There Is No Aggregation Bias: Why Macro Logit Models Work
Author(s): Greg M. Allenby and Peter E. Rossi
Source: *Journal of Business & Economic Statistics*, Vol. 9, No. 1 (Jan., 1991), pp. 1-14
Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association
Stable URL: http://www.jstor.org/stable/1391935

Journal of Business & Economic Statistics, January 1991, Vol. 9, No. 1

There Is No Aggregation Bias: Why Macro Logit Models Work

Greg M. Allenby
College of Business, Ohio State University, Columbus, OH 43210

Peter E. Rossi
Graduate School of Business, University of Chicago, Chicago, IL 60637

In this article, we examine the aggregation properties of (nested) logit models to understand their exceptional macro-level performance. The problem of aggregating micro logit models involves integrating nonlinear functions of model parameters over a distribution of consumer heterogeneity. The aggregation problem is analyzed using a mixture of analytic and simulation techniques, with the simulation experiments using actual panel data to calibrate the distribution of heterogeneity. We conclude that the practice of fitting aggregate logit models is theoretically justified under the following three conditions: (1) All consumers are exposed to the same marketing-mix variables, (2) the brands are close substitutes, and (3) the distribution of prices is not concentrated at an extreme value. These conditions are frequently met in store-level scanner data.

KEY WORDS: Elasticity; Heterogeneity; Scanner data.

Aggregate-market-share models are used extensively in marketing. Retailers use aggregate-market-share data to determine price and promotional strategies. Consider, for instance, the problem of pricing private-label brands in a grocery-store chain. In many product classes, the national and private labels are close substitutes with similar attributes and price. Aggregate-market-share models calibrated with store scanner data are used to estimate the response of the private-label market share to changes in its price and the prices of national brands. It is critical to retailers that the predicted responses to various pricing strategies do not suffer from aggregation bias.

Aggregate-market-share models usually take on one of three functional forms—(1) the linear form (market share = $\mathbf{x}'\boldsymbol{\beta} + \varepsilon$), (2) the multiplicative form (ln market share = $\ln \mathbf{x}'\boldsymbol{\beta} + \varepsilon$), or (3) the logit or market-attraction form (market share = $\exp[\mathbf{x}_i'\boldsymbol{\beta}]/\Sigma_j \exp[\mathbf{x}_j'\boldsymbol{\beta}]$). Surprisingly, studies have shown (see Leeflang and Reuyl 1984) that the three models perform similarly in practice, despite the severe restrictions that the independence of irrelevant alternatives (IIA) property of the logit model imposes on the data. In addition, practitioners find that the logit model, which was developed for micro-level data, performs well with aggregate-market-share data.

In this article, we examine the aggregation properties of logit models to understand their exceptional macro-level performance. A nested logit model that includes the standard logit model as a special case (see McFadden 1978, 1981, 1984) is studied. It is shown that when heterogeneous consumers are all exposed to the same marketing variables, the resulting aggregate market shares exhibit a structure nearly identical to one obtained from a logit model that assumes the existence of a representative individual. In other words, there is almost no aggregation bias.

To study aggregation bias, a model of consumer heterogeneity must be selected. Traditionally, heterogeneity is modeled by the inclusion of demographic variables in the micro model to control for differences in taste (e.g., see Maddala 1983). Practitioners usually find this approach to be inadequate, however, and attempt to remedy this by including variables that measure past buying behavior (Guadagni and Little 1983; Krishnamurthi and Raj 1988). In this article, we employ a random-coefficients model in which a distribution of heterogeneity is assumed (Heckman 1982) in the analysis of the aggregation problem. We find the random-coefficients approach to be conceptually appealing and very flexible in the modeling of consumer heterogeneity.

The problem of aggregation across heterogeneous consumers has been studied extensively in the economic literature. One approach taken by economists is to assume that the fundamental microeconomic relation is linear (cf. Muellbauer 1975; Theil 1954). The nested logit model, however, is a fundamentally nonlinear relation between a discrete dependent variable and the characteristics of households. Stoker (1982) provided a characterization of when the micro function and the distribution of consumer heterogeneity form what he terms a "complete aggregation structure" in which there is a unique correspondence between the macro and micro functions. Since it is difficult to observe and characterize the distribution of consumer heterogeneity, it may be difficult in practice to verify that a complete aggregation structure is in place. Moreover, there may exist situations in which there is an aggregation problem but the magnitude of the problem is small.

This article attacks the aggregation problem in a different manner. The micro model is linearized (with IIA restrictions imposed) and then aggregated over a heterogeneous group of consumers. The nonlinear nature of the cross-elasticity expressions complicates the aggregation of the micro linear model. The results of the analysis indicate that for retail scanning data, in which all consumers are exposed to the same prices, advertisements, and in-store displays, a restricted linear model is capable of accurately representing the true aggregated model. Since the restrictions can be defined in terms of readily observed aggregate statistics, we conclude that knowledge of the distribution of heterogeneity is not required to use either aggregate logit models or their restricted linear approximation with aggregate retail scanning data.

The article is organized as follows: In Section 1 we investigate the degree to which a micro linear model, subject to the restrictions implied by the (nested) logit model, is able to represent the true choice probabilities. In Section 2, we aggregate the micro linear model by integrating over the distribution of consumer heterogeneity. Cross-elasticity estimates computed from the aggregated linear model closely agree with estimates computed using an approximation involving average market shares. The approximations in Sections 1 and 2 are then studied jointly in Section 3, where a simulation experiment is used to show that the macro linear model inherits the restrictions implied by the micro logit model and is capable of accurately representing data arising from a heterogeneous process. Section 4 contains some concluding remarks.

1. LINEARIZATION OF THE NESTED LOGIT MODEL

A nested logit model is postulated for individual consumer choice behavior. Proposed by McFadden (1984), the nested logit model is a flexible choice model that is used in situations in which the goods can be grouped into local clusters or submarkets of similar products. The nested logit model avoids applying the IIA property of the standard conditional logit model to all goods. In this section, we investigate the extent to which the nested logit model can be approximated by a restricted linear model.

The nested logit model is derived from a random-utility model with errors distributed according to the generalized extreme-value distribution. The choice probabilities can be decomposed into the probability of choice given that a particular submarket has been chosen times the probability of selecting that submarket:

$$\Pr(i \mid A) = \frac{\exp[V_i/\lambda_s]}{\sum_{j\in A_{s'}} \exp[V_j/\lambda_s]} \times \frac{\left[\sum_{j\in A_s} \exp[V_j/\lambda_s]\right]^{\lambda_s}}{\sum_{s=1}^{S} \left[\sum_{j\in A_s} \exp[V_j\lambda_s]\right]^{\lambda_s}}, \quad (1)$$

where i is an element of submarket A_s, $V_i = \beta_{0i} - \beta_p \ln p_i$ is the price-adjusted nonstochastic level of utility, and λ_s is a measure of the similarity of the brands within submarket A_s and varies between 0 and 1. Since the second term is equal to $\sum \Pr(i \mid A)$ $(i \in A_s)$ the equation provides a nested (logit) decomposition of the purchase probability: $\Pr(i \mid A) = \Pr(i \mid A_s) \cdot \Pr(A_s \mid A)$.

Cross-elasticities of the nested logit model, of which the logit model is a special case, can be shown to be of the form

$$\frac{\partial \ln \Pr(i)}{\partial \ln p_j} = \beta_p \Pr(j) \left[\frac{\delta_{ij}}{\lambda_s \Pr(j)} - 1 - \frac{\delta_{ss'}(1 - \lambda_s)}{\lambda_s \Pr(A_s)}\right], \quad (2)$$

where $\Pr(i)$ is the choice probability of the ith good, p_j is the price of the jth good, β_p is the log(price) coefficient and can be interpreted as a measure of price sensitivity, $\Pr(A_s)$ is the probability of choosing any one of the brands in submarket A_s [i.e., $\sum_{j\in A(s)} \Pr(j)$], i is assumed to be a member of submarket A_s, j is a member of submarket $A_{s'}$, and δ_{ab} is the Kronecker delta in which $\delta_{ab} = 1$ if $a = b$ and 0 otherwise (see McFadden 1984).

