Printed: 18 Oct 2001

Reflecting Uncertainty about Economic Theory when Estimating Consumer Demand

by

Alan L. Montgomery

Carnegie Mellon University
Graduate School of Industrial Administration
255A Posner Hall
5000 Forbes Avenue
Pittsburgh, PA 15213-3890

email: alan.montgomery@cmu.edu

October 2001

The author would like to acknowledge Peter Rossi, Peter Boatwright, and Wei Zhang for their contributions to research upon which this paper is based.

Reflecting Uncertainty about Economic Theory when Estimating Consumer Demand

Abstract:

Economic theory provides a great deal of information about demand models. Specifically, theory can dictate many relationships that expenditure and price elasticities should fulfill. Unfortunately, analysts cannot be certain whether these relationships will hold exactly. Many analysts perform hypothesis tests to determine if the theory is correct. If the theory is accepted then the relationships are assumed to hold exactly, but if the theory is rejected they are ignored. In this paper we outline a hierarchical Bayesian formulation that allows us to consider the theoretical restrictions as holding stochastically or approximately. Our estimates are shrunk towards those implied by economic theory. This technique can incorporate information that a theory is approximately right, even when exact hypothesis tests would reject the theory and ignore all information from it. We illustrate our model with an application of this data to a store-level system of demand equations using supermarket scanner data.

Keywords: Approximations, Demand Estimation, Hierarchical Bayesian Models, Random Coefficient Models

Journal of Economic Literature Classification Codes: C11, C31, C5, M3

1 Introduction

A basic goal of many marketing analysts and econometricians is the estimation of consumer demand models. More specifically analysts might be interested in estimating price and promotional elasticities that can be used in developing better marketing strategies. Economics provides a large body of theory to guide an analyst in constructing a consumer demand model. Unfortunately, the analyst can never be entirely confident that this theory is correct. In practice many marketing analysts may assume that nothing is known about expenditure and price elasticities due to their uncertainty about whether all theoretical assumptions are met. However, even if the assumptions of these theories are not met exactly the theory might still be approximately correct. It is this notion of approximation that we formalize in this chapter.

The focus of many econometric studies is to determine the extent that the data supports a particular theory. Classical approaches to testing lead the analyst to an all or nothing approach. If the data provides strong confirmatory evidence then the analyst usually proceeds under the assumption the theory is correct and estimates the model. However, if the theory is rejected then the analyst simply rejects the theory and ignores all information from the theory. Sharp tests of null hypotheses in large datasets frequently lead to rejection if the tolerance for type I errors is not increased with the sample size. Large datasets can result in very precise tests that often miss the fact that the theory may not be perfect but provides a reasonable approximation to the true process.

In this paper we propose a Bayesian framework in which uncertainty about a theory is directly represented in the model. Our procedure prescribes treating the theory as a prior and follows recent work by Montgomery and Rossi (1999). The prior is centered over the theory, so the mean is what would be expected under a restricted model in which the theory holds exactly. The variance of the prior is allowed to vary depending upon the analyst's confidence about the theory. For example, Slutsky symmetry may require equating two parameters. In our methodology we can represent these two parameters as two draws from a common distribution, which we call the hyper-prior. If we are certain that the theory holds exactly then the variance of this hyper-prior is zero, and the restrictions are implicitly fulfilled. However, we wish to entertain the notion that the theory may only be approximately correct. Hence we allow the variance of the hyper-prior to vary, perhaps substantially. We may be uncertain about the exact values of the parameters of this distribution and place a prior on the parameters of this hyper-prior.

The analyst can incorporate prior beliefs about the adequacy of the theory and gain useful information even if the theory is technically wrong, but is approximately right. It is this notion of approximation that we are especially interested in representing. The estimator proposed results in adaptive shrinkage towards the theory. Adaptivity refers to the ability of the model to decrease the amount of shrinkage if the data disagrees with the prior. As more information or data is observed less shrinkage occurs and we can learn more about how good an approximation the theory provides to the observed data. Our framework allows the flexibility to mimic the estimates of a model achieved by an economist who holds to theory dogmatically, an analyst who ignores theory entirely, or an analyst's whose beliefs fall in between by choosing the prior appropriately. Our framework also contrasts with statistical formulations of shrinkage estimators in marketing that move estimates towards one another due to empirical similarities without any theoretical justification (Blattberg and George 1991, Montgomery 1997).

Economic theory provides many possible sources of information. First, it can provide information about relationships that elasticities should satisfy, such as adding up or Slutsky symmetry. Second, specific assumptions about utility may result in more parsimonious demand models. For example, the assumption of additive utility results in a very parsimonious model. Many marketing models, like logit choice models and conjoint models, are based upon the assumption of an additive utility model. Third, elasticity estimates for one economic agent may be similar to those of other agents. Finally, previous empirical research may enable us to directly postulate priors on the parameters, i.e., the elasticity matrix is made up of negative elements on the diagonal (negative own-price elasticities) and small positive cross-diagonal elements (modest direct substitution between products within a category). In this paper we show how these prior sources of information can be parameterized and incorporated into a hierarchical Bayesian framework.

Previous research in marketing has considered economic restrictions in demand models (Berndt and Silk 1993), restricted relationships between elasticities (Allenby 1989) in the context of market structure, and the use of hierarchical models to shrink estimates across stores and households (Blattberg and George 1991, Allenby and Rossi 1993, Montgomery 1997). Our framework provides a unifying treatment to these ideas. By evaluating these components together we can appreciate the significant gains in measuring demand that can be had by incorporating theory in a stochastic manner.

The outline of this paper is as follows. First we present our demand model in section 2 and the restrictions implied by economic theory. Section 3 goes on to show how these restrictions can be incorporated stochastically in a hierarchical Bayesian model. A short example is given to illustrate these restrictions. The estimation of this Bayesian treatment is presented using the Gibbs Sampler in Section 4. Section 5 provides an empirical example of shrinkage of price elasticities towards those restrictions implied by an additive utility model. This example estimates store level demand systems using weekly UPC scanner data for the refrigerated orange juice category at Dominick's Finer Foods (DFF), a major Chicago supermarket chain. Section 6 considers a further application of this framework by considering changes in market structures. We conclude the paper in section 7 with a discussion of these results along with suggestions for implementing these techniques in other problems.

2 Sales Response Modeling

We begin not with a formal theory of consumer behavior from which we derive a model of demand as in customary in econometrics, but with a sales response model. Both models try to capture the relationship between quantity and price, the essential difference is in terms of interpretation. A sales response model is a model motivated by statistical considerations, for example a logarithmic relationship between quantity and price is commonly observed by marketing researchers, and is not justified on theoretical grounds. For a discussion of sales response modeling from a marketing perspective see Blattberg and Neslin (1990). On the other hand an econometric model places many restrictions upon the functional form and parameters. The strength of the econometric model is our ability to estimate more parsimonious forms, while its weakness is the requirement to make many assumptions that may be suspect or untestable. In contrast, these strengths are reversed for a sales response model. It makes fewer assumptions about demand, but this flexibility comes at the price of an increased number of parameters.

To begin our analysis of demand we choose a double log functional form for our sales response model. This form is chosen since previous empirical work has shown it to be a good one that captures the logarithmic relationship between quantity and price. Our technique is quite general and does not rely upon a logarithmic functional form, in fact it could be applied to many demand models, such as the AIDS, translog, or Rotterdam model. Our sales response model can be written in vector form:

$$\ln(\boldsymbol{q}_{st}) = \alpha_s + \mu_s \ln(x_s) + \mathrm{H}_s \ln(\boldsymbol{p}_{st}) + \boldsymbol{e}_{st}\ , \quad \boldsymbol{e}_{st} \sim N(\mathbf{0}, \Sigma_s) \tag{1}$$

Where there are M products in store s at week t, $\boldsymbol{q}_{st}$ and $\boldsymbol{p}_{st}$ are vectors of movement and price, and x_{st} is store expenditures ($x_{st}=\Sigma_i p_{ist} q_{ist}$, the ith subscript denotes the ith product in the vector). Our framework is parameterized by the store subscript s, although this index can be interpreted quite generally as an index for different households, markets, or industries, depending upon the application.

The basic problem one encounters in estimating model (1) is the large number of parameters. For example, if there are 10 products and 100 stores as would be found in one small category of a moderately sized retailer, this results in more than 10,000 parameters that must be estimated. In a typical supermarket retailing application perhaps two or three years of weekly observations would be available. While this is a large amount of data, if the retailer wishes to estimate demand for each store separately then it may be difficult to estimate store-level demand with any degree of statistical precision. This problem becomes acute if the retailer wishes to formulate an elasticity based pricing strategy, since the high degree of parameter uncertainty may result in strange pricing prescriptions. For example, positive own-price elasticities may result in undefined optimal prices, or erroneously signed cross-price elasticities may result in higher overall levels of prices.

2.1 An Economic Interpretation of the Sales Response Model

We can reinterpret our sales response model in (1) as a system of demand equations. The H represents uncompensated price elasticities and the μ are expenditure elasticities. Usually x would represent income, and demand would be defined over all products consumed. However, we do not have a measure of weekly income for consumers that shop at store s. Therefore, we use store expenditures[1] and consider (1) as a subset demand model for the products in store s. Subset demand models possess all the usual properties

---

1. Our dataset in sections 5 and 6 consists of 26 categories with over 5,000 UPC's. This dataset accounts for 25% of total store sales. It is this dataset that we use to compute store expenditures. While it would be desirable to have use all products in a store, many products are not scanned, like produce and meat which account for 50% of store sales. Therefore, our expenditure variable can be thought of largely as grocery sales.

of full demand models, although the income elasticities are now interpreted as store expenditure elasticities. For a further discussion of subset demand models see Deaton and Muellbauer (1983).

A store expenditure elasticity states how product sales are effected as store shoppers purchase more groceries. Specifically μ_i states the effect of an increase of store expenditures on the movement for product i. If $\mu_i<0$ then product sales decrease as expenditures grow (an inferior product), and when $\mu_i>1$ product sales garner a larger share of overall sales. Since, this expenditure elasticity is conditional upon store sales, it cannot be used to determine how store traffic is affected by competition and cross-category promotions.

The price elasticity matrix can be decomposed into expenditure and price effects:

$$\mathrm{H}_s = \mathrm{E}_s - \mu_s \boldsymbol{w}_s' \tag{2}$$

Where the uncompensated cross elasticity (H_s) for store s is the sum of a substitution effect, the compensated cross elasticity matrix (E_s), and an income effect, which is the outer product of the income elasticities (μ_s) and the budget or market shares ($\boldsymbol{w}_s$). The ith element of the market share vector is defined as $w_{is}=p_{ist}q_{ist}/x_{st}$. We use the usual definition of substitutes ($[\mathrm{E}_s]_{ij}>0$), complements ($[\mathrm{E}_s]_{ij}<0$), and independent products ($[\mathrm{E}_s]_{ij}=0$) that rely upon compensated elasticities.

