conditional choice probability and the choice probability without

[3] Thus, $P(\theta_h+e_h<\kappa\,|\,\theta_h+e_h\geq\theta_j+e_j,j\in M)=\exp[-e^{-\kappa}G(e^{\theta})]$, $h\in M$, which implies that the conditional distribution of θ_h+e_h is a univariate extreme value distribution with location parameter $-\ln G(e^{\theta})$.

the zero-consumption possibility equal $G^h(e^{\theta})e^{\theta_h}/G(e^{\theta})$.[4]

This conclusion cannot be obtained when the e_j have a joint normal distribution because the conditional probability, $P(\theta_h+e_h>\kappa|\theta_h+e_h\geq\theta_j+e_j, j\in M)$, is specific to the alternative rather than a common constant.

3.2. Demand With Self-Selection

To check for differences between the expected demand functions derived with and without zero-consumption possibilities we must specify functional forms for both the utility function and the distribution of the random terms. This is because, with or without the zero-consumption option, the expectation of the demand function (2.11) conditional on the selection of an alternative, involves the conditional expectations of ψ_h and ψ_{k+1}. Without either an explicit functional form for (2.11) or a joint distribution for the random component of ψ_j, e_j, $j=1,...,k+1$, unambiguous analytical results cannot be obtained. Following Lee and Pitt (1986) we choose the Indirect Translog Utility function and the GEV distribution for the e_j since this distribution preserves the conditional choice probabilities.

Let $V(v_f/\psi_f, v_{k+1}/\psi_{k+1})$ denote the Indirect Translog (ITL) utility function. Then,

$$V=\alpha_1(\ln v_f-\ln\psi_f)+\alpha_2(\ln v_{k+1}-\ln\psi_{k+1})+\frac{1}{2}\beta_{11}(\ln v_f-\ln\psi_f)^2+\frac{1}{2}\beta_{22}(\ln v_{k+1}-\ln\psi_{k+1})^2 + \beta_{12}(\ln v_f-\ln\psi_f)(\ln v_{k+1}-\ln\psi_{k+1}) \tag{3.9}$$

where $\alpha_1+\alpha_2=-1$ and v_f and v_{k+1} denote the prices of the two goods normalized by y. The

[4] When $\xi_f(z_0^*)$ is stochastic but independent of the ψ_j, the above derivation will involve an additional integration over the range of the distribution for $\xi_f(z_0^*)$. The same conclusion follows nevertheless.

function V is also assumed to be homogeneous, so that $\beta_{11}+\beta_{12}=0$ and $\beta_{12}+\beta_{22}=0$.[5,6] As before, $\psi_j=\exp(w_j'\tau+e_j)$ for $j\in M$ and in addition $\psi_{k+1}=\exp(w_{k+1}'\tau+e_{k+1})$ where the $w_{k+1}'\tau$ and the e_{k+1} have analogous interpretations. Also, let e_{k+1} be independent of the e_j and have an extreme value distribution.

Let $s_h^*=p_h x_h^*/y$ denote the expenditure share equation for alternative h. Thus, the expected conditional expenditure share equations corresponding to the selection of alternative h with and without zero-consumption option are $E(s_h^*\mid\Omega)$ and $E(s_h^*\mid\Omega')$, respectively, where by Roy's identity

$$(3.10) \qquad s_h^* = -\alpha_1 + \beta_{11}(w_h'\tau - \ln p_h + e_h) + \beta_{11}(-w_{k+1}'\eta + \ln p_{k+1} - e_{k+1})$$

and Ω and Ω' are the sets of e which satisfy the respective decision rules. With our specifications for ψ_j, $j=1,\dots,k$, $\Omega=\{e\mid\theta_h+e_h\ge\theta_j+e_j, j\in M$ and $\theta_h+e_h\ge-\ln\xi_f(z_0^*)\}$ and $\Omega'=\{e\mid\theta_h+e_h\ge\theta_j+e_j, j\in M\}$.

The expression $-\ln\xi_f(z_0^*)$ in Ω can be found in one of two ways: either from the first-order condition (2.4a) or, following Lee and Pitt (1986), by imposing zero-consumption on the demand equation for the first good and solving for the corresponding virtual price. In the latter method, equation (3.10) becomes

$$(3.11) \qquad 0 = -\alpha_1+\beta_{11}(-\ln\xi_f(z_0^*)) + \beta_{11}(-w_{k+1}'\eta + \ln p_{k+1} - e_{k+1})$$

and therefore, $-\ln\xi_f(z_0^*)=\theta_{k+1}+e_{k+1}$ with $\theta_{k+1}=(\alpha_1/\beta_{11})+w_{k+1}'\eta-\ln p_{k+1}$. Consequently, $\Omega=\{e\mid\theta_h+e_h\ge\theta_j+e_j, j\in M$ and $\theta_h+e_h\ge\theta_{k+1}+e_{k+1}\}$.