The first term inside the parentheses is nonzero when $i = j$, corresponding to a self-elasticity. The second term is what is normally found in the traditional logit model, and the third term alters the expression when i and j are members of the same submarket. Equation (1) reverts back to that of the traditional logit model when all lambdas are equal to 1. Since i only enters the right side of (1) through the term $\delta_{ss'}$, a 1% change in the price of alternative j induces an equal percentage of change in $\Pr(i)$ for any i in the same submarket as j and an equal but smaller percentage of change in $\Pr(i)$ for any i in a different submarket as j. This is often referred to as the *proportional-draw* property. In addition, when i and j are members of the same submarket, the term in the brackets is the same, implying that $\eta_{ij}/\eta_{ik} = \Pr(j)/\Pr(k)$. This restriction will be referred to as the *proportional-influence* property.

1.1 A Taylor Series Approximation

The linearization of the model takes place at the level of the consumer. Since the attributes of the products and the consumers do not change in the short run, only price and other promotional variables are assumed to change over time. For ease of exposition, it will be assumed that only price changes. These results can be extended for the other promotional variables such as advertising.

Expanding the log choice probability as a Taylor series about the average value of log-price yields for the kth individual

$$\begin{aligned}\ln \Pr(i \mid \ln \mathbf{p}_t, \boldsymbol{\beta}_k) &= \ln \Pr(i \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta}_k) \\ &+ (\ln \mathbf{p}_t - \ln \bar{\mathbf{p}})' \left.\frac{\partial \ln \Pr(i \mid \ln \mathbf{p}_t, \boldsymbol{\beta}_k)}{\partial \ln \mathbf{p}_t}\right|_{\ln \bar{\mathbf{p}}} \\ &+ \tfrac{1}{2}(\ln \mathbf{p}_t - \ln \bar{\mathbf{p}})' \left.\frac{\partial^2 \ln \Pr(i \mid \ln \mathbf{p}, \boldsymbol{\beta}_k)}{\partial \ln \mathbf{p}_t \partial \ln \mathbf{p}_t'}\right|_{\ln \mathbf{p}^*} \\ &\times (\ln \mathbf{p}_t - \ln \bar{\mathbf{p}}), \qquad (3)\end{aligned}$$

where $\ln \mathbf{p}^* \in (\ln \mathbf{p}_t, \ln \bar{\mathbf{p}})$ and $\boldsymbol{\beta}_k' = (\boldsymbol{\beta}_{0k}, \beta_{pk}, \boldsymbol{\lambda})$, where $\boldsymbol{\beta}_{0k}$ is a vector of utilities or preferences for the brands under study, and the other terms are as defined previously. Combining terms that are constant over time and regarding the remainder term as an error term results in the usual regression model in which the log of purchase probability regressed on log of prices is obtained.

The linearization of the nested logit choice probabilities provides a system of m linear equations (one for each brand choice). The elasticity expressions in (2) provide the basis for restricting the coefficients of the approximate linear-regression system. It must be verified that the restricted linear system is capable of adequately representing the choice probabilities. The way we measure the accuracy of the approximation depends on the ultimate use of the choice model. Since retailers routinely change prices by more than 30%, both the global and local properties of the linear approximation should be examined.

It should be pointed out that this situation is somewhat different from the standard approximation problems studied in the econometric literature. The only source of error in the linear model is approximation error; there is no random-error term in the underlying nonlinear relationship as in the work of White (1980), who provided sufficient conditions under which the least squares estimates from the linear model will consistently estimate the derivatives of the true micro model. These sufficient conditions are very restrictive and are not satisfied in the case of the nested logit micro model [see White (1980, theorem 1, p. 153) for a discussion of sufficient conditions].

In our case, there is a direct relationship between the goodness of fit of the least squares approximation as measured by mean squared error and the accuracy of approximation of the derivatives. For the two-brand case, it is possible to show that a given level of goodness of fit implies an upper bound on the discrepancy between the slope of the least squares approximation and the derivatives of micro function. This result relies heavily on the special structure of this situation: (a) The only source of error is the approximation error; (b) the log-probability locus is concave and differentiable, ensuring that there are no inflection points; (c) the price distribution has bounded support; and (d) the derivatives over the support of the price distribution are bounded.

The simulation experiment examines both the global and local aspects of the linear approximation. Global approximation error is measured by using the standard error of the linear-model residual. If the log-probability locus is linear over the range of prices, then the errors in the linear model will be small. In addition, comparisons between theoretical elasticities and least squares estimates are made to verify that the local properties of the least squares approximation are adequate and to corroborate our theoretical results. The error term in the linear model is a quadratic form in log prices and the second derivative matrix of the log probability locus. Thus we might expect that the error term is heteroscedastic and correlated with the included log-price regressors. The least squares residuals are inspected for heteroscedasticity and deviations from normality. In addition, the asymptotic bias in the least squares coefficients due to correlation between the errors and included variables is estimated.

1.2 The Simulation Experiment

The range of the dependent variable is a function of two factors, (a) the fluctuations in the price series $\ln \mathbf{p}_t$ and (b) the magnitude of the elasticities. Referring to Equation (2), it is seen that the cross-elasticity expression is in turn a function of β_p, the price sensitivity coefficient; $\boldsymbol{\beta}_0$, the vector of preferences; and the number of brands in the choice set. The design of the simulation experiment must therefore consider the interaction of the factors $\ln \mathbf{p}_t$, β_p, $\boldsymbol{\beta}_0$, and the number of brands.

Actual data on the prices of eight brands of stick margarine were used to form the pattern of prices used in the simulations. These data were obtained from a major supermarket chain in the Chicago area over 50 weeks. Five of the brands are nationally distributed, two are private label, and the last is a generic brand. Since the error term in Equation (3) is a function of the total variability in the dependent variable, the magnitude of the price changes are not, in themselves, of interest. The reported values of β_p induce a widely varying range in the dependent variable. The generality of the results of this experiment are therefore a function of the pattern and not the level of the independent variables.

The experiment was designed with six different "typical" consumers in mind. The number of brands in the choice set, as well as the level of the choice probabilities, were varied across the cells in the experiment shown in Figure 1. The upper left cell assumes that the

Number of Brands	Probabilities: Equal	Unequal	Dominated
8 (nested logit)	$\Pr(i) \approx .12$	$\Pr(1) \approx .30$ $\Pr(j) \approx .10$ $(j \neq 1)$	$\Pr(1) \approx .50$ $\Pr(j) \approx .07$ $(j \neq 1)$
2 (logit)	$\Pr(i) \approx .50$	$\Pr(1) \approx .70$ $\Pr(2) \approx .30$	$\Pr(1) \approx .90$ $\Pr(2) \approx .10$

Figure 1. Experimental Design.

consumer is actively evaluating all eight brands in a nested logit model and that the probabilities of the eight brands are approximately equal. The lower left cell assumes that the consumer is evaluating the first two national brands, which are assumed to be in the same submarket, resulting in a standard logit model. The remaining cells in the figure have one brand with a higher probability than the others. The cells with one dominant brand are designed to represent the situation in which there is considerable consumer loyalty to one brand, whereas the cells with equal choice probabilities represent the case of nonloyal brand-switching consumers.

The price-sensitivity coefficient, β_p, was systematically varied in each of the experimental cells, taking on values of -1.0, -2.0, and -4.0. Very large values of the β_p coefficient induce high variability in purchase probabilities, which results in larger approximation error than for smaller values. The data were generated assuming that the five national brands compose a submarket, the two private label brands make up a second submarket, and the generic brand is in its own submarket. The similarity parameter λ was set to .7 for the national brands, .5 for the private label brands, and 1.0 for the generic brand.

The simulation experiment proceeded as follows:

1. For a given cell in the design table (see Fig. 1), one of the three values of the price-sensitivity coefficient was chosen.
2. The intercept parameters were adjusted to produce the approximate choice probabilities indicated in the design table.
3. Choice probabilities were computed using a nested logit model specification (the logit model is a special case of this) for each of 50 different price vectors.