Substituting (2) into (1) yields a demand system in terms of compensated elasticities. We also augment this model with cross- feature and deal variables to control for other marketing mix effects. Finally, we assume that the category employed in our analysis is independent of other categories, so our system only uses the set of products within a category. The final form of the demand model that we employ in this paper is:

$$\ln(\boldsymbol{q}_{st}) = \boldsymbol{\alpha}_s + \boldsymbol{\mu}_s \ln(x_s/P_{st}) + \mathrm{E}_s \boldsymbol{p}_{st} + \Theta_s \boldsymbol{f}_{st} + \Psi_s \boldsymbol{d}_{st} + \boldsymbol{\epsilon}_{st}\,, \quad \boldsymbol{\epsilon}_{st} \sim N(\mathbf{0}, \Sigma_s) \tag{3}$$

Where $P_{st}=\exp\{\Sigma_i w_{ist}\ln(p_{ist})\}$ is a Divisia price index, $\boldsymbol{f}_{st}$ and $\boldsymbol{d}_{st}$ are the vectors of feature and display variables for store s during week t.

2.2 Economic Theory

If we interpret (3) not as a sales response model, but as a system of demand equations then economic theory is very informative about the parameters or more specifically the conditions that the price elasticities must satisfy. These restrictions follow as a consequence of underlying assumptions about utility: reflextivity, completeness, transitivity, continuity, and nonsatiation. In our discussion we only express the consequences of these assumptions on demand and do not provide their derivations. For additional reference we refer the reader to Deaton and Muellbauer (1983, pp. 43-46).

Adding-Up: The budget constraint imposes the following condition on demand:

$$\boldsymbol{p}_s' \boldsymbol{q}_s = x_s \tag{4}$$

This equation can be differentiated with respect to price and expenditures to yield the following:

$$\boldsymbol{w}_s' \boldsymbol{\mu}_s = 1 \tag{5}$$

and

$$\boldsymbol{w}_s' \mathbf{H}_s = \boldsymbol{w}_s \quad \Rightarrow \quad \boldsymbol{w}_s' \mathbf{E}_s = \mathbf{0} \tag{6}$$

These restrictions reduce our demand system by 1 and *M* parameters respectively.

Homogeneity: The assumption of homogeneity implies that if we double all prices and income then the budget shares remain unchanged (no money illusion):

$$\mathbf{H}_s \boldsymbol{\iota} = \boldsymbol{\mu}_s \quad \Rightarrow \quad \mathbf{E}_s \boldsymbol{\iota} = \mathbf{0}\ , \quad \text{where } \boldsymbol{\iota} = (1 \;\; 1 \;\; \cdots \;\; 1)' \tag{7}$$

This restriction reduces our demand system by an additional *M* parameters.

Symmetry: The symmetry restriction is derived from the double differentiability of the cost function or the symmetry of the Slutsky matrix[2], and implies that the compensated elasticity matrix when weighted by the budget shares is symmetric:

$$diag(\boldsymbol{w}_s)\, \mathbf{E}_s = \mathbf{E}_s'\, diag(\boldsymbol{w}_s) \tag{8}$$

Notice that symmetry results in a large reduction in the order of the demand system, specifically by ½M(M-1) terms or a 45% reduction in the cross-price elasticities with 10 products (*M*=10).

---

2. The (*i,j*)th element of the Slutsky matrix is defined as $s_{ij} = \frac{\partial q_i}{\partial x} q_j + \frac{\partial q_i}{\partial p_j}$.

Many marketers may worry that Slutsky symmetry may be too restrictive. It is well established in marketing (Blattberg and Wiesniewski 1989 and Kamakura and Russell 1989) that uncompensated price elasticity matrices are asymmetric. For example price changes of higher quality brands effect sales of lower quality brands, but price changes of lower quality brands have only small effects on high quality brands. These asymmetries are consistent with economic theory and can be explained by differences in market shares and expenditure elasticities, and do not require asymmetries in the compensated elasticity matrix. Consider an example with three brands (premium, national, and store brands) and the following parameters:

$$\boldsymbol{\mu} = \begin{bmatrix} 1.5 \\ 1.0 \\ .5 \end{bmatrix}, \quad \boldsymbol{w} = \begin{bmatrix} .4 \\ .3 \\ .3 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} -2.0 & .50 & .50 \\ .67 & -2.0 & .50 \\ .67 & .50 & -2.0 \end{bmatrix}$$

Employing (8) we find the uncompensated price elasticity matrix becomes:

$$\mathbf{H} = \begin{bmatrix} -2.6 & .05 & .05 \\ .27 & -2.3 & .20 \\ .47 & .35 & -2.15 \end{bmatrix}$$

The asymmetry in the compensated elasticity matrix (E) between the premium and national brands is due to market share differences (***w***), while there is no assymetry between the national and store brands. However, upon evaluation of the uncompensated elasticity matrix (H), we find pronounced price asymmetries between these three brands. The asymmetry in price elasticities is due to expenditure effects (μ), i.e., as expenditures grow people purchase higher quality brands.

Sign-Restrictions: Downward sloping demand curves require the Slutsky matrix to possess a negative semi-definite property:

$$\forall \delta \quad \delta' S_s \delta \leq 0 \tag{9}$$

In addition to the usual consequence that the own-price elasticities must be non-positive, it further implies that any linear bundle of products must also have a non-positive elasticity. A common concern in marketing is that price elasticities can frequently be of the wrong sign.

2.3 Weak Separability and Market Structure

Another component of economic theory that can induce relationships among price elasticities are ones about the relationships between products. Many marketing researchers have suggested a hierarchical structure for market competition (to name just a few see Allenby 1989, Vilcassim 1989, Srivastava et al 1981). This hierarchy is illustrated in Figure 1. For example, a consumer first decides whether to buy liquid or dry laundry detergent, and then considers which product to buy within the subcategory. Products at the same level within a branch are strong substitutes, while competition between items in different branches is weaker and have the same general pattern.

Insert Figure 1 about here.

At the heart of most research on market structure is weak separability of the utility function. Frequently these hierarchical structures are justified by assuming that consumers engage in some type of hierarchical budgeting process. Allocating budget shares to large groups of products like groceries, housing, transportation, etc., and then deciding upon allocations to individual products within each category. This broad budget allocation process allows us to break the problem into smaller units by assuming groups of products within a category can be weakly separated from one another. Categories can be partitioned into subcategories, until finally we reach the individual product level. The general form of the utility function for an individual household is of the form:

$$utility \ = \ \nu\left(\nu_1(q_1), \nu_2(q_2), \cdots, \nu_C(q_C)\right) \tag{10}$$

Where $\boldsymbol{q}_i$ is the vector of quantities for all items in the ith category of which there are C categories.

The hierarchy in the cost or utility functions naturally imposes a structure in the demand model. It can be shown that weak separability imposes the following restriction on the elasticity matrix:

$$\epsilon_{ijs} \ = \ \kappa_{GHs}\, \mu_{is}\, \mu_{js}\, w_{js} \quad \text{if } i \in G \text{ and } j \in H \tag{11}$$

where ϵ_{ijs} is the i,jth element of the matrix E and κ_{GHs} is a parameter that may depend upon x. In other words, the elasticities that capture substitution within a category can take on a general form, but those elasticities representing intra-category substitution must follow a restricted pattern that is common for all items in the subcategories.

2.4 Strong Separability and Additive Utility

The restrictions discussed in the previous subsection hold for many families of utility functions. If the analyst is willing to make stronger assumptions about a specific form of utility then this can also result in much simpler forms to demand. One possibility is to assume utility is additive or strongly separable across products:

$$utility = \nu\left(\textstyle\sum_i \nu_i(q_i) \right) \tag{12}$$

where q_i is the quantity of the *i*th product consumed. Additivity has a long history in economic models (Lancaster 1966) and empirical applications in marketing like logit modeling (Guadagni and Little 1983) and conjoint analysis (Green and Rao 1971). Often additivity is argued at the attribute level in logit and conjoint applications and not the higher, product level as we have suggested.

Additive utility models result in parsimonious—but restrictive—demand models:

$$\mathbf{E}_s = \phi_s\, diag(\mu_s) - \phi_s \mu_s (\mu_s \odot \boldsymbol{w}_s)' \tag{13}$$

Notice that the cross-elasticity matrix is populated solely by the expenditure elasticities (μ), market shares (w), and a general substitution parameter (ϕ). This restricted elasticity matrix has M+1 parameters, not including the market shares, as opposed to M^2+M for the unrestricted form. Additionally, the elasticity matrix in (13) will satisfy the properties of demand models given in the previous subsections. However, the incredible parsimony of the additive model also comes a high penalty. Namely, either all products must be substitutes or complements, and the level of substitution or complementarity is dictated by a single parameter (ϕ) and the expenditure elasticities.

It might seem odd to many economists to propose an additive utility structure, since many econometric studies have rejected additivity (Barten 1969, Deaton 1974, Theil 1976, Deaton 1978). However, we are proposing an additive utility structure at a very low-level (e.g., similar products within a single category), while most have considered additivity at high levels in a hierarchical structure (food, clothing, housing). Additive utility implies that the utility gained from one product is unaffected by the utility of other products. For example, there is no interaction in utility from purchasing Minute Maid and Tropicana orange juice together. This makes a great deal of sense for products within a category, which are typically direct substitutes and not used together. However, additivity may not make sense across products from different categories that when combined together can interact, such as bread and peanut butter.

2.5 Pooling and Heterogeneity

The last set of restrictions that we propose are not really theoretical ones, but ones motivated from practice. It is quite common to observe multiple agents, either consumers or stores as in our case. A common assumption is to simply pool the observations across all agents and assume identical elasticities as in the following relationship:

$$\mathbf{E}_s = \mathbf{E}, \quad \boldsymbol{\mu}_s = \boldsymbol{\mu} \tag{14}$$

Recently there has been a great deal of research in marketing studying heterogeneity, for a recent review refer to Rossi and Allenby (2000). One technique is to capture heterogeneity in a random coefficient model:

$$\mathbf{E}_s = \mathbf{E} + U_s, \quad \boldsymbol{\mu}_s = \boldsymbol{\mu} + \boldsymbol{u}_s \tag{15}$$

This specification has been studied extensively starting with the early work by Swamy (1970) from a frequentist perspective and by Lindley and Smith (1972) from a Bayesian interpretation as a hierarchical Bayesian model.