The difference between $E(s_h^*\mid\Omega)$ and $E(s_h^*\mid\Omega')$ clearly depends upon the difference between $E(e_h-e_{k+1}\mid\Omega)$ and $E(e_h-e_{k+1}\mid\Omega')$. Theorem B.2.2 in Dubin (1985) shows that

---

[5] The Indirect Translog Function needs to satisfy the coherency conditions (see Soest and Kooreman (1990)). In the bivariate case, the homogeneity conditions and the positivity of β_{11} are consistent with the coherency conditions.

[6] The homogeneity restrictions yield simple functional forms for the expenditure equations.

$$E(\epsilon_h \mid \Omega) = \mu + \ln G^h(e^\theta) - \ln P(D=h, I=1)$$

$$E(\epsilon_{k+1} \mid \Omega) = \frac{G(e^\theta) + e^{\theta_{k+1}}}{G(e^\theta)} \left[P(I=1)\, \mu + P(I=0)\, \ln P(I=0) \right]$$

where μ is the Euler's constant and $P(D=h, I=1)$ and $P(I=0)$, respectively, are

$$P(D=h, I=1) = \frac{G^h(e^\theta) \cdot e^{\theta_h}}{G(e^\theta) + e^{\theta_{k+1}}} \qquad h = 1, \ldots, k \tag{3.12}$$

$$P(I=0) = \frac{e^{\theta_{k+1}}}{G(e^\theta) + e^{\theta_{k+1}}} \tag{3.13}$$

Hence, using (3.12), (3.13) and $P(I=0)=1-P(I=1)$, gives

$$\begin{aligned} E(\epsilon_h - \epsilon_{k+1} \mid \Omega) &= \mu + \ln G^h(e^\theta) - \ln\left[\frac{G^h(e^\theta) e^{\theta_h}}{G(e^\theta) + e^{\theta_{k+1}}}\right] - \\ &\frac{G(e^\theta) + e^{\theta_{k+1}}}{G(e^\theta)} \left[\left(1 - \frac{e^{\theta_{k+1}}}{G(e^\theta) + e^{\theta_{k+1}}} \right) \mu + \frac{e^{\theta_{k+1}}}{G(e^\theta) + e^{\theta_{k+1}}} \ln\left(\frac{e^{\theta_{k+1}}}{G(e^\theta) + e^{\theta_{k+1}}} \right) \right] \\ &= -(\theta_h - \theta_{k+1}) - \frac{\ln P(I=0)}{P(I=1)} \end{aligned} \tag{3.14}$$

Applying the same theorem, $E(\epsilon_h \mid \Omega')$ is given by

$$E(\epsilon_h \mid \Omega') = \mu + \ln G^h(e^\theta) - \ln\left(\frac{G^h(e^\theta) e^{\theta_h}}{G(e^\theta)} \right)$$

whereas $E(\epsilon_{k+1} \mid \Omega') = \mu$ since ϵ_{k+1} is not in Ω'. Thus,

$$E(\epsilon_h - \epsilon_{k+1} \mid \Omega') = \ln G^h(e^\theta) - \ln\left(\frac{G^h(e^\theta) e^{\theta_h}}{G(e^\theta)} \right) \tag{3.15}$$

Subsequently, the expenditure share equation (3.10) with and without the zero-consumption option can be expressed, respectively, as

$$
\begin{aligned}
(3.16)\qquad s_h^* &= E(s_h^* \mid \Omega) + u_h \\
&= \beta_{11}(\theta_h - \theta_{k+1}) + \beta_{11}E(e_h - e_{k+1} \mid \Omega) + u_h \\
&= \beta_{11}\left(\frac{-\ln P(I=0)}{P(I=1)}\right) + u_h
\end{aligned}
$$

and

$$
\begin{aligned}
(3.17)\qquad s_h^* &= E(s_h^* \mid \Omega') + u_h' \\
&= \beta_{11}(\theta_h - \theta_{k+1}) + \beta_{11}E(e_h - e_{k+1} \mid \Omega') + u_h' \\
&= \beta_{11}[\ln G(e^\theta) - \theta_{k+1}] + u_h'
\end{aligned}
$$

where $u_h = \beta_{11}\{e_h - e_{k+1} - E(e_h - e_{k+1} \mid \Omega)\}$ and $u_h' = \beta_{11}\{e_h - e_{k+1} - E(e_h - e_{k+1} \mid \Omega')\}$.