A multivariate linear model was used to fit the data subject to the constraints on the cross-elasticities implied by Equation (2),

$$\ln \Pr(i)_t = \alpha_i + \boldsymbol{\eta}_i' \ln \mathbf{p}_t + \varepsilon_{it}, \qquad i = 1, \ldots, 7, \quad (4)$$

which was subject to the IIA restrictions on the collection of η_i vectors. The average values of the choice probability over the 50 time periods were used in the within-equation proportional-influence restriction. One equation was arbitrarily dropped from the analysis to avoid logical consistency problems, since the simulated probabilities sum to 1. For the nested logit model the generic brand was eliminated, and for the logit model the second brand was dropped from the analysis.

The simulation experiment was also used to investigate the ability of the linear model to represent choice probabilities (as opposed to log probability). The re-

Table 1. Simulation Results

Model	Price coefficient (β_p)	Average range[a] (ln prob.)	Gain hi/low[b] (prob.)	Residual standard error[c]: ln prob.	prob.
1. Nested logit (n = 8) equal prob.	−1.0	.77	2.16	.0071	.0020
	−2.0	1.53	4.62	.0292	.0082
	−4.0	3.06	21.32	.1193	.0304
2. Nested logit (n = 8) unequal prob.	−1.0	.79	2.20	.0073	.0021
	−2.0	1.57	4.81	.0299	.0078
	−4.0	3.08	21.76	.1244	.0294
3. Nested logit (n = 8) dominated prob.	−1.0	.85	2.34	.0065	.0017
	−2.0	1.69	5.42	.0259	.0069
	−4.0	3.30	27.11	.1106	.0280
4. Logit (n = 2) equal prob.	−1.0	.47[d]	1.60	.0076	.0003
	−2.0	.96	2.61	.0298	.0023
	−4.0	1.96	7.10	.1103	.0137
5. Logit (n = 2) unequal prob.	−1.0	.26	1.30	.0061	.0027
	−2.0	.55	1.73	.0246	.0100
	−4.0	1.24	3.45	.0962	.0316
6. Logit (n = 2) dominated prob.	−1.0	.09	1.09	.0026	.0021
	−2.0	.20	1.22	.0111	.0086
	−4.0	.53	1.70	.0501	.0339

[a] Calculated as the average value of $F_i^{-1}(.95) - F_i^{-1}(.05)$ over the i brands used in estimation, where $F_i(\cdot)$ is the empirical distribution of ln Pr(i).
[b] Calculated as exp[average range].
[c] ln prob. uses log probability as the dependent variable in the model; prob. uses probability as the dependent variable.
[d] Calculated for the one brand used in estimation—that is, brand number 1.

strictions associated with the log-log model were modified accordingly, and, as before, the average choice probabilities over the 50 time periods were used to impose the restrictions.

1.3 Results of the Experiment

The results of the simulation experiment are contained in Table 1. Comparisons are made for values of the price-sensitivity parameter at -2.0 (this is the modal value from analysis of panel data in Sec. 2) and at -4.0, which represents an extremely high price sensitivity. Residual standard errors are reported for both the log-log and semi-log models. The 90% interpercentile range and the implied "gain" in the choice probabilities, defined as Pr(high)/Pr(low), are also reported. Although results for both the log-log model and the semi-log model are reported, only the former will be discussed in detail.

The results indicate that the restricted log-log model provides an excellent approximation to the true model. Even for very large ranges in choice probabilities, the residual standard error is very small. As indicated pre-

Table 2. Simulation Results: Elasticity Biases

	$\beta_p = -2.0$		$\beta_p = -4.0$	
Cross-elasticities[1]	*Theoretical*[2]	*Estimated*[3]	*Theoretical*	*Estimated*
Nested logit model				
Equal-probability case				
Brand 1 within group	.384	.406 (.012)	.729	.769 (.044)
Brand 1 across group	.219	.242 (.011)	.420	(.037)
Brand 6 within group	1.273	1.736 (.018)	2.581	4.186 (.069)
Brand 6 across group	.263	.390 (.020)	.542	1.054 (.089)
Brand 8 across group	.347	.363 (.023)	.624	.732 (.092)
Unequal-probability case				
Brand 1 within group	1.041	1.114 (.029)	1.848	2.037 (.109)
Brand 1 across group	.627	.683 (.026)	1.120	1.265 (.092)
Brand 6 within group	1.226	1.668 (.019)	2.502	4.089 (.071)
Brand 6 across group	.217	.338 (.021)	.463	.973 (.092)
Brand 8 across group	.284	.271 (.025)	.517	.489 (.097)
Dominated-probability case				
Brand 1 within group	1.692	1.774 (.023)	3.004	3.327 (.110)
Brand 1 across group	1.073	1.141 (.025)	1.898	2.145 (.094)
Brand 6 within group	1.169	1.582 (.016)	2.405	3.918 (.064)
Brand 6 across group	.159	.256 (.015)	.366	.849 (.077)
Brand 8 across group	.207	.171 (.018)	.391	.285 (.083)
Logit Model[4]				
Equal	1.008	1.054 (.025)	2.204	2.196 (.092)
Unequal	.561	.541 (.020)	1.207	1.343 (.080)
Dominated	.198	.213 (.009)	.479	.538 (.042)

NOTE: Standard errors are in parentheses.

[1] For the nested logit model, brands (1, 2, 3, 4, and 5) form group A, and brands 6 and 7 form Group B. Elasticities are as follows:
brand 1 within group: $-\beta\,\Pr(1)[1 + ((1 - .5)/(.5\,\Pr(A))]$
brand 1 across group: $-\beta^p \Pr(1)$
brand 6 within group: $-\beta^p \Pr(6)[1 + ((1 - .7)/(.7\,\Pr(B))]$
brand 6 across group: $-\beta^p \Pr(6)$
brand 8 across group: $-\beta_p^p \Pr(8)$

[2] The theoretical value of the elasticity is computed using the appropriate nested logit or simple logit formula and taking the expectation of this formula with respect to the empirical distribution of brand prices.

[3] The estimated value is the least squares coefficient estimated in a restricted regression model fitted to the computed probabilities.

[4] Elasticities are reported for the two-brand logit model using brands 1 and 2. Reported are the actual and estimated cross-elasticities of brand 1 with respect to the price of brand 2.

viously, it is expected that the approximation error in this model is positively related to the range of the choice probabilities. The simulation results show that data arising from a nested logit model with a gain of 5:1 is associated with a residual standard error of less than .03. This is true for the case in which all of the probabilities are equal, as well as when one of the probabilities is greater than the rest. A residual standard error of .03 in units of log probability can be interpreted as a 3% difference in the actual versus the fitted value of the probability. This is a reasonable level of accuracy for many problems faced in practice. A gain of 5:1 for the equal-probability model implies that the choice probabilities vary from approximately .07 to .35. Larger gains can be obtained at the expense of a poorer fit. For instance, a gain of 20:1 is associated with a residual standard error of .12.

Similar results are obtained for the logit model in which the two-choice situation is studied. For the equal-probability case, a residual standard error of .05 corresponds to a gain in the probabilities of about 4.0, or changes in the level from .20 to .80, and changes from