3 A Bayesian Specification

The economic theory proposed in section 2 is simply that, a theory. As with any theory the assumptions upon which it is based are subject to question. One technique is to inject randomness into the axioms upon which the theory is constructed, namely reflexivity, completeness, transitivity, continuity, nonsatiation, and convexity. However, our belief is that the theory is quite reasonable. But we also realize that there are many reasons to believe that this model may not be entirely correct. Our theory is at an individual level, but our data is at an aggregate level. We know that aggregate demand models will satisfy additivity and other economic properties only under certain conditions (see Deaton and Muellabuer (1983, pg. 148-166) for a discussion of the conditions for exact aggregation to hold). Additionally, these theoretical relationships are abstractions that omit certain effects (intertemporal substitution, savings, nonlinear budget constraints, etc.) or be subject to measurement errors.

Our belief is that the theory should be a reasonable approximation to the observed process, but will hold approximately or in a stochastic manner and not exactly. This contrasts with the usual pre-testing approach which would test whether these effects hold exactly and then totally discard them if they do not meet a specific p-value. An essential difference is that in our framework—even if the theory is not entirely supported by the data—the information implied by the theory will not be completely ignored. To explicitly incorporate the notion of approximation into our model follow the approach proposed by Montgomery and Rossi (1999). First, we assume that the price elasticities have the following prior distribution:

$$\epsilon_s \mid \beta_s, \phi \sim N(\bar{\epsilon}_s, \Delta) \tag{16}$$

where

$$\beta_s = [\beta_{1s}' \;\; \beta_{2s}' \;\; \cdots \;\; \beta_{Ms}']' \;, \quad \beta_{is} = [\alpha_{is} \;\; \mu_{is} \;\; \theta_{is} \;\; \psi_{is}]' \tag{17}$$

This distribution will be centered around the restrictions implied by our theory, $\bar{\epsilon}_s$, and the variance around these restrictions represents our confidence in this approximation. Δ can be interpreted as the degree to which an approximation is valid. If Δ is small then these restrictions will effectively be enforced. Conversely large values of Δ will result in estimates that may bear little resemblance to the restricted parameter estimates, i.e., unrestricted parameter estimates.

We are not able to assess the parameters of this prior directly, so we place a prior on this prior. To avoid confusion the prior in (16) is called the hyper-prior. Additionally, we assume that an elasticity matrix that conforms to an additive utility structure is reasonable, which implicitly satisfies all the relationships outlined in section 2. We propose the following relationship:

$$\epsilon_s = \mathrm{vec}(\mathbf{E}_s) \;, \quad \bar{\epsilon}_s = \mathrm{vec}(\bar{\mathbf{E}}_s) \;, \quad \bar{\mathbf{E}}_s = \phi_s \mathrm{diag}(\mu_s) - \phi_s \mu_s (\mu_s \odot \boldsymbol{w}_s)' \tag{18}$$

We place the usual multivariate normal prior on the remaining store parameters:

$$\phi_s \sim N(\bar{\phi}, \lambda_\phi) \tag{19}$$

$$\beta_s \sim N\left(\bar{\beta}_s, \Lambda\right) \tag{20}$$

An important reason for expressing the prior on E_s conditionally upon μ_s and ϕ_s is to avoid problems on nonlinearity. Notice that while E_s is conditionally linear upon μ_s, unconditionally our prior is nonlinear in μ_s. Additionally the prior implies that the price elasticity elements will be correlated, which can help counter the effect of multicollinearity in a typical price dataset.

Notice that our priors on E_s, μ_s, and ϕ_s are exchangeable across stores. It is this exchangeability that will drive the shrinkage of one stores parameter estimates towards another. The store to store variation of the expenditure elasticities (μ_s) is governed by Λ, and variation in the price elasticity matrix (E_s)—both across store and deviations from the theory—is governed by the Δ matrix. If Λ and Δ are zero then there will be no random variation across stores and the cross elasticity matrix will be held to its restricted pattern, i.e., the estimates will be close to a pooled restricted model. If Λ and Δ are large then the information from the hyper-distribution will be discounted and the parameter estimates will be close to individual store models.

Since we cannot directly evaluate Λ and Δ, we formulate a prior on these matrices, and use the data to make inferences about the variation present in the data. In our Bayesian framework we assume independent Wishart priors for each of these matrices:

$$\Delta^{-1} \sim Wishart(\nu_\Delta, V_\Delta^{-1}) \ , \quad \Lambda^{-1} \sim Wishart(\nu_\Lambda, V_\Lambda^{-1}) \tag{21}$$

We parameterize the prior on these priors as: $V_\Delta = \nu_\Delta\, k_\Delta\, \bar{V}_\Delta$ and $V_\Lambda = \nu_\Lambda\, k_\Lambda\, \bar{V}_\Lambda$, so that these priors are centered over $\bar{V}_\Delta^{-1}/k_\Delta$ and $\bar{V}_\Lambda^{-1}/k_\Lambda$, respectively.

The use of independent priors on Λ and Δ as in Montgomery and Rossi (1999) provides an important point of divergence with previous work in marketing research that uses a single joint Wishart prior on these matrices (Blattberg and George 1991, Montgomery 1997). The problem with a single inverted Wishart prior on the variance of β_s and ϵ_s is a lack of flexibility. Once the mean of the distribution is set, the dispersion around this mean is controlled by a single scaling parameter. However, we want a prior that will allow for differential degrees of freedom on how tight the prior should be on β_s and ϵ_s. Specifically in our problem we wish to have a prior that may allow differential amounts of shrinkage across stores and towards the theory. For example, we may wish to have more cross-store shrinkage than shrinkage towards the theory, i.e., $\Lambda > \Delta$.

To illustrate this problem consider Figure 1 which illustrates the inverted Wishart prior for two diagonal elements in the corresponding panels. Once the dispersion is set for the first element, the dispersion for the second element is automatically fixed, as denoted by a solid line. If we wish to loosen up the prior on the first element to increase the amount of shrinkage (there is an inverted relationship), this would also increase the shrinkage of the second element, as denoted by the dashed line. However, we wish to have the ability to tighten up the prior on the first element without altering the second element, i.e., choose the dashed line for the first parameter and the solid line for the second parameter. The introduction of two independent priors allows for this type of differential shrinkage.

Figure 1 about here

Recent work by Barnard et al (2000) on decomposing the prior on the covariance matrix into the standard deviations and correlation matrices can also allow differential shrinkage.

3.1 An Example

To illustrate the framework presented in the previous subsection consider an example with three products. We use our demand model from (3) without promotional variables:

$$\begin{bmatrix} \ln(q_{1ts}) \\ \ln(q_{2ts}) \\ \ln(q_{3ts}) \end{bmatrix} = \begin{bmatrix} \alpha_{1s} \\ \alpha_{2s} \\ \alpha_{3s} \end{bmatrix} + \begin{bmatrix} \mu_{1s} \\ \mu_{2s} \\ \mu_{3s} \end{bmatrix} \ln(x_{ts}/P_{ts}) + \begin{bmatrix} \epsilon_{11,s} & \epsilon_{12,s} & \epsilon_{13,s} \\ \epsilon_{21,s} & \epsilon_{22,s} & \epsilon_{23,s} \\ \epsilon_{31,s} & \epsilon_{32,s} & \epsilon_{33,s} \end{bmatrix} \begin{bmatrix} \ln(p_{1ts}) \\ \ln(p_{2ts}) \\ \ln(p_{3ts}) \end{bmatrix} + \begin{bmatrix} e_{1ts} \\ e_{2ts} \\ e_{3ts} \end{bmatrix}, \quad \boldsymbol{e} \sim N(\mathbf{0}, \Sigma) \tag{22}$$

The hyper-parameters are:

$$\bar{\mu} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \bar{\alpha} = \begin{bmatrix} 8 \\ 7 \\ 6 \end{bmatrix}, \quad \Lambda = .1 I_6 , \quad \Delta = .01 I_9 , \quad \bar{\phi} = -3 , \quad \lambda_{\phi} = 1$$

Suppose the draw for an individual store is:

$$\phi_s = -3 \ , \ \ \mu_s = \begin{bmatrix} 1.4 \\ .8 \\ .8 \end{bmatrix}, \ \ \boldsymbol{w}_s = \begin{bmatrix} .33 \\ .33 \\ .33 \end{bmatrix}, \ \ \mathrm{E}_s = \begin{bmatrix} -2.3 & 1.1 & 1.1 \\ 1.1 & -1.7 & .6 \\ 1.1 & .6 & -1.7 \end{bmatrix}$$

The restricted price elasticity implied by this specific model would be:

$$\phi_s = -3 \ , \ \ \boldsymbol{w}_s = \begin{bmatrix} .33 \\ .33 \\ .33 \end{bmatrix}, \ \ \bar{\mathrm{E}}_s = \begin{bmatrix} -2.3 & 1.1 & 1.1 \\ 1.1 & -1.7 & .6 \\ 1.1 & .6 & -1.7 \end{bmatrix}$$

Notice that this restricted price elasticity reflects the high own-price sensitivity and small cross-price elasticities that is usually observed in empirical work.

The price elasticity estimates for this individual store will be shrunk towards the restricted price elasticity matrix. This contrasts with Blattberg and George (1991) who propose shrinking all own price terms (deflated by relative prices) to a single value. Their structure would result in price terms being shrunk towards:

$$\begin{bmatrix} -2 & .5 & .5 \\ .5 & -2 & .5 \\ .5 & .5 & -2 \end{bmatrix}$$

Notice that Blattberg and George (1991) can be thought of as a special case of our framework. The shrinkage pattern they suggested is the same as ours when market shares and expenditure elasticities are equal. However, market shares are rarely equal and we may expect some brands to benefit from category expenditures more than others (unequal expenditure elasticities). An advantage of our framework is that we can evaluate the shrinkage of the estimates in terms of theoretical properties of our model, and not rely upon empirical justifications. This is an important distinction since it permits evaluation of shrinkage in terms of utility and not ad hoc empirical justifications.