If the model with the zero-consumption option is the correct model, equation (3.17) can be further rewritten as

$$
\begin{aligned}
(3.18)\qquad s_h^* &= \beta_{11}\{\ln\left(\frac{G(e^\theta)}{G(e^\theta) + e^{\theta_{k+1}}}\right) - \ln\left(\frac{e^{\theta_{k+1}}}{G(e^\theta) + e^{\theta_{k+1}}}\right)\} + u_h' \\
&= \beta_{11}\big(\ln P(I=1) - \ln P(I=0)\big) + u_h'
\end{aligned}
$$

where, since $\ln P(I=1) < 0$, the following inequalities hold:

$$
\big(\ln P(I=1) - \ln P(I=0)\big) < \frac{\ln P(I=1) - \ln P(I=0)}{P(I=1)} < \frac{-\ln P(I=0)}{P(I=1)}
$$

Comparing (3.16) and (3.18) and using this inequality, the conditional expected expenditure share without the zero-consumption option is biased downward since β_{11} is positive by the coherency conditions. Thus, though the discrete choice probabilities without the zero-consumption option are not biased with the GEV distribution, the conditional expected demands under this distribution are biased.

4. Two-Stage Estimation

With the ITL utility function and the GEV distribution, we have shown that the model yields an endogenous switching regression model with a limited dependent variable, characterized by the continuous and discrete choices given, respectively, by equations (3.16) and (3.12), and the zero-consumption probability of equation (3.13). We found that rather than being an alternative to the set of discrete choices, zero-consumption can be regarded as a member of the set of discrete choices. In this case, then, including binding non-negativity constraints to the discrete/continuous choice framework yields probabilities which resemble those of the GEV discrete choice model.[7]

Furthermore, for this specification an alternative estimation procedure, the two-stage method, can be used. Since (3.12) and (3.13) yield a GEV model, the first stage is to obtain consistent estimates of τ, η and α_1/β_{11} by maximum likelihood method from the choice probabilities. The estimated values of these parameters can then be used to form the regressor in (3.16). The second stage uses a sub-sample, in which each consumer chooses the same alternative, to obtain a consistent estimate of β_{11} by OLS. Since there are k alternatives the second stage yields k estimations for β_{11} and α_1.

In general, the two-stage method avoids a complicated likelihood function but suffers from the possibility of multicollinearity in the second stage. Also, the estimates obtained by the two-stage method are inefficient since the data on quantity are not used in the first stage. In addition the derivation of the corrected standard error is complicated. On the other hand, the computational burden of the MLE method is significantly reduced by available computer hardware and software. We found that using the estimates from the two-stage method as the initial values for the MLE method reduces the number of iterations slightly.

---

[7] The Linear Expenditure System can also yield similar results (Chiang (1988)).

5. Empirical Application and Comparisons

We apply the model described in section 3.2 to data on the demand for coffee. Then, for purpose of comparison, we also apply the model with zeros excluded to the same data. Both are estimated by the MLE method.

5.1 Data

IRI scanner panel data for coffee purchase is used for estimation and validation. The analysis is restricted to data for ground caffeinated coffee in one market (Pittsfield, MA) for the first 20 weeks of the record. The data is from 253 households, with at least a female adult, who purchase regular coffee in this 20-week period at any one of six different stores at least one time.

Group I is a set of 130 panelists who are randomly selected, and group II is comprised of the remaining 123 panelists. Estimation of the model uses only the data from the first 10 weeks for group I. The last 10 weeks of group I data and the first 10 week of group II data are reserved as the holdout samples for model testing.

Regular coffee is the focus of this application, and so purchases of nonregular coffee alone are regarded as zeros.[8] Hence, among the 612 observations involved in the estimation, there are 187 zeros and 425 consumption events.[9]

The data does not include the total expenditure, including noncoffee products, of a panelist at each purchase occasion. Therefore, the total expenditure is estimated.[10]

[8] The model, in principle, needs data from each shopping trip including those which do not involve any kind of coffee purchase. Unfortunately, IRI coffee data does not record those trips. Therefore, the author recognizes that the number of zeros are under-reported.

[9] Obervations with multi-brand purchases are about 1.8% of the total sample. These observations can be eliminated with a higher degree of brand aggregation. We discard these as exceptional cases.

[10] The estimation of weekly expenditure is based on Ching-Fan Chung (1987). Chung's model is adopted for its intensive use of demographic variables, which is compatible with the information provided by IRI data. The standard errors caused by this expenditure estimation can be easily

5.2 Brands and Variables

The three major brands of regular coffee are denoted A, B and C. D represents an aggregate of five minor brands, while E is an aggregate of all generic brands.[11]

The explanatory variables are now defined as follows:

1) Variables included in w_j' are:

$$w_j^1 = \begin{cases} 1 & \text{for brand } j, \; j = A,B,C,D \\ 0 & \text{otherwise} \end{cases}$$

In the subsequent tables, w_j^1 is denoted Dummy j. The brand-specific dummy is suppressed for Brand E as a standard normalization.