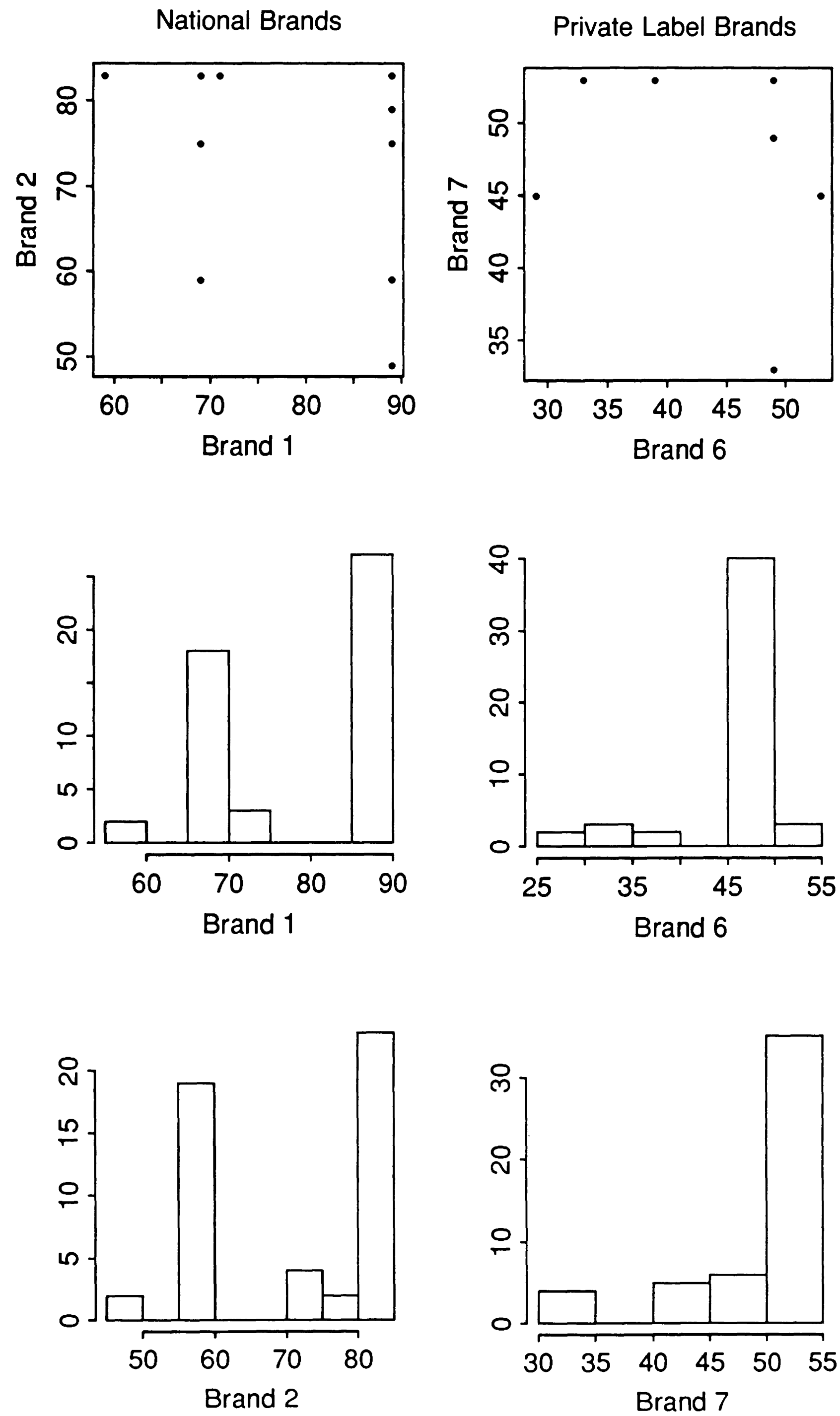

Figure 2. Distribution of Prices in Brands 6 and 7.

.50 to .99 when the average probability is .90. In other words, the log-log model adequately represents the situation in which on average one good is preferred to the other by a factor of 9:1 and price changes reduce this factor to 1:1.

In Table 2, theoretical elasticities are compared to the linear least squares estimates. The nested logit elasticity expressions in Equation (2) are functions of the price level. The expected value of the theoretical elasticity averaged over the empirical distribution of prices is compared to the constant-elasticity estimates from the linear approximation. The linear least squares estimates closely agree with the theoretical values in all cases except for the cross-elasticity estimates involving the house brands 6 and 7.

Brands 6 and 7 have an unusual distribution of prices, as is shown in Figure 2. Both brands rarely go on sale and remain mostly at one "regular" price. This contrasts dramatically with the national brands represented in the figure by brands 1 and 2. The national brands have a more uniform price distribution with frequent deep discount sales. Although it is generally true that store brands are less frequently and less deeply discounted than national brands, it is more common to find store brands on sale with greater regularity than in this case. We have examined other private-label price distributions and found this to be the case for products such as vegetable oil, paper towels, and tuna. For these product categories, the estimated and theoretical elasticities agree closely.

Close agreement between theoretical and estimated elasticities suggests that any bias in the least squares estimates due to correlation between the regressors and the error term is small. To provide more direct evidence on the magnitude of this bias, however, we approximated the asymptotic bias for the equal-probability nested logit and standard logit cases. A direct estimate of the remainder term in the Taylor series expansion was calculated by evaluating the matrix of the quadratic form at the sample average of prices and evaluating the quadratic form for each of the 50 observations. In the equal-probability case, the Hessian matrix of the log-probability locus does not change much as the price point is moved. Given this estimated error vector, $\hat{\varepsilon}$, we approximate the asymptotic bias by computing $(P'P)^{-1}P'\hat{\varepsilon}$, where P is the matrix of observed prices. For the nested logit model, equal-probability case, the biases in the cross-elasticities for brand 1 with respect to all other brands range between $-.048$ and .034 for $\beta_p = -2.0$ and between $-.19$ and .14 for $\beta_p = -4.0$. In the nested logit case, these approximate biases are at most 20% of the theoretical elasticities (see Table 2). For the logit-model case, the bias in the cross-elasticity of brand 1 with respect to brand 2 is .07 for $\beta_p = -2.0$ and .15 for $\beta_p = -4.0$. In the logit case, the approximate biases are less than 10% of the theoretical elasticities. We conclude that the bias is small for all practical cases.

The simulation results show that the restricted linear log-log model provides an excellent fit to data generated from a nested logit model, in terms of both goodness of fit and elasticity estimates. The dominant feature of the nested logit model, the proportional-draw and proportional-influence property, are both adequately captured by the log-log version of the model. In addition, the proposed constant-elasticity model yields residuals that exhibit little correlation with included variables and very little evidence of heteroscedasticity. It is important to note that these favorable results for the linearization of the logit model depend crucially on the assumption that all consumers face the same prices, the concavity of the log probability locus, and the price distribution not being concentrated at an extreme point.

2. AGGREGATION

The simulation results in Section 1 illustrate that a Taylor series approximation to the log choice probability is capable of accurately representing data arising from a nested logit model. Although linear models of this type can always be viewed as first-order approximations to any price-dependent choice model, the substance of the nested logit model is retained through the two sets of restrictions implied by Equation (2). The focus of this section is to investigate the extent to which these restrictions are altered when the micro model is aggregated over consumers with differing tastes.

2.1 Aggregating the Linear Model

It is assumed that the linear model is valid for each consumer but that consumers differ in their evaluation of the goods, $\boldsymbol{\beta}_0$, and in their aversion to price, β_p. Since the cross-elasticities are functions of these $\boldsymbol{\beta}$ parameters, different consumers will have different cross-elasticities. In this subsection, we focus on cross-elasticity measures and compare elasticity estimates computed by aggregating the micro-level values to estimates computed from aggregate statistics.

To perform the aggregation, we will initially assume that the parameters associated with the similarity of brands ($\boldsymbol{\lambda}$) are fixed for all consumers. The sensitivity of the results to this assumption will be examined in more detail in Section 2.4 and Appendix A. It is also assumed that all consumers are exposed to the same marketing variables. This assumption is justified, for instance, for aggregate scanning data from a particular store. When prices are the same for all consumers, the aggregated linear model [see Eq. (2)] becomes

$$\frac{\sum_{k=1}^{K} \ln \Pr(i \mid \ln \mathbf{p}_t, \boldsymbol{\beta}_k)}{K} = \text{cnst}^* + \ln \mathbf{p}_t' \left. \frac{\sum_{k=1}^{K} \dfrac{\partial \ln \Pr(i \mid \ln \mathbf{p}_t, \boldsymbol{\beta}_k)}{\partial \ln \mathbf{p}}}{K} \right|_{\ln \bar{\mathbf{p}}} + \text{error}^*, \quad (5)$$

where, as before, k is used to index the consumers. By the law of large numbers, the aggregate elements of the cross-elasticity matrix in Equation (5) are equal to the expected value of their micro counterparts, with the expectation taken over the distribution of consumer heterogeneity. Typically, aggregate scanner data involve thousands of different households so that the average over households and the expectation will agree closely. For goods i and j that are members of different submarkets,

$$\eta_{ij}(\text{macro}) = \mathrm{E}_{\boldsymbol{\beta}}[-\beta_p \Pr(j \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta})],$$

and for goods i and j that are members of the same submarket,

$$\eta_{ij}(\text{macro}) = \mathrm{E}_{\boldsymbol{\beta}}\left[-\beta_p \Pr(j) - \frac{1-\lambda_s}{\lambda_s} \cdot \beta_p \Pr(j \mid A_s)\right],$$

where $\Pr(j) = \Pr(j \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta})$ and $\Pr(j \mid A_s) = \Pr(j \mid A_s, \ln \bar{\mathbf{p}}, \boldsymbol{\beta})$. The aggregate elasticities are integrals of nonlinear functions and cannot be aggregated using well-known results for the linear-regression model. In Section 2.2, a distribution of consumer heterogeneity is generated by using micro-level scanner data. This distribution will then be used to integrate the micro elasticity expressions to produce aggregate elasticity measures.