4 Estimation

We rewrite our model in SUR form:

$$\boldsymbol{y}_s \sim N(\ X_s\beta_s\ ,\ \Sigma_s \otimes I_T\ )\ ,\ \ \Sigma_s^{-1} \sim W(\ \nu_\Sigma\ ,\ \bar{V}_\Sigma^{-1}\ ) \tag{27}$$

In this case the s subscript denotes an individual store, and the dimension of the $\boldsymbol{y}_s$ vector is M brands by T weeks. In rewriting the model we have implicitly stacked the vector of observations for each brand on top of one another in the following manner:

$$\boldsymbol{y_s} = \begin{bmatrix} \boldsymbol{q_{1s}} \\ \vdots \\ \boldsymbol{q_{Ms}} \end{bmatrix}, \quad \boldsymbol{q_{is}} = \begin{bmatrix} q_{i1s} \\ \vdots \\ q_{iTs} \end{bmatrix}, \quad X_s = \begin{bmatrix} X_{1s} & & \\ & \ddots & \\ & & X_{Ms} \end{bmatrix}, \quad X_{is} = \begin{bmatrix} 1 & \ln(x_{1s}/P_{1s}) & p_{11s} & \cdots & p_{M1s} & f_{i1s} & d_{i1s} \\ 1 & \ln(x_{2s}/P_{2s}) & p_{12s} & \cdots & p_{M2s} & f_{12s} & d_{i2s} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 1 & \ln(x_{Ts}/P_{ts}) & p_{1Ts} & \cdots & p_{MTs} & f_{1Ts} & d_{iTs} \end{bmatrix} \quad (28)$$

The second stage of our hierarchical model refers to the hyper-distribution from which the vector of parameters for each store is drawn:

$$vec(\mathrm{E}_s) \mid \mu_s, \phi_s \sim N(\, vec(\bar{\mathrm{E}}_s)\, ,\, \Delta\,) \text{ for } s = 1,\cdots,S \quad , \quad \Delta^{-1} \sim W(\, \nu_\Delta\, ,\, V_\Delta^{-1}\,) \quad (29)$$

where the expected price elasticity matrix is the restricted one implied by an additive utility model:

$$\bar{\mathrm{E}}_s = \phi_s\, diag(\mu_s) - \phi_s\, \mu_s\, (\mu_s \odot \boldsymbol{w}_s)' \quad (30)$$

The remaining parameters are drawn from:

$$\beta_s \sim N(\, \bar{\beta}_s\, ,\, \Lambda\,) \text{ for } s = 1,\cdots,S \quad , \quad \Lambda^{-1} \sim W(\, \nu_\Lambda\, ,\, V_\Lambda^{-1}\,) \quad (31)$$

The third stage of our model expresses the prior on the hyper-distribution:

$$\bar{\beta} \sim N(\, \theta\, ,\, V_\theta\,) \quad (32)$$

4.1 Estimation Using the Gibbs Sampler

Our goal is to compute the posterior distribution of the model parameters. The posterior distribution contains all the information from our sample given our distributional assumptions. From the posterior distribution we can compute the means, which are commonly used as point estimates, along with any other measures of the distribution that are of interest. The following data and parameters are supplied by the analyst:

$$X\,,\,Y \quad ; \quad \theta\,,\,V_{\theta}\,,\,\nu_{\Lambda}\,,\,V_{\Lambda}\,,\,\nu_{\Delta}\,,\,V_{\Delta}\,,\,\nu_{\Sigma}\,,\,V_{\Sigma}\,,\,\nu_{\phi}\,,\,V_{\Phi}\,,\,\bar{\bar{\phi}}\,,\,V_{\bar{\phi}} \tag{33}$$

The general procedure for finding the marginal posterior distribution is to compute the joint posterior and then integrate out all parameters except those of interest. In this case the joint distribution of our model can be written as:

$$\begin{aligned} &p(\beta_1,\cdots,\beta_S,\mathrm{E}_1,\cdots,\mathrm{E}_s,\phi_1,\cdots,\phi_S,\Sigma_1,\cdots,\Sigma_S,\bar{\beta},\Lambda,\Delta,\bar{\phi},\lambda_{\phi} \mid \textbf{data, priors}) \propto \\ &\prod_{s=1}^{S} like(\beta_s,\mathrm{E}_s,\Sigma_s \mid \Delta,\Lambda)\; p(\bar{\beta} \mid \beta_1,\cdots,\beta_S,\Lambda)\; p(\Lambda)\; p(\Delta)\; p(\theta)\; p(\bar{\phi})\; p(\lambda_{\phi}) \end{aligned} \tag{34}$$

If we wanted to find the marginal posterior distribution of θ we would need to solve:

$$\begin{aligned} &p(\bar{\beta} \mid \theta, V_{\theta}, \nu_{\Lambda}, V_{\Lambda}, \nu_{\Delta}, V_{\Delta}, \nu_{\Sigma}, V_{\Sigma}, \nu_{\phi}, V_{\phi}, \bar{\bar{\phi}}, V_{\bar{\phi}}) \\ &= \int p(\beta,\epsilon,\Sigma,\phi,\bar{\beta},\Lambda,\Delta,\bar{\phi},\lambda_{\phi})\, d\beta\, d\mathrm{E}\, d\Sigma\, d\phi\, d\bar{\phi}\, d\lambda_{\phi} \end{aligned} \tag{35}$$

The analytic solution to this integral is not known even with natural conjugate priors. To understand the difficulty in solving this integral, we refer the reader to the simpler case of trying to solve a single stage SUR model (Zellner 1971, pp. 240-6) for which the analytic solution is not known either. Therefore we will have to rely upon numerical procedures to find the solution. Unfortunately the high dimension of the integral makes it difficult to find a solution using conventional numerical integration techniques.

An alternate method is through the use the Gibbs sampler. The Gibbs sampler requires the solution of the conditional distributions, which can be easily derived due to the hierarchical structure of the model. For a good introduction to the Gibbs sampler see Casella and George (1992). We do not advocate the use of Gibbs sampler based on computational efficiency, instead we advocate its use because of its ease of implementation. The most desirable solution would be an analytical one, but given that this solution does not exist in closed form we satisfy ourselves with a numerical solution.

The Gibbs sampler employed in this paper requires sequentially randomly sampling from each of the conditional distributions. It has been shown by Gelfand and Smith (1990) and Gelfand et al (1990) that this draws converge to the posterior marginal distributions. The general outline of the procedure is:

1. Select starting values for the parameters of the marginal posterior distributions. In our practice the least squares estimates of these parameters provide good starting points.

2. Generate M_1+M_2 sets of random numbers with each set being drawn in the following manner:

$$\beta_s^{(k)} \veebar p(\,\beta_s \mid \mathbf{E}_s^{(k-1)}, \Sigma_s^{(k-1)}, \cdots) \quad \text{for } s = 1, \cdots, S \tag{36}$$

$$\mathbf{E}_s^{(k)} \veebar p(\,\mathbf{E}_s \mid \beta_s^{(k)}, \Sigma_s^{(k-1)}, \phi_s^{(k-1)}, \cdots) \quad \text{for } s = 1, \cdots, S \tag{37}$$

$$\phi_s^{(k)} \veebar p(\phi_s \mid \beta_s^{(k)}, \mathbf{E}_s^{(k)}, \cdots) \quad \text{for } s = 1, \cdots, S \tag{38}$$

$$\Sigma^{(k)} \veebar p(\Sigma_s \mid \beta_s^{(k)}, \mathbf{E}_s^{(k)}, \cdots) \tag{39}$$

$$\bar{\beta}^{(k)} \veebar p(\,\bar{\beta} \mid \beta_1^{(k)}, \cdots, \beta_S^{(k)}, \Lambda^{(k-1)}, \cdots) \tag{40}$$

$$\bar{\phi}^{(k)} \veebar p(\,\bar{\phi} \mid \phi_1^{(k)}, \cdots, \phi_S^{(k)}, \lambda_\phi^{(k-1)}, \cdots) \tag{41}$$

$$\Delta^{(k)} \veebar p(\Delta \mid \mathbf{E}_1^{(k)}, \cdots, \mathbf{E}_S^{(k)}, \beta_1^{(k)}, \cdots, \beta_S^{(k)}, \phi_1^{(k)}, \cdots, \phi_S^{(k)}, \cdots) \tag{42}$$

$$\Lambda^{(k)} \veebar p(\Lambda \mid \beta_1^{(k)}, \cdots, \beta_S^{(k)}, \bar{\beta}^{(k)}, \cdots) \tag{43}$$

$$\lambda_\phi^{(k)} \veebar p(\lambda_\phi \mid \phi_1^{(k)}, \cdots, \phi_S^{(k)}, \bar{\phi}^{(k)}, \cdots) \tag{44}$$

Where the symbol $x \veebar p(x)$ means that the x is a simulated realization or draw from the density $p(x)$ and k denotes the iteration number. The above conditional distributions are understood to also depend upon the prior parameters and the data.

3. Use the last M_2 sets of draws to estimate the posterior marginal distributions.

This means that the problem reduces to solving the conditional distributions of each of the parameters in the posterior distribution. These solutions are readily available due to the hierarchical structure of our model and the affine nature of the normal and Wishart distributions. The solution of the conditional densities are:

1. Draw the parameter vector in the first-stage in two parts to avoid the nonlinearity induced by the additive separable prior:

    (a) Since we know the price elasticities, we can rewrite the model as below:

$$\left[\ln(q_{its}) - \sum_j \epsilon_{ijs}\ln(p_{jts})\right] = \alpha_{is} + \mu_{is}\ln(x/P_{ts}) + \theta_{is}f_{its} + \phi_{is}d_{its} + e_{its} \tag{45}$$

    The β_s vector can be drawn using the usual SUR result.

    (b) Since we know the β_s vector we can rewrite the model as below:

$$[\ln(q_{its}) - \{\alpha_{is} + \mu_{is}\ln(x/P_{ts}) + \theta_{is}f_{its} + \psi_{is}d_{its}\}] = \sum_j \epsilon_{ijs}\ln(p_{jts}) + e_{its} \tag{46}$$

    The E_s matrix can be drawn using the usual multivariate regression result.

2. Draw the ϕ parameter. Notice that conditional upon $\mathbf{E}_s$ and μ_s we have the following univariate regression:

$$\epsilon_{ijs} = (\delta_{ij} - \mu_{js}w_{js})\mu_{is}\,\phi_s + u_{ijs}\ ,\ \boldsymbol{u}_s \sim N(\mathbf{0},\Delta) \tag{47}$$

    Hence, ϕ_s can be drawn using the usual univariate regression result.