Two marketing promotional variables which record whether a brand was featured in a store flier/ad (Feature) or on special store display (Display) are included for their potential influence on consumer decision-making. The corresponding coefficients are assumed the same across brands.

$$w_j^2(t) = \begin{cases} 1 & \text{if brand } j \text{ was on feature at time of purchase} \\ 0 & \text{otherwise} \end{cases}$$

$$w_j^3(t) = \begin{cases} 1 & \text{if brand } j \text{ was on display at time of purchase} \\ 0 & \text{otherwise} \end{cases}$$

2) Variables included in w_{k+1}' are:

Several demographic variables are included to reflect consumer heterogeneity in scaling

incorporated into the model.

[11] Each brand may have various sizes and unit prices. The weekly representative price in dollar per 16 ounce of each brand is calculated with the conventional weight scheme over these brand/size variations.

the composite good. These variables are

$w^1_{k+1} =$ the number of family members (Family Size).

$w^2_{k+1} =$ 1 if the household income is greater than \$25k, 0 otherwise (Income).

$w^3_{k+1} =$ 1 if the female adult has a full-time job, 0 otherwise (Female Working Status).

$w^4_{k+1} =$ 1 if the age of the female adult is over 35, 0 otherwise (Female Age).

$w^5_{k+1} =$ 1 if the female adult is at least high school educated, 0 otherwise (Female Education).

It is presumed that the female adult shops for each family and hence we restrict our attention to their characteristics.

Since each panelist may or may not appear each week, the data is unbalanced. We assume shopping frequency is determined independently from the decision framework in our model. Therefore, let the dummy variable S_{it} equal one if consumer i shops during week t, and equal zero otherwise. Let P_{si} denote the probability of $S_{it}=1$ and I_{jit} denote the event indicator for which $I_{jit}=1$ if alternative j is chosen and $I_{jit}=0$ otherwise. Then, the sample likelihood function of each observation, which is indexed by i and t, is

$$(5.1) \qquad L_{it}(\pi) = \left[\prod_{j\in M}\left(L_{it}(D=j,I=1\,|\,\pi,S_{it}=1)\right)^{I_{jit}}\left(L_{it}(I=0\,|\,\pi,S_{it}=1)\right)^{1-I_{jit}}\cdot P_{si}\right]^{S_{it}}\cdot(1-P_{si})^{1-S_{it}}$$

where $L(D=j,I=1\,|\,\pi,S_{it}=1)$ and $L(I=0\,|\,\pi,S_{it}=1)$ are, respectively, represented by (2.12) and (2.10) and π denotes the parameter.

Summing, across both individuals and time, the logarithm of equation (5.1), it is clear that the portion of the likelihood function involving π is identical to the likelihood function given in (2.13). However, because the serial correlation of choices made by each panelist at different time periods is not specified, this likelihood function is a pseudo-likelihood function.[12] Hence, the variance-covariance matrix of $\hat{\pi}$ is adjusted accordingly (see White (1982)).

5.3 Coefficients and Relevant Statistics

Table 1 provides the likelihood estimates using the ITL utility function with and without the zero-consumption option. The two cases are labeled ITL_1 and ITL_2 respectively. Assuming that the $\epsilon=(\epsilon_1,...,\epsilon_{k+1})$ have the GEV distribution with $G(t)=[\Sigma_{j=1}^{k} t_j/(1-\delta)]^{1-\delta}+t_{k+1}$, both the discrete choice and zero-consumption probabilities have a multinomial nested logit structure.

Though the signs of coefficients are the same for the two sets of estimates, the magnitudes are dissimilar. ITL_1 has larger coefficient estimates for Feature and Display, two major variables in the brand-specific index, but has, overall, smaller absolute values of the estimates for the variables in the composite good index. The smaller log-likelihood value of ITL_1 is due to the negative log-density values contributed by zero-consumption observations.

We found that both types of promotion have strong positive effects on brand choice and consequently on demand. This implies that, for a given market price, promotions effectively enhance perceived quality and increase the probability of selection. By equation (3.16) the expected demand also increases.

The number of family members has a significant positive coefficient. Thus, as family size increases then, as shown in (3.13) and (3.16) respectively, the likelihood of a regular coffee purchase and the expected demand decrease. This may reflect the greater importance of the composite good for large families having the same budget as smaller families.

[12] It is plausible that an error components model (Balestra and Nerlove (1966)) can deal with this problem, but this method complicates the derivations of choice probabilities and increases the computational burden. The method of simulated moments proposed by McFadden (1989) maybe an alternative method for the estimation.

Other things being equal, higher income households have less tendency to purchase regular coffee and have lesser expected demand when purchases do occur. This may reflect the substitution by these households to premium grade coffee.