2.2 A Distribution of Heterogeneity

The distribution of consumer heterogeneity, $f(\boldsymbol{\beta})$, plays a crucial role in the aggregate cross-elasticity expressions. To provide realistic evidence on the effect of heterogeneity, empirical estimates of the distribution are employed that are based on individual household data. The distribution was generated from a data set of 8,866 purchases of stick margarine over the course of a year (1985–1986) by households in Springfield, Missouri. Five brands of one-pound stick margarine, with a total market share of .74, were used to define the relevant choice set. The average price and market shares of the brands are given in Table 3. Demographic characteristics of the households were used to define 39 "homogeneous" consumer types with which to generate the empirical distribution. Family size, family income, education of the head of household, and occupational status were used to define these groupings. The creation of these 39 groups provides some information about the extent of heterogeneity in the consumer population. Because of the assumption that the consumers are homogeneous within each group, our estimated distribution of taste parameters will likely be less dispersed than the actual distribution of taste parameters across households. In Section 2.4, we dramatically increase the amount of heterogeneity in a sensitivity analysis of the results.

Table 3. Margarine Product Information

Brand	Average price[1]	Market share[2]
1. Imperial	$.85	.035
2. House brand 1	.55	.118
3. House brand 2	.43	.111
4. Blue Bonnet	.64	.056
5. Parkay	.59	.418
	Total	.736

NOTE: Panel-level data set, $N = 8{,}866$.
[1] Average shelf price over all 8,866 purchases in the data set.
[2] Percentage of purchases of all one-pound stick margarine.

Table 4. Aggregate Cross-elasticities Using the Empirical Distribution of Figure 3

Brand (j)	True value[1]	Approximation[2]
Same submarket		
1. Imperial		
2. House brand 1	.42	.45
3. House brand 2		
4. Blue Bonnet	.23	.23
5. Parkay	2.10	2.07
Different submarket		
1. Imperial	.08	.08
2. House brand 1	.23	.24
3. House brand 2	.29	.30
4. Blue Bonnet	.13	.13
5. Parkay	1.16	1.13

[1] $\eta_{ij}(\text{true}) = \sum_{k=1}^{39} N_k[-\beta_{pk}\Pr(j)_k - ((1-\lambda_s)/\lambda_s)\beta_{pk}\Pr(j \mid A_s)_k]$ (same submarket) and $\eta_{ij} = \sum_{k=1}^{39} N_k[-\beta_{pk}\Pr(j \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta}_k)]$ (different submarket), where N_k is the household count for the kth demographic grouping.
[2] $\eta_{ij}(\text{approx.}) = -\bar{\beta}_p\text{mkt}(j) - ((1-\lambda_s)/\lambda_s)\bar{\beta}_p\text{mkt}(j \mid A_s)$ (same submarket), and $\eta_{ij} = -\bar{\beta}_p\text{mkt}(j)$ (different submarket).

The brands of margarine include three national brands, a house brand, and a generic brand. Three separate submarkets of goods are defined and used in the estimation of the nested logit model. Imperial margarine, with an average price of $.85, was assumed to be in its own submarket. Parkay, Blue Bonnet, and a house brand (1) of margarine make up the second submarket, and a second house brand (2) makes up the third submarket. The assignment of brands to submarkets is based on their average price (see Table 3). This assignment of brands to submarkets requires the specification of one λ parameter for the three brands in the second submarket. As mentioned before, λ was initially assumed to be the same for each of the 39 groups.

Maximum likelihood estimates (MLE's) of λ and $\boldsymbol{\beta}_k$ ($k = 1, \ldots, 39$) were obtained using a scoring method that approximates the Hessian by the expected value of the outer product of the gradient vector (in Appendix B, problems of estimation in the nested logit model are discussed). General expressions for the gradients and outer products associated with the nested logit model can be found in the work of McFadden (1981, pp. 252–260). The coefficient for Blue Bonnet was arbitrarily set to 0 to achieve statistical identification, and a histogram of the resulting parameter estimates is provided in Figure 3. The log-likelihood for the 8,866 observations was maximized at a value of $-9{,}900.1$ when $\lambda = .60$.

2.3 The Effect of Heterogeneity

In Section 2.1, expressions for the aggregate-linear-model elasticities were presented. In this section, the

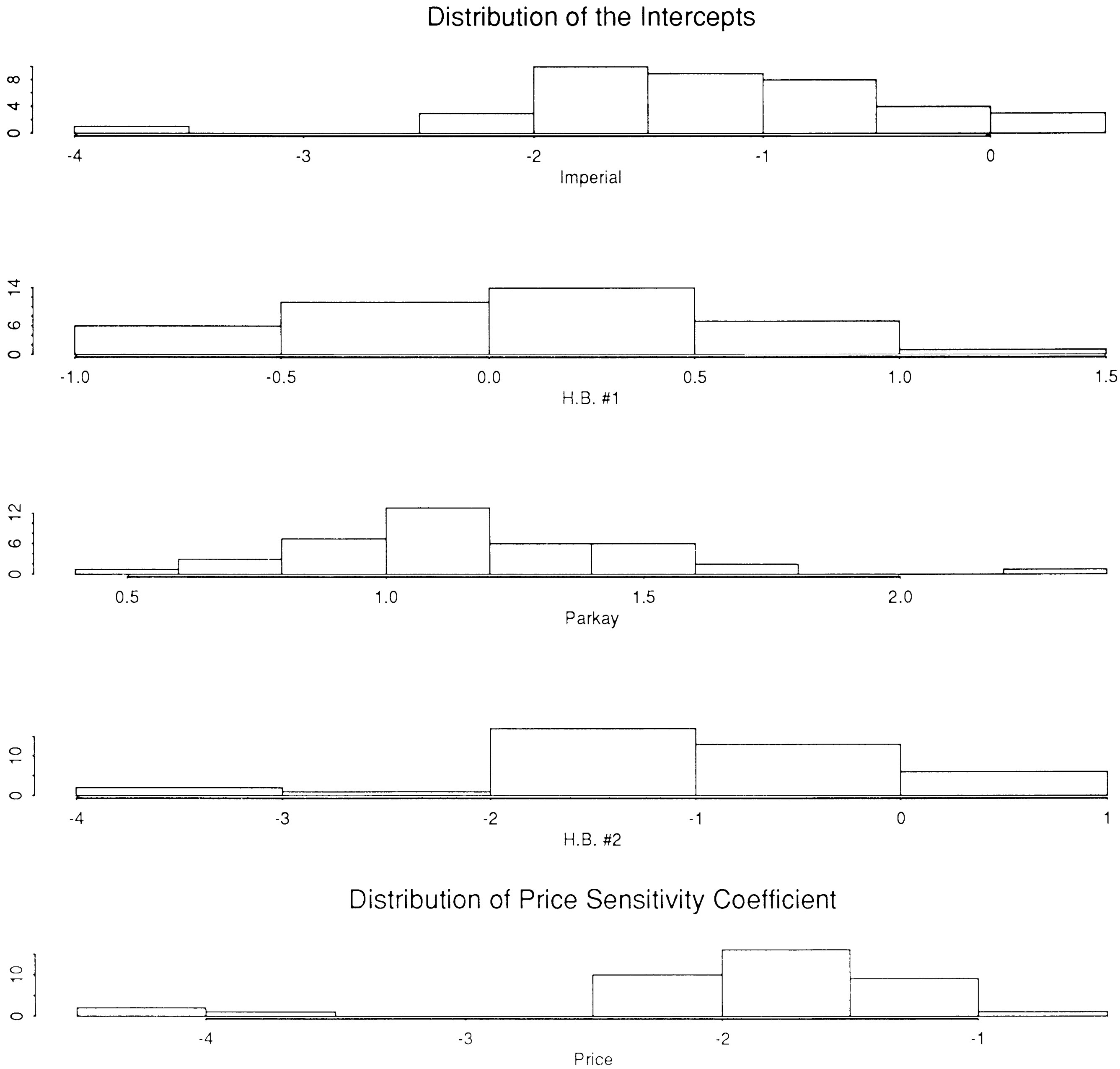

Figure 3. A Histogram of Parameter Estimates With the Coefficient for Blue Bonnet Arbitrarily Set to Zero.

distribution of heterogeneity is used to evaluate these expressions. Potential aggregation problems arise from the fact that the micro elasticity [see Eq. (2)] is the product of β_p and the probability, which is also a function of β_p.