3. Σ_s is drawn from an inverted Wishart distribution

$$\Sigma_s^{-1} \sim W(\nu_\Sigma + T_s\ ,\ (V_\Sigma + \hat{E}_s'\hat{E}_s)^{-1}) \quad , \quad \hat{E}_s[,i] = \boldsymbol{y}_{is} - X_{is}'\boldsymbol{\beta}_{is} \tag{48}$$

4. $\bar{\beta}$ is a multivariate regression

$$\bar{\beta} \sim N(\ H(\Sigma_s \Lambda^{-1}\beta_s + V_\theta^{-1}\theta)\ ,\ H\ ) \quad , \quad H = (S\Lambda^{-1} + V_\theta^{-1})^{-1} \tag{49}$$

    $\bar{\phi}$ is a univariate regression

$$\bar{\phi} \sim N\left(H\left[\frac{\Sigma_s \phi_s}{\lambda_\phi} + \frac{\bar{\bar{\phi}}}{V_{\bar{\phi}}}\right], H \right) \quad , \quad H = \left[\frac{S}{\lambda_\phi} + \frac{1}{V_{\bar{\phi}}}\right]^{-1} \tag{50}$$

5. Since Δ and Λ are independent they can be drawn separately from inverted Wishart distributions:

$$\Lambda^{-1} \sim W(\nu_\Lambda + S\ ,\ V_\Lambda + \textstyle\sum(\beta_s - \bar{\beta}_s))(\beta_s - \bar{\beta}_s))'\ ) \tag{51}$$

$$\Delta^{-1} \sim W(\nu_\Delta + S\ ,\ V_\Delta + \textstyle\sum(\boldsymbol{vec}(\mathrm{E})_s - vec(\bar{\mathrm{E}}_s))(\boldsymbol{vec}(\mathrm{E})_s - vec(\bar{\mathrm{E}}_s))'\ ) \tag{52}$$

$$\lambda_\phi^{-1} \sim W(\nu_\phi + S\ ,\ V_\phi + \textstyle\sum(\phi_s - \bar{\phi})(\phi_s - \bar{\phi})'\ ) \tag{53}$$

5 Application to Scanner Data from the Refrigerated Orange Juice Category

We apply our methods to store level scanner data collected from 83 stores from Dominick's Finer Foods chain in Chicago, IL. This data is collected from point-of-sale computers that record quantity and prices of purchased items. Our data is reported at the weekly level for each store. We have 120 weeks of data which is split for the purposes of model validation into a sample for estimation and another for out-of-sample predictive validation. We consider products in the refrigerated orange juice category. Table 1 lists the items under study, average price and market share. The 11 items represent well over 70% of the revenue in this category and cover the range from premium national brands to lower quality store brands. Our expenditure variable (x) is calculated from a subset of 26 store categories with over 5,000 UPC's. These categories account for over 25% of total store ACV.

Table 1 about here

The settings of the priors are chosen to be relatively uninformative relative to the data except for priors on Δ and Λ. The prior on Δ controls the amount of shrinkage towards the theory, and the prior on Λ controls the amount of shrinkage across the stores. A judicious choice of prior settings on these variables can result in estimates that closely proxy the restricted or unrestricted models, or fulfill our desire to fall somewhere in-between these estimates. We evaluate the impact of the prior over a range of settings.

5.1 How good is our theory?

We remind the user that we can actually think of our model as providing two dimensions of restrictions. The first is to employ the restrictions on the price elasticity matrix implied by an additive utility model as described in section 2. The second is to pool the observations across stores, which would restrict the estimates of one store to be equal to one another. A natural starting point is to perform a classical test to determine whether the restrictions hold exactly. We summarize the number of parameters, in- and out-of-sample MSE, log-likelihood, and Schwarz information criterion (SIC) in Table 2. The restrictions implied by an additive utility model, pooling assumption, or both are all overwhelming rejected ($p<.0001$) by standard likelihood ration tests. Upon an initial evaluation it might appear that the neither the theory nor pooling is helpful. An alternative model selection criterion would be to use SIC as an asymptotic argument to justify the choice of models. Using the Schwarz information criterion (SIC) would lead to the choice of restricted store-level models. The out-of-sample predictions imply that the parameter bias induced by the restricted store model is well worth the reduced variance of the parameter estimates.

Table 2 about here

Table 2 clearly shows that, either in terms of in-sample or out-of-sample fit, pooled models are inferior to more unrestricted models. This is because of the large heterogeneity in this population of stores. It is important to note that the out-of-sample validation results indicate that this is not just the result of over-fitting. The next most important conclusion is that the restrictions of the additive utility theory are useful in improving predictive accuracy. The Bayes model performs the best in out-of-sample predictive validation and offers the flexibility of store level models without the dangers of over-parameterization. In this data set, it appears that the restrictions of additive utility theory hold fairly well. In addition, there are large and detectable store differences so that the Bayes model adapts to something fairly close to the restricted store models. A more formal measure to determine the best model is to compute the posterior odds of our Bayesian models. We follow Newton and Raftery's (1994) technique to compute the posterior odds and we find overwhelming support that a model with a strong prior on the theory and weak prior on commonalities across stores has the highest posterior probability.

5.2 Impact on price elasticity estimates

In Table 3 we illustrate the similarity and differences in the point estimates for the expenditure and price elasticities of four selected products. First, note the wide variation in the magnitude of the unrestricted store models. A common complaint amongst analysts is that a large number of elasticities may be incorrectly signed and even the magnitudes may be suspect. Notice four of the twelve parameters have unexpected signs and the magnitudes of the own-price elasticities vary widely from -2.2 to -3.7, given the similarity of the products we might expect more similar estimates. In contrast the restricted pooled model which implements pooling across the stores and the exact restrictions as prescribed by an additive utility model eliminates both of these criticisms. However, we have lost all heterogeneity in the estimates across the stores and the theoretical restrictions are rigidly enforced. Both of these assumptions are rejected by standard statistical tests. The estimates from the Bayes model offer a compromise solution in which the only on of the cross-price elasticity is incorrectly signed, and the range of the elasticities are reduced. A judicious choice of our prior can result in estimates that can mimic these restricted estimates, or result in estimates that fall in between these estimates. Again we note that the data provides strong support that a compromise solution is superior both in the form of improved out-of-sample predictions and high posterior odds.

Table 3 about here

6 Incorporating Information about Market Structure

The analysis of the previous subsection which uses a prior based upon an additive prior may seem overly restrictive. One concern is that a category may have two subcategories that are only weakly related.

For example, the laundry detergent category may consist of liquid and powder forms. Substitution within a subcategory may be high, but between these subcategories it may be weak. To allow increased flexibility we consider combining the strong and weak separability arguments from section 2 into a single model. If we assume that utility is additive or strongly separable within a category but weakly separable across categories, then utility can take the following structure:

$$utility = \nu_{\left(\sum_i \nu_{1i}(q_{1i}),\ \sum_i \nu_{2i}(q_{2i}),\ \cdots,\ \sum_i \nu_{Ci}(q_{Ci})\right)} \tag{54}$$

where q_{ci} is the quantity of the ith product in the cth category. This will result in the following restrictions on the price elasticities:

$$\epsilon_{ijs} = \begin{cases} \phi_{GG}\,\mu_{is} - \phi_{GG}\,\mu_{is}\,\mu_{js}\,w_{js} & \text{if } i \in G \text{ and } j \in H,\ i \neq j \\ \phi_{GG}\,\mu_{is}\,\mu_{js}\,w_{js} & \text{if } i \in G \text{ and } j \in G,\ i = j \\ \phi_{GH}\,\mu_{is}\,\mu_{js}\,w_{js} & \text{if } i \in G \text{ and } j \in H \end{cases} \tag{55}$$

Notice one change from our previous formulation is that we have dropped the store subscript on ϕ. This change is necessitated by the increased computational requirements of the model. However, we believe this is a sensible restriction, since the ϕ's permit differences in market structures and we presume that the market structure in each store is the same.

This structure permits more flexibility in the price elasticity matrix, but still is a fairly parsimonious structure, perhaps overly so for many analysts. If $\phi_{GH}=\phi$ for all G and H then (55) will reduce to the restrictions induced by an additive utility structure in (13). While these structures can be similar, our hope is that by incorporating models that are closer to the true generating process of the data this should result in better approximations and shrinkage patterns. On the other hand, the added flexibility may not be necessary since the model already permits substantial departures from the theory embedded within the prior.

This type of structure has been considered previously in marketing in the context of market structures. Allenby (1989) proposed identifying market structures using a restricted additive utility model—albeit in nested logit form. If we assume that the expenditure elasticities within a market segment are constant, we can derive the same market structure proposed by Allenby. As an illustration suppose there are two submarkets each with 3 brands. The uncompensated elasticity matrix will be:

$$
H = \left[\begin{array}{ccc|ccc}
\eta_1 & \xi_{aa}w_2 & \xi_{aa}w_3 & \xi_{ab}w_4 & \xi_{ab}w_5 & \xi_{ab}w_6 \\
\xi_{aa}w_1 & \eta_2 & \xi_{aa}w_3 & \xi_{ab}w_4 & \xi_{ab}w_5 & \xi_{ab}w_6 \\
\xi_{aa}w_1 & \xi_{aa}w_2 & \eta_3 & \xi_{ab}w_4 & \xi_{ab}w_5 & \xi_{ab}w_6 \\
\hline
\xi_{ba}w_1 & \xi_{ba}w_2 & \xi_{ba}w_3 & \eta_4 & \xi_{bb}w_5 & \xi_{bb}w_6 \\
\xi_{ba}w_1 & \xi_{ba}w_2 & \xi_{ba}w_3 & \xi_{bb}w_4 & \eta_5 & \xi_{bb}w_6 \\
\xi_{ba}w_1 & \xi_{ba}w_2 & \xi_{ba}w_3 & \xi_{bb}w_4 & \xi_{bb}w_5 & \eta_6
\end{array}\right] \tag{56}
$$

Where $\eta_i=\xi_{ab}w_i-\mu_i\phi$, $\xi_{ab}=-\phi\mu_i\mu_j-\mu_i$, a and b denote the submarket for products i and j. The restricted elasticity matrix of (56) is the same as that given in Allenby's (1989) figure 1.

6.1 Constructing a Prior on Market Structure

The first step in constructing a Bayesian model is to develop a prior assessment of the probability for each market structure. For example if we have a category with three products: A, B, and C, then there are five possible market structures: {(A,B,C)}, {(A,B),(C)}, {(A),(B,C)}, {(A,C),(B)}, {(A),(B),(C)}. The most direct solution would be to assume a particular market structure and simply replace the $\bar{\epsilon}$ used in (18), which was based upon an additive utility model with the model proposed in (55). In keeping with the theme of this paper we would like to allow some uncertainty about the market structure and allow deviations away from this market structure. Our prior must attach a probability to each of these possible market structures. As the number of products increases there is a combinatorial explosion of possible market structures, perhaps allowing millions of models. Computationally it is not possible to compute the posterior distribution if all these markets must be considered as would happen with a flat prior. Therefore theory or some expertise must be used to guide in identifying likely market structures. If we are totally agnostic then we will not be able to find a solution.

One technique used by Allenby (1989) is to simply enumerate category structures based upon the product attributes, like brand, size, flavor, etc. For example, choose a market structure induced by blocking all brands with the same size together. This technique results in a small number of market structures. Unfortunately, grouping upon individual attributes alone may not be satisfactory. We would like to propose a more flexible approach, that allows grouping based upon multiple attributes, say size and quality.