All characteristics of the female adult are shown to be without significance. Thus, coffee drinking habits of a family appear to be separate from that of their shopper, possibly reflecting the fact that the collective family traits are more relevant.

The coefficient δ is interpreted as a measurement of the similarities among brands. For both ITL_1 and ITL_2 data, δ is significantly different from zero. This implies that the data does not support a multinomial logit structure. The Lagrange Multiplier tests for the homogeneity restrictions in the utility function are not rejected at the significance level of 5% in either ITL_1 or ITL_2 (χ^2-3.24 and χ^2-2.45 respectively with d.f-2).

Table 2 presents estimates of price elasticities for the unconditional expected demand equations. For ITL_1, the total elasticity is decomposed into three portions: changes in the expected quantity conditional on both brand selection and purchase occurrence (η_{pq}), changes in the brand choice probability conditional on purchase occurrence (η_{pb}), and changes in the probability of purchase (η_{pp}). These elasticities are calculated at the point which represents average family size, zeros for all demographic variables, an absence of promotions, and average market prices.[13] Since the estimates indicate that $|\eta_{pb}| > |\eta_{pq}| > |\eta_{pp}|$, an average consumer switches brands instead of consuming less or forgoing regular coffee when the price increases. For ITL_2, the total elasticity is only decomposable into two parts, η_{pq} and η_{pb}. Both elasticities as well as the total elasticity are larger for ITL_2. Thus, in spite of the results of Section 3.1, which show that the GEV distribution yields functional forms for the conditional discrete choice

[13] There are minute differences in the calculated results when demographic variables are individually set equal to one.

probabilities which are identical to that for the discrete choice probabilities without zero-consumption, the calculated elasticities differ. This reflects the misspecification caused by biased sample selection and its effect on the parameter estimates.

Tables 3 and 4, respectively, show the effects of feature and display promotions. Both ITL_1 and ITL_2 show that the most pronounced effect of promotions is on the brand choice probabilities. This suggests that both promotional means can effectively increase market share by attracting consumers away from competing brands. Also, the promotional effects on brand choice and quantity are stronger for ITL_1. This, again, is caused by the biased sample selection.

5.4 Testing and Comparison

Discrete choice predictions for the holdout sample are given in Table 5, where the outcome with highest calculated probability is the predicted choice (Domencich and McFadden (1977)). The model correctly predicts approximately 50% of the choices. In contrast, a naive approach which tabulates the frequencies of outcomes across individuals, week by week, and uses these as predicted probabilities for every individual, correctly predicts only about 26.5%. Excluding zeros, the predictions of ITL_1 and ITL_2, though not shown here, are almost identical, with 60% and 62% correct respectively.

Figure 1 shows the predicted and actual demand quantities in the holdout sample for ITL_1. The one-step ahead total quantity prediction for brand j is obtained by multiplying the conditional expected demand for brand j, equation (3.16), by its choice probability, equation (3.12), for each individual and then summing across individuals. The standard error of the quantity prediction for each brand is derived as follows. Let π denote the underlying parameters of the model and let $\tilde{\pi}$ denote the estimates of π. Let $q_{ji}(\tilde{\pi})$ denote the expected quantity given in (3.16) multiplied by y/p_j so that the actual demand for brand j by consumer i is $q_{ji}(\pi)+v_{ji}(\pi)$, where $v_{ji}(\pi)=y\beta_{11}u_{ji}(\pi)/p_j$. $P_{ji}(\tilde{\pi})$ represents the choice probability of brand j by individual i, as defined in (3.12), and I_{ji} indicates the action taken by consumer i with respect to brand j. The estimated and actual total quantity, Q_j and Q_j^*, can be expressed by $Q_j=\sum_i q_{ji}(\tilde{\pi})P_{ji}(\tilde{\pi})$ and $Q_j^*=\sum_i [q_{ji}(\pi)+v_{ji}(\pi)]I_{ji}$. Therefore,

$$(5.2) \qquad Q_j-Q_j^*=\left[\sum_{i=1}^{N} q_{ji}(\tilde{\pi})P_{ji}(\tilde{\pi})-\sum_{i=1}^{N} q_{ji}(\pi)P_{ji}(\pi)\right]+\sum_{i=1}^{N} q_{ji}(\pi)P_{ji}(\pi)-\sum_{i=1}^{N} [q_{ji}(\pi)+v_{ji}(\pi)]\cdot I_{ji}$$

$$=\left[\frac{\partial\sum_{i=1}^{N} q_{ji}(\tilde{\pi})P_{ji}(\tilde{\pi})}{\partial\pi}\right]\cdot(\tilde{\pi}-\pi)+\sum_{i=1}^{N} q_{ji}(\pi)[P_{ji}(\pi)-I_{ji}]-\sum_{i=1}^{N} v_{ji}(\pi)I_{ji}$$