In Table 4 on page 8, macro expressions for the cross-elasticities were obtained by aggregating the micro expression with respect to the distribution of heterogeneity. The aggregate values for the cross-elasticities are compared to values obtained by using expressions that substitute $\text{mkt}(j)$ for $\Pr(j)$ and $\text{mkt}(j \mid A_s)$ for $\Pr(j \mid A_s)$, where $\text{mkt}(j)$ refers to the average aggregate market share of brand j and $\text{mkt}(j \mid A_s) = \text{mkt}(j)/[\sum_{k\in A_s}\text{mkt}(k)]$. It is seen that these latter expressions closely approximate the true values and are made up of readily observed aggregate statistics.

The two expressions in Table 4 agree because, conditional on the mean-price vector, the choice probabilities, $\Pr(j \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta})$, are approximately constant with respect to changes in β_p. The derivative of the probability expression with respect to β_p for the logit model can be shown to be

$$\frac{\partial \Pr(j \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta})}{\partial \beta_p} = \Pr(j \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta}) \left[\ln \bar{p}_j - \sum_i \ln \bar{p}_i \Pr(i \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta}) \right],$$

and for the nested logit model the expression is equal to

$$\frac{\partial \Pr(j)}{\partial \beta_p} = \frac{\Pr(j)}{\lambda_{s'}} \left[\ln \bar{p}_j - \sum_{i \in A_{s'}} \ln \bar{p}_i \Pr(i \mid A_{s'}) \right]$$

$$+ \Pr(j) \left[\sum_{i \in A_{s'}} \ln \bar{p}_i \Pr(i \mid A_{s'}) - \sum_{s} \sum_{i \in A_s} \ln \bar{p}_i \Pr(i) \right],$$

where the arguments of the probability expression have been suppressed for clarity and where brand j is a member of submarket $A_{s'}$. For nearly perfect substitutes, the difference between $\bar{p}_j$ and the weighted average of prices of the other goods will tend to be small, resulting in conditional elasticity expressions that are essentially constant with respect to β_p.

Since the probability expression is approximately constant over the relevant range of β_p, we can assume that β_p and $\Pr(i)$ are independent, conditional on price. Therefore, the expectation of the product is equal to the product of the expectations. For two brands in different submarkets,

$$\begin{aligned} \eta_{ij}(\text{macro}) &= \mathrm{E}_{\boldsymbol{\beta}}[-\beta_p \Pr(j \mid \ln \bar{\mathbf{p}}, \boldsymbol{\beta})] \\ &\doteq -\bar{\beta}_p \mathrm{E}_{\boldsymbol{\beta}}[\Pr(j \mid \ln \bar{\mathbf{p}}, \beta_p, \boldsymbol{\beta}_0]. \end{aligned}$$

Furthermore, as illustrated in the simulation experiment in Section 1.3, the logit model's probability locus, conditional on $\boldsymbol{\beta}$, is well approximated by a linear model. Therefore, the preceding expression can be rearranged as follows:

$$\begin{aligned} \eta_{ij}(\text{macro}) &\doteq -\bar{\beta}_p \mathrm{E}_{\boldsymbol{\beta}}[\mathrm{E}_{\ln \mathbf{p} \mid \boldsymbol{\beta}}[\Pr(j \mid \ln \mathbf{p}, \beta_p, \boldsymbol{\beta}_0]] \\ &= -\bar{\beta}_p \mathrm{E}_{\boldsymbol{\beta}, \ln \mathbf{p}}[\Pr(j \mid \ln \mathbf{p}, \beta_p, \boldsymbol{\beta}_0)] \\ &= -\bar{\beta}_p \mathrm{mkt}(j). \end{aligned}$$

The same arguments can be applied to show that the cross-elasticity expressions for brands in the same market can be approximated by

$$\begin{aligned} \eta_{ij}(\text{macro}) &= \mathrm{E}_{\boldsymbol{\beta}} \left[-\beta_p \Pr(j) - \frac{1 - \lambda_s}{\lambda_s} \cdot \beta_p \Pr(j \mid A_s) \right] \\ &\doteq -\bar{\beta}_p \mathrm{mkt}(j) - \frac{1 - \lambda_s}{\lambda_s} \cdot \bar{\beta}_p \mathrm{mkt}(j \mid A_s) \end{aligned}$$

where $\mathrm{mkt}(j \mid A_s) = \mathrm{mkt}(j)/\Sigma_{k \in A_s} \mathrm{mkt}(k)$. The results of the numerical integration suggest that a minimal amount of bias is incurred when using the model on the aggregate level. The probabilities in the cross-elasticity expression, Equation (2), can be replaced with average market shares, resulting in a bias with an order of magnitude of .01.

2.4 Sensitivity Analysis

The results of the previous sections imply that the two sets of restrictions of the micro model can be extended to the macro level using market shares in place of choice probabilities. Two assumptions were used in the calculations that support this conclusion—(a) the similarity parameter, λ, was the same for all consumers and (b) the empirical distribution of $\boldsymbol{\beta}$ was representative of other distributions of heterogeneity.

In Appendix A, the sensitivity of the results to these two assumptions is investigated. The analysis indicates that the results of the previous section are robust to more extreme forms of heterogeneity in the λ and $\boldsymbol{\beta}$ distributions. In fact, the distribution of heterogeneity was expanded by a factor of five, yielding cross-elasticities of about 4.0, and the approximation is shown to remain accurate.

3. APPROXIMATION PROPERTIES OF THE MACRO LINEAR MODEL

In Section 1, the ability of a restricted linear model to represent a micro nested logit model was examined. The linear model was then aggregated in Section 2 to study effects of heterogeneity. The small aggregation errors found can be traced to the special structure of the nested logit model, as well as the assumption that the brands under consideration are close substitutes. Although the aggregation and linearization biases were found to be very small, the analyses in these sections do not provide direct evidence that a linear model with IIA restrictions would serve as a good approximation to aggregate data. The purpose of this section is to study the joint effect of these two approximations.

A simulation experiment is used to gauge the usefulness of the macro linear model. Multinomial data is generated on the household level, aggregated, and then fit with the restricted linear model. Goodness of fit and predictive performance measures are used to evaluate the adequacy of the linear approximation.

The simulation experiment proceeded as follows:

1. For each of 1,000 households, a β vector was drawn independently across households from the following distribution:
   a. β_p was drawn from a uniform distribution on $(-4, 0)$. This provides a wide range, which would include very price-sensitive behavior (see Table 1).
   b. Each of the intercepts, β_{0i}, were drawn from independent uniform distributions with a range of four units. The means of these uniform distributions were set so as to produce approximately equal mean-choice probabilities averaged over the 1,000 households. This range of intercepts covers all of the cells in Figure 1.
2. For each household, the β vector was used to generate choice probabilities for each of 50 price vectors used in Section 1.
3. Draws from a multinomial distribution with choice probabilities computed in steps 1 and 2 were made to simulate household choice behavior.