Additionally, we would like to permit some deviations away from this structure. For example, one product that has the same size as those products in one subcategory should be placed with another subcategory due its similarity on quality.

We use a conditional approach to specify our prior that a product belongs to a subcategory. The conditional approach assumes that we know the assignments of the N-1 other items in the category and are interested in assigning one additional product. Our problem becomes one of predicting the probability that this unassigned product should be assigned to a new k+1 subcategory or one of the existing k subcategories. This conditional specification makes it easy to incorporate it into our Gibbs sampling algorithm. The marginal probabilities of each model can be computed using simulation.

We begin by considering the probability that a new category should be created. We would like this probability to reflect the similarity of the existing groups. If the existing subcategories are quite similar then they will offer low discriminatory value, and we would argue that it is likely that a new category should be opened. On the other hand, if the unassigned product has a high probability of belonging to one subcategory versus the others then this indicates a high discriminatory power of the existing structure, and we would argue that it is less likely that a new category should be created. Additionally, as more categories are created we wish to decrease the probability that a new category should be opened. Another function of this conditional probability is to serve as a penalty function and avoid creating too many subcategories, which would result in an overparameterized model.

Suppose there are k existing subcategories, and the conditional probability that a new product is assigned to subcategory g is p_g and the probability that it is assigned to a new subcategory is p_{k+1}. We begin by defining the probability of opening a new subcategory:

$$p_{k+1} = \frac{1}{\gamma} \exp\{ - \delta\, o \} \tag{57}$$

where γ is a parameter that scales the overall probability and is positive, δ is a function of the number of categories that currently existing, and o is the entropy of the current subcategory classification probabilities. We define entropy as follows:

$$o = - \sum_{g=1}^{k} p_g \log_2(p_g) \tag{58}$$

Entropy is a measure of how much disparity there is in the attributes of the existing categories. If all the probabilities are (p_g) are close then entropy is low, as the probabilities diverge entropy increases. Notice that entropy is always positive. Additionally, the scaling function of entropy (δ) is defined as follows:

$$\delta = \delta_1 \left(1 + \frac{k}{\delta_2} \right) \tag{59}$$

where δ_1 and δ_2 are scaling parameters and are positive. δ_1 scales the entropy, and δ_2 increases this penalty as the number of existing categories grows.

In constructing the probability that an unassigned item belongs to an existing category we wish to reflect the similarity of the unassigned product with the existing categories. If an attribute of an unassigned product matches those in an existing category then it is likely that this product belongs to this category. We begin by defining the probability that given attribute i the unassigned product belongs to category g:

$$p_{gi} = \frac{c_g + \omega}{n_g + \omega} \tag{60}$$

Where c_g is the number of products within subcategory g that have the same ith attribute and n_g is the total number of products in the subcategory. The role of the parameter ω is to prevent zero probabilities. If we assume that the M attributes of a product are independent of one another, then the probability that the new product belongs to gth group is proportional to:

$$p_g \propto \prod_i^M p_{gi} \tag{61}$$

It might seem like an independence assumption may be questionable, but since highly correlated attributes can be omitted independence may be a reasonable assumption.

One further extension that we wish to incorporate is to place additional weight on one particular attribute. We modify (61) by raising the probability of the correspond attribute by τ and raising the other attributes by $1/\tau$. In our problem a priori we are uncertain as to which attribute will be more important, therefore we consider a mixture prior in which an attribute has an equal probability of being the important attribute. In summary our model is:

$$p_g = (1 - p_{k+1}) \frac{\sum_{j=1}^{M} \prod_{i=1}^{M} p_{gi}^{\kappa_{ij}}}{\sum_{l=1}^{k} \sum_{j=1}^{M} \prod_{i=1}^{M} p_{li}^{\kappa_{ij}}} \;, \quad \kappa_{ij} = \begin{cases} \tau & if\, i = j \\ 1/\tau & otherwise \end{cases} \tag{62}$$

Where (1-p_{k+1}) reflects the probability that a new category is not created or one of the existing categories is selected.

Example: Consider the following example to illustrate this prior. Our problem is to determine whether the eleventh product, Minute Maid - Regular - 96 Oz, should be assigned to subcategory A, B, C, or a new subcategory D given the assignments of the other ten products as listed in Table 4. Notice subcategory A appears to be premium products, B is made up of regular products of various brands and sizes, while C is made up of store brands. We set the parameters of this prior as follows: ω=.001, δ_1=.25, δ_2=10, τ=2, and γ=100. The results indicate that there is a 99% probability that Minute Maid - Regular - 96 Oz. should be assigned to subcategory B, a 1% chance that it should be assigned to a new category, and a negligible probability of being assigned to subcategory A or C. This conforms well with our intuition that subcategory B is made up of various national brands. Intuitively the prior strongly predicts that the product belongs to subcategory B because the quality attribute matches perfectly and there one match in the brand category, otherwise category C would have been highly favored. As the ω parameter is increased to .3 the odds of the product being assigned to subcategory C rise significantly to 41%, subcategory B's probability drops to 58%, and the odds of a new category drop to .7%. If the ω parameter is set to zero then unless there is at least one match of the unassigned attribute to the products in the subcategory there is no probability of the unassigned attribute being assigned to that subcategory.

Insert Table 4 about here.

For the 11 products listed in Table 4 there are almost 40 million possible permutations of market structures. However, many of these permutations result in structures that are essentially the same except for the labeling of the subcategories. For example, the market structure {(A,B),(C)} is the same as {(C),(A,B)}.

To insure the identifiability of the market structures we only allow those structures in which the lowest product rank as given in Table 1 is less than those of the subcategories that follow it. In the previous example, the permutation {(C),(A,B)} would not be allowed. This identifiability condition results in about 500,000 possible market structures.

We simulate our prior using 100,000 iterations, and list the parameter settings and the number of subcategory structures identified in Table 5[3]. Setting 1 favors those category structures that allow more subcategories and includes the extreme case that all products are assigned to different subcategories. Settings 2 through 6 include most of the usual candidate structures that are blocked by attributes: brand, size, quality, and all products in the same category. These priors tend to result in subcategories that have more products and result in those subcategories that have similar attributes.

Insert Table 5 about here.

In our subsequent analysis we use the prior that corresponds with setting 6. To acquaint the reader with the types of structures that this model identifies we list the top ten models along with their prior probability in Table 6. A priori the best market structure is the one in which there are two subcategories. One with the store brands (which match on brand and quality) and all others. Many models are slight deviates from one another, in which one product will switch to a different subcategory. These top ten models account for 59% of the probability in the prior. The market structure in which all items are assigned to the same category was ranked 15th.

Insert Table 6 about here.

To better demonstrate the association of the products using our prior we compute the conditional probability that each pair of products will be included in the same subcategory in Table 7.

---

3. If a market structure does not occur in the simulation we assume that its probability is zero. Effectively, we are truncating our prior.

Insert Table 7 about here.

We readily acknowledge that this prior is only one out of the multitudes that could be constructed. For example, we could imagine using a flat prior, and enumerate all possible models and allow each to have an equal probability of being selected. However, this is computationally infeasible. Another suggestion would be to simply count the number of categories and place a prior that would penalize models based upon the number of parameters. This may result in a prior that yields a penalty function that is that same as the Schwarz information criterion. The benefit of our prior is that it uses brand attribute information and results in model structures that seem plausible without eliminating too many combinations. We conducted many tests of the sensitivity of the prior and found that the information from the likelihood function tends to dominate the information in the prior. Therefore, the basic function of the prior is simply to identify which models are considered, so the censoring property of the prior is its most critical function (i.e., most market structures have zero probability).

6.2 Estimating the model

To estimate this model we can create a Gibbs sampler to simulate draws from the marginal posterior distribution. The estimation structure we proposed in section 4 can be readily adapted to this new structure. We divide the sampler into two components. The first is to simulate the model conditional upon the market structure. The second component is to simulate the market structure conditional upon the parameter estimates. Since this first component is similar to the algorithm described in section 4, we will not discuss it in depth. The critical difference is that the mean of the hyper-distribution is based on the restrictions given weak separability across the subcategories as described in (55) and not the restrictions implied by an additive utility model as given in (13). Again our intent is to allow some variation around the restricted model, but induce strong shrinkage towards the theoretical restrictions.

A new component of our algorithm is to simulate the market structure conditional upon the parameter values. The motivation is to randomly select one of the products, compute the probability that it should be remain in the same subcategory, be reassigned to another subcategory, or a new subcategory created. These probabilities form a multinomial distribution from which we simulate a value and reassign the product to the appropriate subcategory and then repeat the first part of the process again which re-estimates all the parameters conditional upon the market structure.

To illustrate this algorithm, suppose that we have four products: A, B, C, and D. At iteration *i* the market structure is {(A),(B),(C,D)}, and we wish to re-evaluate the assignment of product A. We need to compute the probability of the following models: {(A),(B),(C,D)}, {(A,B),(C,D)}, and {(B),(A,C,D)}. In other words, what is the chance of no change (i.e., product A staying as a separate subcategory) or product A being merged with one of the existing subcategories. The market assignment of product A at iteration *i* is defined as M_i. In our example M_i can take on one of three values: {(A),(B),(C,D)}, {(A,B),(C,D)}; and {(B),(A,C,D)}. Our problem is to compute the posterior probability of M_i:

$$p(M_i \mid \Theta) \propto p(\Theta \mid M_i)\, p(M_i) \tag{63}$$

where Θ is the set of all parameters in the model to be estimated, $p(\Theta|M_i)$ can be computed from the likelihood function given the market structure, and the prior $p(M_i)$ is the prior probability as defined in the previous subsection. Equation (63) will take on a multinomial distribution which can be sampled easily.

Table 8 about here

We apply this estimation procedure to the same dataset described in section 5. We evaluate the posterior using three different priors on the amount of shrinkage towards the theory, as captured by Δ, that should be done: strong (ν_Δ=dim(Δ)+3+5**S*, V_Δ=.000025), moderate (ν_Δ=dim(Δ)+3, V_Δ=.01), and weak (ν_Δ=dim(Δ)+3, V_Δ=.25). The purpose is to guage the sensitivity of the posterior to this prior specification. Table 8 provides the out-of-sample forecasting results. The moderate prior performs the best, but all the priors have superior out-of-sample forecasting results compared with the unrestricted models. In comparison to the market structure restricted models the predictive results are similar. However, there are substantial differences in the price elasticity estimates induced by the differences in market structures. Tables 9 through 11 provide the posterior probability of the top ten market structures for the strong, moderate, and weak priors. The most likely market structure in the strong prior contains the 64 ounce cartons, 96 ounce cartons, Tree Fresh, and the store brands. Again it is unlikely that a priori an analyst would have guessed such a structure since this classification cannot be derived from a single attribute. The only question seems to be whether the subcategory with the 96 ounce cartons should be split. As the prior on Δ is weakened the posterior distribution becomes more diffuse and it is more difficult to identify a single market structure. This is quite important since it suggests that if the analyst is unwilling to be aggressive in stating his beliefs that the theory is correct, relying upon the data using a pre-testing method will lead to biased market structure estimates, and hence price elasticity estimates. Regardless of the analyst's beliefs the data has quite a bit of information and can move the prior on the market structures significantly even with a weak prior as the posterior probabilities that pairs of products will be assigned in the same subcategory shows in Table 12. Table 12 can be contrasted with the prior probabilities given in Table 7.