By (5.2), the asymptotic variance is:

$$(5.3) \qquad \mathrm{Asy.Var}(Q_j)=T\Delta_\pi T'+\sum_{i=1}^{N} q_{ji}(\pi)\Delta_j q_{ji}(\pi)'+\sum_{i=1}^{N} \mathrm{var}(v_{ji}(\pi))I_{ji}$$

where $T=\Sigma_{i=1}^{N}\,\partial q_{ji}(\tilde{\pi})P_{ji}(\tilde{\pi})/\partial\pi$, Δ_π is the asymptotic variance of $(\tilde{\pi}-\pi)$, Δ_j is the variance of the Bernoulli variable $(P_{ji}(\pi)-I_{ji})$, and $\mathrm{Var}(v_{ji}(\pi)=(\beta_{11}y/p_j)^2\mathrm{var}(u_{ji}(\pi))$. The variance of $u_{ji}(\pi)$ can be derived using theorem B.2.2 of Dubin (1985). By substituting $\tilde{\pi}$ into (5.3) the consistent estimate of the asymptotic variance of Q_j can be obtained.

As shown in Figure 1, the predictions are reasonably accurate at 90% confidence intervals. The model also follows market turns for demand especially well. Figure 2 shows the predictions of ITL_2 against the actual data. Comparing these figures, model performances are very similar for each brand. The aggregate prediction, which includes all brands, has mean sum of squared errors (MSSE) equal to 103.44 and 130.07 for ITL_1 and ITL_2 respectively, and so the ITL_1 model is more accurate but has greater variances.

Figure 1 Quantity Prediction (ITL_1)