Table 5. Results of the Simulation Experiment of the Performance of an Aggregate Linear Model

	Market share		Adjusted R^2		Forecast mean squared error	
Brand	Mean	Range	Unrestricted	Restricted	Unrestricted	Restricted
1	.112	(.06, .21)	.85	.84	.0079	.0070
2	.112	(.07, .25)	.95	.93	.0077	.0071
3	.124	(.05, .24)	.95	.95	.0150	.0150
4	.117	(.06, .21)	.95	.95	.0075	.0065
5	.102	(.06, .20)	.86	.85	.0079	.0083
6	.132	(.06, .33)	.94	.93	.0080	.0076
7	.120	(.05, .30)	.94	.93	.0120	.0089
8	.180	(.11, .28)	—	—	—	—

NOTE: The likelihood ratio test is LR = 63.5, (44 df) p = .28. The sample size for estimation was 50 aggregate observations (formed by aggregating over 1,000 households); for prediction, it was 50 aggregate observations.

4. The actual choice data were then summed over households to form aggregate market shares.
5. Steps 2–4 were repeated to generate a hold-out sample for predictive validation of the model using fitted coefficients from the estimation sample.
6. The linear model was fitted with and without restrictions to the first 50 observations, and predictive testing was performed on the hold-out sample.

The results of the simulation experiment are presented in Table 5. The mean market shares reported in the table were used to impose the restrictions on the linear model. Note that there is a considerable range of market-share values, indicating substantial price sensitivity. In spite of the large range of market share and brand price, the linear model serves as a very good approximation. The R^2 of the unrestricted model for each of the seven brands is very high. More important, when the restrictions suggested by the micro model are imposed, the goodness of fit is virtually identical to the unrestricted model.

The ultimate test of model performance is to use the fitted coefficients from one sample to predict market shares in a hold-out sample (this is precisely how this model would be used in practice). A substantial reduction in the mean squared error of prediction results from using the restricted linear model. In six out of seven brands, the mean squared error is lower. The superior predictive performance is due to the reduction in the number of cross-elasticities that need to be estimated. In the unrestricted model, there are 49 cross-elasticities, but in the restricted model there are only five. In Table 6, the elasticity estimates and standard errors are given. A threefold to fourfold reduction in standard errors is achieved by imposing the restrictions.

Table 6. Elasticity Estimates and Standard Errors for the Simulation Experiment

Brand 1	Brand 2	Brand 3	Brand 4	Brand 5	Brand 6	Brand 7	Brand 8
			Unrestricted elasticities				
−1.6581	.3029	.2973	.3717	.3257	.1871	.2868	.1216
.2007	−1.6794	.2049	.2988	.6211	.2608	.2644	.2217
.1961	.3143	−1.7862	.3055	.3337	.3148	.2462	.3224
.2302	.3719	.3982	−1.7648	.0513	.2872	.2719	.3002
.5001	.4542	.4487	.3297	−1.9102	.3601	.3182	.2617
.2303	.1325	.2628	.1722	.4425	−1.9383	.9345	.5321
.2059	.1212	.2092	.1965	.0600	.9684	−1.8762	.6828
			Standard errors				
.1241	.0905	.0721	.0807	.1853	.1029	.1091	.1204
.0893	.0651	.0519	.0581	.1334	.0741	.0786	.0867
.0982	.0716	.0570	.0639	.1466	.0814	.0863	.0953
.0920	.0671	.0535	.0599	.1375	.0764	.0810	.0893
.1088	.0794	.0633	.0708	.1626	.0903	.0957	.1056
.1057	.0771	.0614	.0688	.1579	.0877	.0930	.1026
.0996	.0727	.0579	.0648	.1488	.0826	.0876	.0967
			Restricted elasticities				
X.XX	.30	.33	.31	.27	.29*	.26	.36*
.30*	X.XX	.33	.31	.27	.29	.26	.36
.30	.30	X.XX	.31	.27	.29	.26	.36
.30	.30	.33	X.XX	.27	.29	.26	.36
.30	.30	.33	.31	X.XX	.29	.26	.36
.22*	.22	.24	.22	.20	X.XX	88	.36
.22	.22	.24	.22	.20	.97*	X.XX	.36

* Standard error: .30 = .01; .29 = .02; .36 = .02; .22 = .02; .97 = .06.

Table 5 also contains a conventional likelihood ratio test of the restrictions. The test gives some weak evidence in favor of rejecting the linear restrictions. There are some problems interpreting this test, however. Since the linear model is only an approximation, the restrictions will never hold exactly, and it is expected that the test will reject the null with probability 1 for large samples. Therefore, the test must be viewed as simply a measure of reduction of in-sample goodness of fit, which results from imposing the restrictions. From a practical standpoint, the predictive performance measures reported earlier tend to be more relevant to users of aggregated choice models.

These simulation results further support our contention that the nested logit model can be linearized on the micro level and aggregated up to the macro level. Not only does the macro linear model fit extremely well, suggesting that approximation error is small, but the restrictions on the cross-elasticity matrix that hold for the micro nested logit model also hold up well on the aggregate level. Generally, it is still possible that a misspecified model could fit well by exploiting a favorable bias-variance tradeoff. In this case, however, the evidence presented in Section 1 on the local properties of the linear model rules out this possibility.

4. CONCLUDING REMARKS

This article provides analytic support to the empirical practice of fitting (nested) logit models to data aggregated across heterogeneous consumers. The results indicate that aggregate logit models work when consumers can all be described by the same underlying model and are all exposed to the same marketing variables (e.g., price). For the type of price series encountered in retail scanner data, the aggregation carries through even when consumers are heterogeneous with respect to price sensitivity and brand preference.

A characteristic aspect of the (nested) logit model is its implications for the competitive structure of demand. This feature is captured in the first-order term of a Taylor series expansion of the model about the price vector, resulting in the proportional-draw and proportional-influence restrictions associated with the IIA property. A simulation experiment was used to demonstrate that a restricted linear model was capable of accurately reflecting choice probabilities arising from these models.

An analysis of the cross-elasticity expressions supports the substitution of market shares for choice probabilities on the aggregate level when brands are close substitutes. An empirical distribution of heterogeneity was used to demonstrate that the aggregated cross-elasticities are well approximated when market shares are substituted for choice probabilities in the cross-elasticity expressions.

Further analysis showed that an aggregate linear model with the IIA restrictions imposed provides an excellent approximation to data aggregated from a heterogeneous micro logit model without requiring knowledge of the distribution of heterogeneity.

We conclude that the practice of fitting aggregate logit models or their linear approximations is justified under three conditions: (a) All consumers are exposed to the same marketing-mix variables, (b) the brands are close substitutes, and (c) the price distribution is not concentrated at an extreme value. These conditions are often met in the analysis of aggregate retail scanning data.

ACKNOWLEDGMENT

We thank the A. C. Nielsen Corporation for supplying the data used in the analysis.

APPENDIX A: SENSITIVITY ANALYSIS

Two assumptions were used to arrive at the results of Section 2.3 and are investigated in this appendix: (a) The similarity parameter, λ, was the same for all consumers and (b) the empirical distribution of $\boldsymbol{\beta}$ was representative of other distributions faced in practice.

The first assumption, a uniform perception of similarity across consumers, was imposed during estimation by requiring the parameter λ to be the same in each of the 39 homogeneous cells. This assumption can be relaxed by allowing λ to take on different values. MLE's for this less restrictive model were obtained, with the log-likelihood of the 8,866 observations equal to $-9{,}875$ as compared to $-9{,}900$ for the restricted model. The unrestricted model has 38 more parameters (λ's) than the unrestricted model, resulting in an insignificant improvement in the fit: $-2(9{,}875 - 9{,}900) = 50 < \chi^2_{38}(.05) = 53.4$.