Tables 9, 10, 11, and 12 about here

7 Conclusions

We have shown how economic theory can be incorporated into estimators of consumer demand. Our purpose is to represent the notion that a theory is approximately correct. Our estimates can be described as shrinking the unrestricted model estimates without the theory towards the restricted estimates implied by the theory. The amount of shrinkage is adaptive and modified by both an analyst's prior beliefs and the amount of support the data has for the theory. Classical approaches to estimating demand by first pre-testing the adequacy of the theory and then proceeding conditionally upon these estimates will bias the estimates. This will either lead to overconfidence in the estimates when the theory is accepted or underconfidence when the theory is disregarded. An important facet of our shrinkage estimates is that the theory can contribute information even when it is rejected by classical testing procedures, since the theory may be approximately correct. Another benefit of our approach is that it provides the analyst a method for understanding the impact of theoretical assumptions on parameter estimates by varying the degree of confidence in the prior. While we have illustrated our technique using logarithmic demand models, this approach can be applied to any functional form, such as an AIDS or Rotterdam model. Additionally, we hope that this research will encourage applications of Bayes methods to other problems like the estimation of supply and production functions.

References

Allenby, G.M. (1989), "A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data", *Marketing Science*, 8, 265-280.

Allenby, G.M. and P.E. Rossi (1993), "A Bayesian approach to estimating household parameters", *Journal of Marketing Research*, 30, 171-182.

Barten, A.P. (1969), "Maximum likelihood estimation of a complete system of demand equations", *European Economic Review*, 1: 7-73.

Barnard, J., R.E. McCulloch, and X-L Meng (2000), "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage", *Statistica Sinica*, 10 (4): 1281-1311.

Berndt, E. and A. Silk (1993), "Consistency Requirements and the Specification of Asymmetric Attraction Models of Aggregate Market Share", Working Paper, MIT.

Blattberg, R.C. and E. George (1991), "Shrinkage Estimation of Price and Promotional Elasticities: Seemingly Unrelated Equations", *Journal of the American Statistical Association*, 86, 304-315.

Blattberg, R.C. and S.A. Neslin (1990), *Sales promotion: concepts, methods, and strategies*, Englewood Cliffs, N.J.: Prentice Hall.

Blattberg, R.C. and K. Wisniewski (1989), "Price-Induced Patterns of Competition", *Marketing Science*, 8, 291-309.

Casella, G. and E. George (1992), "Explaining the Gibbs Sampler", *American Statistician* 46, 167-174.

Deaton, A.S. (1974), "The analysis of consumer demand in the United Kingdom, 1900-1970", *Econometrica*, 42: 341-67.

Deaton, A.S. (1978), "Specification and testing in applied demand analysis", *Economic Journal*, 88: 526-36.

Deaton, A.S. and J. Muellbauer (1983), *Economics and Consumer Behavior*, Cambridge: Cambridge University Press.

Gelfand, A., S. Hills, A. Racine-Poon, and A. Smith (1990), "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling", *Journal of the American Statistical Association*, 85, 972-985.

Gelfand, A. and A. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, 85, 398-409.

Green, P.E. and V.R. Rao (1971), "Conjoint Measurement for Quantifying Judgmental Data", *Journal of Marketing Research*, 8 (3), 355-363.

Guadagni, P.M. and J.D.C. Little, (1983), "A Logit model of brand choice calibrated on scanner Data", *Marketing Science*, 2 (Summer), 203-38.

Kamakura, W. and G. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure", *Journal of Marketing Research* 26, 379-90.

Lancaster, K. J. (1966), “A new approach to consumer theory”, *Journal of Political Economy*, 10 (2), part 2, 1-92.

Lindley, D.V. and A.F.M. Smith (1972), “Bayes Estimates for Linear Model”, *Journal of the Royal Statistical Society: Series B-Statistical Methodology*, 34 (1): 1-&.

Montgomery, A.L. (1997), “Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data”, *Marketing Science*, 16 (4), 315-337.

Montgomery, A.L. and P.E. Rossi (1999), “Estimating Price Elasticities with Theory-Based Priors”, *Journal of Marketing Research*, 36, 413-423.

Newton, M. and A. Raftery (1994), “Approximate Bayesian Inference with the weighted Likelihood Bootstrap”, *Journal of the Royal Statistical Society: Series B*, 56, 3-48.

Rossi, P.E. and G.M. Allenby (2000), “Statistics and Marketing”, *Journal of the American Statistical Association*, 95, 635-638.

Srivastava, R.K., R.P. Leone, and A.D. Shocker (1981), “Market Structure Analysis: Hierarchical Clustering of Products Based on Substitution-In-Use”, *Journal of Marketing*, 45, 38-48.

Swamy, P.A.V.B. (1970), “Efficient Inference in a Random Coefficient Regression Model”, *Econometrica*, 38 (2): 311-&.

Theil, H. (1976), *Theory and Measurement of Consumer Demand*. Amsterdam: North-Holland.

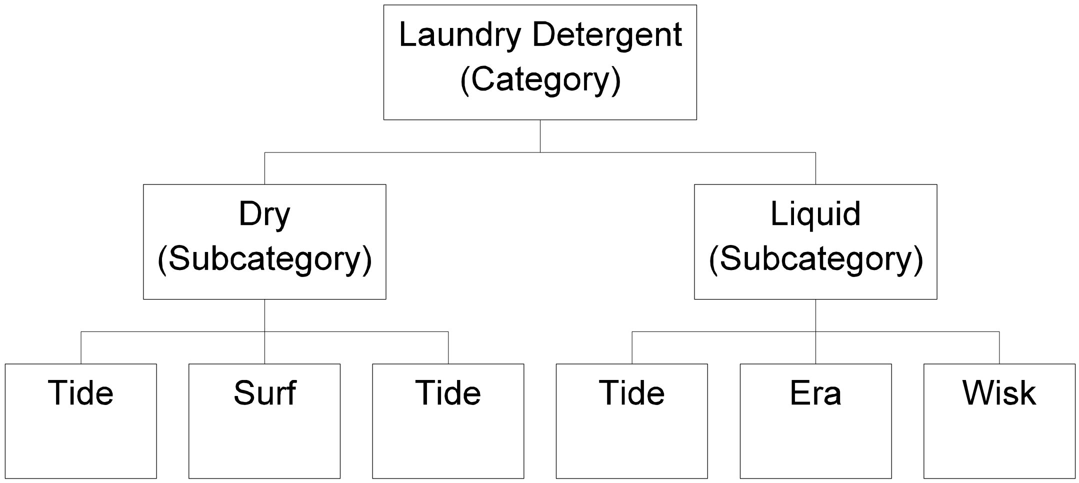

Figure 1. Example of a hierarchical market structure for laundry detergent with dry and liquid subcategories.

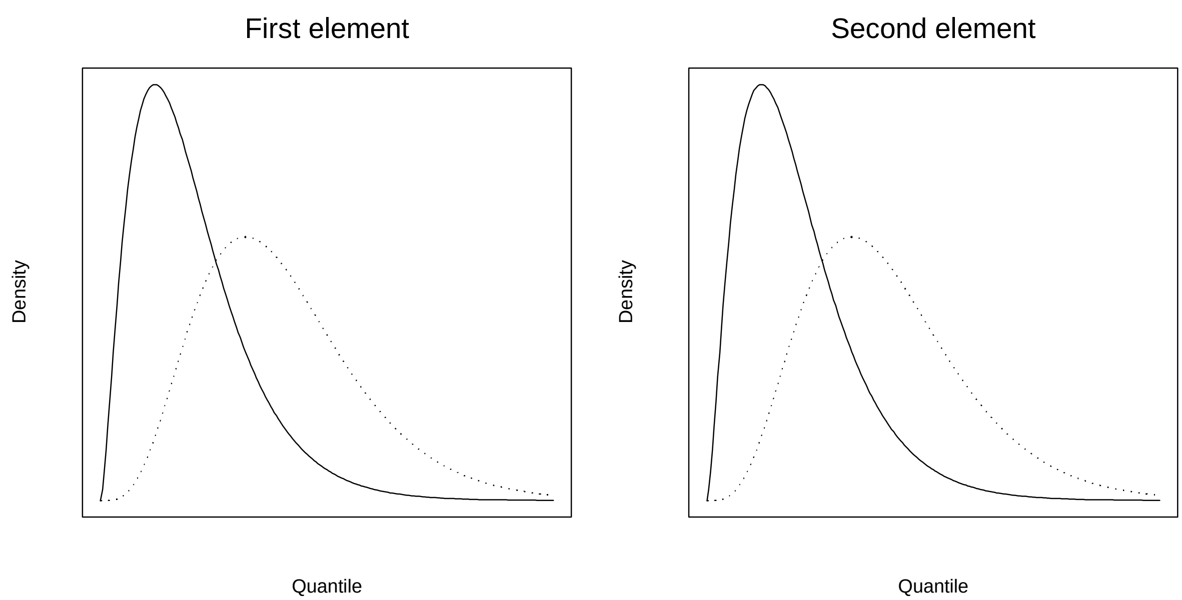

Figure 2. Two selected elements of an inverted Wishart distribution.

Item	Abbreviation	Average Price	Market Share
Tropicana Premium 64	TropP64	2.87	16.1
Tropicana Premium 96	TropP96	3.12	10.7
Florida's Natural 64	FNat64	2.86	4.0
Tropicana 64	Trop64	2.27	15.8
Minute Maid 64	MMaid64	2.24	16.9
Minute Maid 96	MMaid96	2.68	5.7
Citrus Hill 64	CHill64	2.32	5.1
Tree Fresh 64	TFresh64	2.18	2.5
Florida Gold 64	FGold64	2.07	2.6
Dominick's 64	Dom64	1.74	13.6
Dominick's 128	Dom128	1.83	6.9

Table 1. Listing of the items used in the study, along with their average price and market share.