-- Mean Prediction; — Actual Quantity; ··· 90% Confidence Intervals

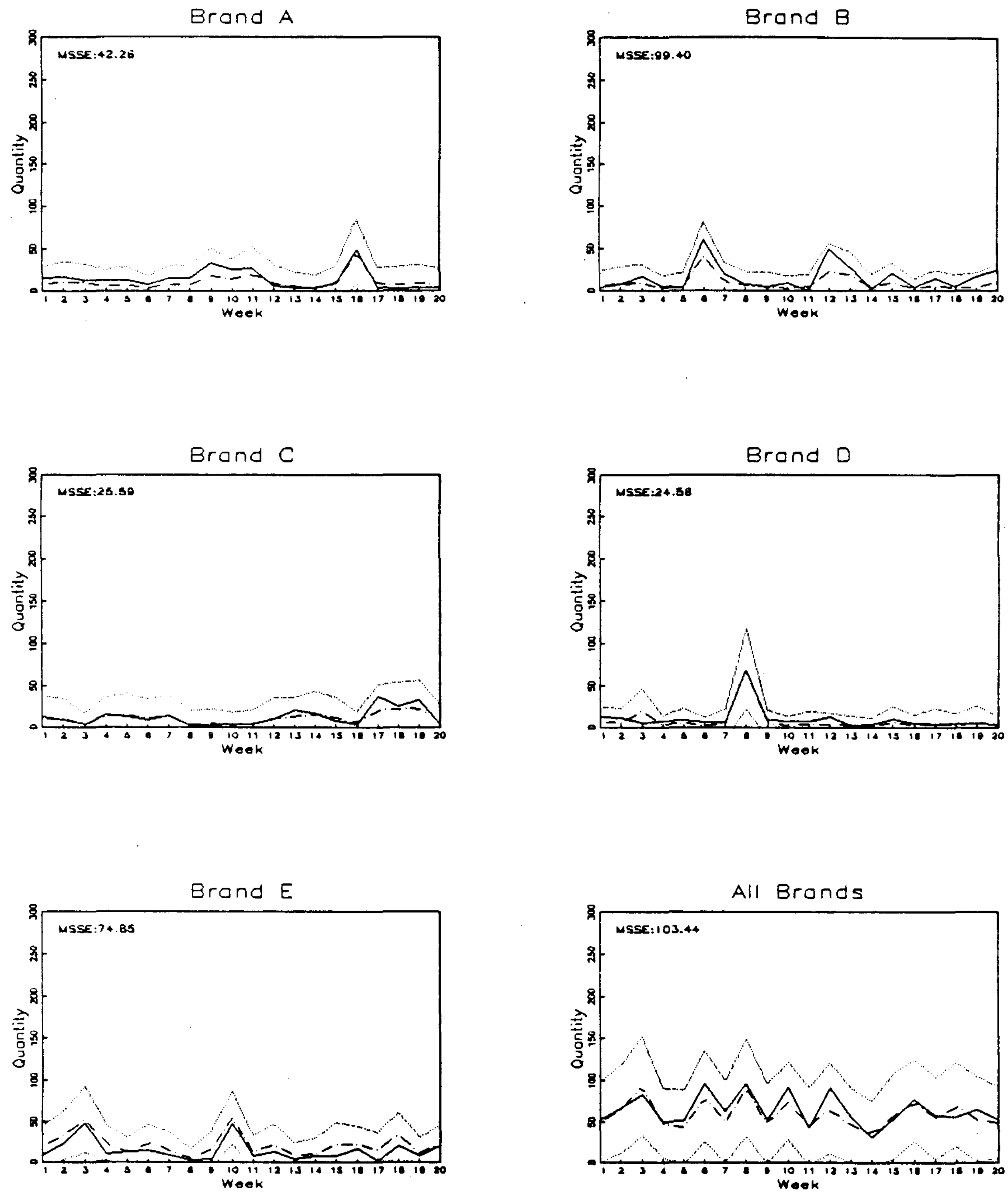


Figure 2 Quantity Prediction (ITL_2)

-- Mean Prediction; — Actual Quantity; ··· 90% Confidence Intervals

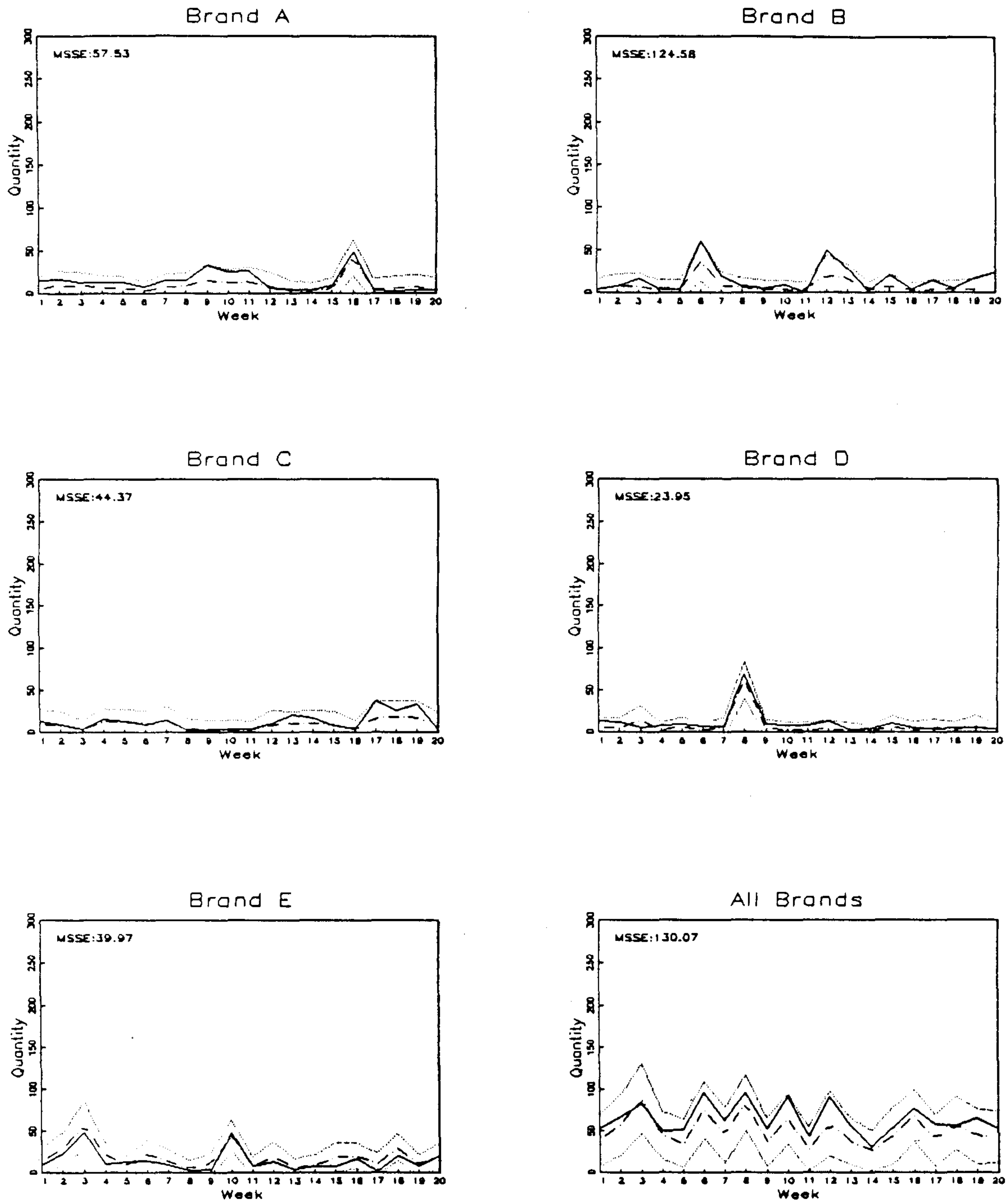
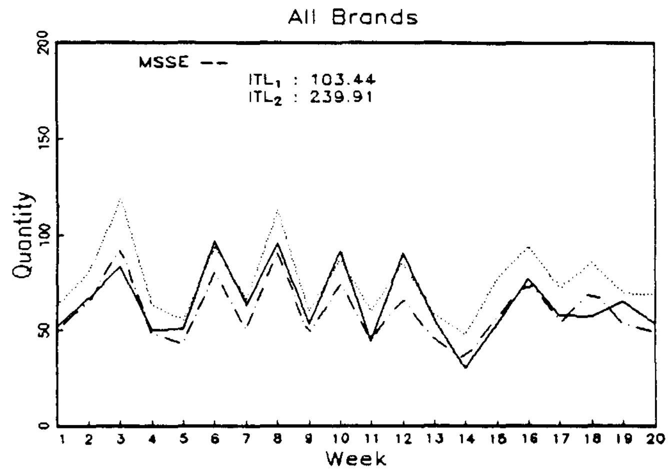