A great deal of collinearity exists between λ and the other parameters ($\boldsymbol{\beta}_0$, β_p) in the model. It is typical to find asymptotic correlations of over .95 between elements of the β vector and the λ parameter. In each of the 39 homogeneous groups, the scoring algorithm took many iterations to converge, tracing out a slowly rising ridge along the λ dimension of the parameter space. We used both scoring and modified Gauss–Newton algorithms with analytic derivatives. In addition, we used transformations on the parameter vector to reduce the collinearity. Fundamentally, in spite of the large variation in prices, it is difficult to obtain precise estimates of the λ parameter. From experience with this data set, it appears that a large number of observations (close to 1,000) are needed to obtain any precision in estimation of the λ parameter. Since the data are relatively uninformative about the value of the λ parameter, we have retained the assumption of constant λ across groups. It sould be pointed out that the common practice of using standard logit models assumes that λ is equal to one for all households. In addition, the practice of using the conditional likelihood (iterated logit) procedure sug-

Table A.1. Sensitivity of Aggregate Cross-elasticity Approximation to the Expansion of the Simulated Distribution

Brand(j)	True value[1]	Approximation[2]
Expansion factor[3] = 1		
Same submarket		
1. Imperial		
2. House brand 1	.15	.14
3. House brand 2		
4. Blue Bonnet	.34	.37
5. Parkay	1.43	1.42
Different submarket		
1. Imperial	.05	.06
2. House brand 1	.19	.21
3. House brand 2	.12	.11
4. Blue Bonnet	.19	.20
5. Parkay	.81	.80
Expansion factor = 2		
Same submarket		
1. Imperial		
2. House brand 1	.62	.58
3. House brand 2		
4. Blue Bonnet	.50	.56
5. Parkay	2.50	2.50
Different submarket		
1. Imperial	.11	.14
2. House brand 1	.33	.31
3. House brand 2	.49	.41
4. Blue Bonnet	.26	.30
5. Parkay	1.34	1.35
Expansion factor = 3		
Same submarket		
1. Imperial		
2. House brand 1	1.35	1.32
3. House brand 2		
4. Blue Bonnet	.56	.64
5. Parkay	3.14	3.16
Different submarket		
1. Imperial	.19	.27
2. House brand 1	.69	.67
3. House brand 2	1.08	.92
4. Blue Bonnet	.26	.32
5. Parkay	1.57	1.60
Expansion factor = 5		
Same submarket		
1. Imperial		
2. House brand 1	2.97	3.07
3. House brand 2		
4. Blue Bonnet	.68	.73
5. Parkay	4.02	3.98
Different submarket		
1. Imperial	.41	.61
2. House brand 1	1.44	1.41
3. House brand 2	2.44	2.10
4. Blue Bonnet	.25	.34
5. Parkay	1.78	1.83

[1] See Table 3 for the definition of "true value" ($N_k = 1$).

[2] "Approximation" is the same as in Table 3, except that average probability is used instead of market share.

[3] "Expansion factor" is the constant used to inflate the range: Imperial (−2.5, .5); house brand 1 (−1, 1); Parkay (.7, 1.7); house brand 2 (−2, .5); price (−2.5, 0).

gested by McFadden (1981) can yield particularly inefficient estimates relative to the full MLE procedure used in this article when λ and the β vector are highly intercorrelated.

The second assumption concerns the representativeness of the empirical distribution and the ability to generalize the results of Section 2. Of primary concern is that coefficients for the 39 homogeneous groups describe averaged as opposed to individual behavior. In particular, consider the situation in which a "switcher," who is highly price sensitive, is combined with a "brand-loyal" buyer with a low price-sensitivity coefficient, resulting in an aggregate consumer who occasionally switches brands (moderate β_p). It is therefore of interest to investigate the sensitivity of the results to more extreme values of β_p.

More price-sensitive distributions were generated using the empirical distribution as a reference. A uniform distribution was used to independently generate the price and the four intercept coefficients with the same range as the coefficients reported in Figure 3. A total of 10,000 draws from this multivariate uniform distribution were used to study the performance of the macro approximation to the aggregated micro cross-elasticities. For each draw from the distribution, the cross-elasticity expression [Eq. (2)] was evaluated at the mean level of price. These values were then averaged over the different draws and compared to the approximation that uses average probabilities (market shares) and the average price coefficient.

The range of the simulated distribution was systematically increased to reflect a more price-sensitive consumer population. The range of each intercept term, $\boldsymbol{\beta}_0$, was expanded symmetrically about its mean, while the range of the price coefficient, which is negative, was increased by expanding its lower limit. The performance of the macro approximation for different ranges of the simulated distribution is reported in Table A.1. The results indicate that the magnitude of the bias is still of order .01 even when the range of the distribution is three times that of the empirical distribution.

In summary, the sensitivity analysis indicates that average market share can be used in place of probability when imposing the across- and within-equation restrictions of Equation (1) on a multivariate linear model. This result appears to be robust to the degree of price sensitivity in the population. Furthermore, the assumption requiring consumers to have the same level of correlation does not significantly affect the fit of the model.

APPENDIX B: MATRIX OF SECOND DERIVATIVES FOR LOGIT AND NESTED LOGIT MODELS

For the logit model, this matrix has the jth diagonal element equal to $\{\eta^L_{jj}\eta^L_{\cdot j}\}$ and the $\{j, k\}$ off-diagonal element equal to $\{\eta^L_{\cdot j}\eta^L_{\cdot k}\}$, where $\eta^L_{jj} = \beta_p[1 - \Pr(j)]$ and $\eta^L_{\cdot j} = -\beta_p\Pr(j)$, expressions equivalent to that of the logit model's self- and cross-elasticity. The second derivative matrix of the nested logit model is somewhat more complicated, with expressions dependent on the submarket membership of the numerator and denominator terms. Assuming that brand i is an element of submarket A_s, the $\{j, k\}$ element of the second derivative matrix is equal to $\{\eta^L_{\cdot j}\eta^{NL}_{jk} + \gamma_{jk}\}$ for i and j, $k \in A_s$; $\{\eta^L_{\cdot j}\eta^{NL}_{jk}\}$ for $i \in A_s$ and j, $k \in A_{s'}$ where $A_{s'}$ is a different submarket, and $\{\eta^L_{\cdot j}\eta^L_{\cdot k}\}$ for j and k each members of a different submarket, where η^{NL}_{jk} is the expression for the nested logit elasticity defined by Equation (1), $\gamma_{jk} = [(1 - \lambda_s)/\lambda_s^2]\eta^L_{\cdot j}\eta^L_{jk}$, $\eta^L_{jk} = \beta_p[\delta_{jk} - \Pr(k)]$, and δ_{jk} is the Kronecker delta. As expected, the second derivative matrix for the nested logit model reduces to that of the logit model when $\lambda_s = 1$. In addition, the matrices are symmetric, since $\eta^L_{\cdot j}\eta^{NL}_{jk} = \eta^L_{\cdot k}\eta^{NL}_{kj}$ and $\eta^L_{\cdot j}\eta^L_{jk} = \eta^L_{\cdot k}\eta^L_{kj}$.

[*Received July 1988. Revised June 1990.*]

REFERENCES

Guadagni, Peter M., and Little, John D. C. (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2, 203–238.

Heckman, James (1982), "Statistical Models for the Analysis of Discrete Panel Data," in *Structural Analysis of Discrete Data: With Econometric Applications*, eds. C. Manski and D. McFadden, Cambridge, MA: MIT Press, pp. 114–178.

Krishnamurthi, Lakshman, and Raj, S. P. (1988), "A Model of Brand Choice and Purchase Quantity Price Sensitivities," *Marketing Science*, 7, 1–20.

Leeflang, P., and Reuyl, J. (1984), "On the Predictive Power of Market Share Attraction Models," *Journal of Marketing Research*, 21, 211–225.

Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, U.K.: Cambridge University Press.

McFadden, D. (1978), "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Residential Location*, eds. A. Karlquist et al., Amsterdam: North-Holland, pp. 75–96.

——— (1981), "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data*, eds. C. Manski and D. McFadden, Cambridge, MA: MIT Press, pp. 198–272.

——— (1984), "Econometric Analysis of Qualitative Response Models," in *Handbook of Econometrics* (Vol. 2), eds. Z. Griliches and M. Intriligator, Amsterdam: North-Holland, pp. 1395–1457.

Muellbauer, J. (1975), "Aggregation, Income Distribution and Consumer Demand," *Review of Economic Studies*, 42, 525–543.

Stoker, T. (1982), "Completeness, Distribution Restrictions, and the Form of Aggregate Functions," *Econometrica*, 52, 887–908.

Theil, H. (1954), *Linear Aggregation of Economic Relations*, Amsterdam: North-Holland.

White, H. (1980), "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21, 149–170.