Figure 3 Purchase Quantity (Market Level)

—— Observed Quantity
-·- ITL_1 Mean Prediction
···· ITL_2 Mean Prediction (With the Complete Holdout Sample)



REFERENCES

Balestra, P., and M. Nerlove (1966), "Pooling Cross-Section and Time Series Data in the Estimation of A Dynamic Model: The Demand For Natural Gas," Econometrica, 34: 586-612.

Barten, A. P. (1964), "Family Composition, Price and Expenditure Patterns," in P. E. Hart, G. Mill, and J. K. Whitaker (ed.), Econometric Analysis for National Economic Planning, London: Butterworth.

Chiang, J. (1988), "Discrete/Continuous Choice and Purchase Decision Econometric Models for Consumer Demand," unpublished dissertation, Department of Economics, University of Minnesota.

Chung, C. F. (1987), "Demand Systems with Demographic Effects: Theory, Estimation and the Testing," unpublished dissertation, Department of Economics, University of Wisconsin, Madison.

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

Deaton, A., and T. Muellbauer (1980), Economics and Consumer Behavior. Cambridge: Cambridge University Press.

Domencich, T. A., and D. McFadden (1975), Urban Travel Demand. Amsterdam: North-Holland.

Dubin, J. A., and D. McFadden (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," Econometrica, 52: 345-362.

Dubin, J. A. (1985), Consumer Durable Choice and the Demand for Electricity, Amsterdam: North-Holland.

Hanemann, W. M. (1984), "Discrete/Continuous Models of Consumer Demand," Econometrica, 52: 541-561.

King, M. A. (1980), "An Econometric Model of Tenure Choice and Demand for Housing as a Joint Decision," Journal of Public Economics, 14: 137-159.

Lau, L. J., W. L. Lin and P. A. Yotopoulos (1978), "The Linear Logarithmic Expenditure System: An Application to Consumption-Leisure Choice," Econometrica, 46: 843-868.

Lee, L. F. and M. M. Pitt (1986), "Microeconometric Models of Consumer and Production Demand With Binding Non-Negativity Constraints: The Dual Approach," Econometrica, 54:1237-1242.

Lee, L. F. and M. M. Pitt (1987), "Microeconometric Models of Rationing, Imperfect Markets, and Non-negativity Constraints," Journal of Econometrics, 36: 89-110.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), Frontier in Econometrics, New York: Academic.

McFadden, D. (1978), "Modelling the Choice of Residential Location," in A. Karlquist et al. (ed.), Spatial Interaction Theory and Residential Location, Amsterdam: North-Holland.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," Econometrica, 57, 995-1026.

Muellbauer, J. (1974), "Household Composition, Engel Curves and Welfare Comparisons Between Household: A Duality Approach," European Economic Review, 10: 103-122.

Neary, J. P. and K. W. S. Roberts (1980), "The Theory of Household Behavior Under Rationing," European Economic Review, 13: 25-42.

Rothbarth, E. (1941), "The Measurement of Changes in Real Income Under Conditions of Rationing," Review of Economic Studies, 8: 100-107.

VanSoest, A. and P. Kooreman (1990), "Coherency of Indirect Translog Demand System with Binding Non-Negativity Constraints," Journal of Econometrics, 44, 391-400.

Wales, T. J. and A. D. Woodland (1983), "Estimation of Consumer Demand Systems With Binding Non-Negativity Constraints," Journal of Econometrics 21: 263-285.

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