Approach	Model	Number of parameters	Log-likelihood	SIC	MSE	Predictive MSE
Classical	Unrestricted store	12,865	49560.9	40778.2	.170	.379
	Restricted store	2,905	24798.1	-18005.8	.247	.318
	Unrestricted pooled	155	20850.9	-40016.3	.314	.385
	Restricted pooled	35	11598.5	-22816.4	.358	.402
	Shrinkage across Stores		*Shrinkage Towards Additive Utility Restrictions*			
Bayes	Strong (k_Λ=.0001)		Strong (k_Δ=.0001)		.251	.318
	Strong (k_Λ=.0001)		Weak (k_Δ=10000)		.209	.301
	Weak (k_Λ=4900)		Strong (k_Δ=.0001)		.182	.337
	Weak (k_Λ=4900)		Weak (k_Δ=10000)		.177	.350
	Moderate (k_Λ=1)		Moderate (k_Δ=1)		.214	.292

Table 2. Comparison of various estimators in terms of number of parameters, log-likelihood, Schwarz information criterion (SIC), and in-sample and out-of-sample MSE estimates. The Bayes estimates for several prior settings that range between weak and moderate settings of the priors that control shrinkage across stores and towards the restrictions of the additive utility model are provided.

Description	Product	Expenditure Elasticity Estimate	Cross-Price Elasticity Matrix Estimates			
			TropP64	TropR64	CHill64	Dom64
Unrestricted Store Model	TropP64	1.1	-2.2	.2	.2	.0
	Trop64	1.7	-.4	-3.7	.6	-.2
	CHill64	.9	.2	.1	-3.1	-.2
	Dom64	1.2	.8	1.5	-.4	-2.3
Restricted Pooled Model	TropP64	1.1	-3.1	.4	.1	.3
	Trop64	1.0	.4	-2.8	.1	.3
	CHill64	1.0	.4	.4	-3.0	.3
	Dom64	1.0	.4	.4	.1	-2.8
Bayes Model	TropP64	1.1	-2.1	.1	.1	.0
	TropR64	1.6	.6	-3.2	.7	.1
	CHill64	1.3	.4	.4	-2.6	-.2
	Dom64	1.0	.4	1.2	.2	-2.3

Table 3. Expenditure and cross-price elasticity estimates for selected products using various estimators.

Product	Brand	Quality	Size	Subcategory
1	Tropicana	Premium	64	A
2	Tropicana	Premium	96	
3	Florida Natural	Premium	64	
4	Minute Maid	Regular	64	B
5	Tropicana	Regular	64	
6	Florida Gold	Regular	64	
7	Citrus Hill	Regular	64	
8	Tree Fresh	Regular	64	
9	Dominicks	Regular	64	C
10	Dominicks	Regular	128	
11	Minute	Regular	96	?

Table 4. The attributes of products and an sample market structure and the product assignments to each subcategory.

Setting	Parameter Values γ	ω	δ	Number of Market Structures
1	.6	.005	2	3,120
2	20	.005	2	16,192
3	20	.005	1	6,669
4	20	.500	5	54,756
5	100	.005	2	6,168
6	100	.001	2	3,662

Table 5. Number of market structures generated by various settings of the prior.

Rank	Subcategory Product Assignments	Prior Probability
1	{TropP64,TropP96,FNat64,Trop64,MMaid64,MMaid96,CHill64,TFrsh64,FGold64}{Dom64,Dom128}	27%
2	{TropP64,TropP96,FNat64,Trop64,FGold64} {MMaid64,MMaid96,CHill64,TFrsh64} {Dom64,Dom128}	11%
3	{TropP64,TropP96,FNat64,Trop64,MMaid64,MMaid96,FGold64} {CHill64,TFrsh64} {Dom64,Dom128}	5%
4	{TropP64,TropP96,FNat64,Trop64,CHill64,FGold64} {MMaid64,MMaid96,TFrsh64} {Dom64,Dom128}	4%
5	{TropP64,TropP96,FNat64,Trop64,TFrsh64,FGold64} {MMaid64,MMaid96,CHill64} {Dom64,Dom128}	3.6%
6	{TropP64,TropP96,FNat64} {Trop64,MMaid64,MMaid96,CHill64,TFrsh64,FGold64} {Dom64,Dom128}	2.7%
7	{TropP64,TropP96,Trop64,MMaid64,MMaid96} {FNat64,CHill64,TFrsh64,FGold64} {Dom64,Dom128}	2.5%
8	{TropP64,TropP96,FNat64,Trop64,CHill64,TFrsh64,FGold64} {MMaid64,MMaid96} {Dom64,Dom128}	1.7%
9	{TropP64,TropP96,FNat64,Trop64} {MMaid64,MMaid96,CHill64,TFrsh64,FGold64} {Dom64,Dom128}	1.5%
10	{TropP64,TropP96,FNat64,Trop64,CHill64,TFrsh64,FGold64} {MMaid64,MMaid96,Dom64,Dom128}	1.3%

Table 6. Top ten market structures and their probabilities according to the prior.

	Trop P64	Trop P96	Fnat 64	Trop 64	MMai d64	MMai d96	CHill 64	TFrsh 64	FGol d64	Dom 64	Dom 128
TropP64	1.00	.95	.86	.85	.47	.49	.45	.45	.74	.05	.05
TropP96		1.00	.82	.82	.47	.50	.43	.43	.70	.06	.06
FNat64			1.00	.77	.45	.45	.50	.50	.86	.03	.03
Trop64				1.00	.55	.56	.55	.54	.78	.03	.03
MMaid64					1.00	.94	.67	.67	.51	.07	.07
MMaid96						1.00	.64	.64	.50	.08	.08
CHill64							1.00	.75	.59	.05	.04
TFrsh64								1.00	.59	.05	.05
FGol64									1.00	.03	.03
Dom64										1.00	.99
Dom128											1.00

Table 7. Prior probability that a pair of products will be assigned to the same subcategory. The product abbreviations are given in Table 1.

Description	MSE	Predictive MSE
Strong (ν_Δ=dim(Δ)+3+5*S, V_Δ=.000025)	.268	.330
Moderate (ν_Δ=dim(Δ)+3, V_Δ=.01)	.211	.320
Weak (ν_Δ=dim(Δ)+3, V_Δ=.25)	.213	.352

Table 8. Comparison of various prior settings for the Bayes model described in Section 6 in terms of in-sample and out-of-sample MSE estimates. The historical period is different than the previous example, and has in-sample MSE of .164 and predictive MSE of .395.

Rank	Subcategory Product Assignments	Prior Probability
1	{TropP64,FNat64,Trop64,MMaid64,CHill64,FGold64} {TropP96,MMaid96} {TFrsh64} {Dom64,Dom128}	55%
2	{TropP64,FNat64,Trop64,MMaid64,CHill64,FGold64} {TropP96} {MMaid96} {TFrsh64} {Dom64,Dom128}	45%

Table 9. Top ten market structures and their posterior probabilities estimating using a strong prior.

Rank	Subcategory Product Assignments	Prior Probability
1	{TropP64,Trop64,MMaid96,TFrsh64} {TropP96,Dom64,Dom128} {FNat64,MMaid64,CHill64} {FGold64}	1.6%
2	{TropP64,TropP96,FNat64,MMaid64,MMaid96,CHill64,Dom64,Dom128} {Trop64,TFrsh64,FGold64}	.6%
3	{TropP64,FNat64,MMaid64,TFrsh64,FGold64} {TropP96,Trop64,MMaid96,CHill64,Dom64} {Dom128}	.6%
4	{TropP64,TropP96,TFrsh64,FGold64} {FNat64,Trop64,MMaid64,MMaid96,CHill64} {Dom64,Dom128}	.6%
5	{TropP64,TropP96,FNat64,FGold64,Dom64} {Trop64,MMaid64,CHill64,TFrsh64} {MMaid96,Dom128}	.6%
6	{TropP64,TropP96,MMaid64,MMaid96,TFrsh64} {FNat64,FGold64} {Trop64,CHill64} {Dom64,Dom128}	.5%
7	{TropP64,FNat64} {TropP96} {Trop64} {MMaid64,MMaid96} {CHill64,TFrsh64,FGold64} {Dom64,Dom128}	.5%
8	{TropP64,TropP96} {FNat64,FGold64} {Trop64,MMaid64,TFrsh64} {MMaid96} {CHill64} {Dom64,Dom128}	.5%
9	{TropP64,TropP96,FNat64,Trop64,MMaid64,MMaid96,CHill64,Dom64,Dom128} {TFrsh64,FGold64}	.5%
10	{TropP64,MMaid64,MMaid96,CHill64} {TropP96,FNat64} {Trop64,TFrsh64,FGold64} {Dom64,Dom128}	.5%

Table 10. Top ten market structures and their posterior probabilities estimating using a moderate prior.

Rank	Subcategory Product Assignments	Prior Probability
1	{TropP64,TropP96} {FNat64,FGold64} {Trop64,CHill64} {MMaid64,MMaid96,TFrsh64} {Dom64} {Dom128}	1.8%
2	{TropP64} {TropP96,Trop64} {FNat64,CHill64,FGold64,Dom64} {MMaid64,TFrsh64} {MMaid96,Dom128}	1.1%
3	{TropP64,TropP96} {FNat64,TFrsh64} {Trop64,MMaid64,CHill64,FGold64} {MMaid96,Dom128} {Dom64}	1.1%
4	{TropP64,TropP96,MMaid96,TFrsh64,Dom64,Dom128} {FNat64,Trop64,MMaid64,FGold64} {CHill64}	.8%
5	{TropP64,TropP96,CHill64,FGold64} {FNat64} {Trop64} {MMaid64,MMaid96,TFrsh64} {Dom64,Dom128}	.6%
6	{TropP64} {TropP96,Trop64} {FNat64,CHill64,FGold64,Dom64,Dom128} {MMaid64,TFrsh64} {MMaid96}	.6%
7	{TropP64,TropP96,FNat64,MMaid64,MMaid96,TFrsh64,FGold64} {Trop64,CHill64,Dom64,Dom128}	.6%
8	{TropP64,Dom64,Dom128} {TropP96,MMaid64,MMaid96} {FNat64} {Trop64,CHill64,TFrsh64,FGold64}	.5%
9	{TropP64,Trop64,CHill64} {TropP96,FNat64,MMaid64,MMaid96,FGold64} {TFrsh64,Dom64,Dom128}	.5%
10	{TropP64,TropP96,CHill64} {FNat64,FGold64} {Trop64,MMaid64,MMaid96,TFrsh64} {Dom64,Dom128}	.5%

Table 11. Top ten market structures and their posterior probabilities estimating using a weak prior.

	Trop P64	Trop P96	Fnat 64	Trop 64	MMai d64	MMai d96	CHill 64	TFrsh 64	FGol d64	Dom 64	Dom 128
TropP64	1.00	.69	.49	.42	.21	.25	.24	.24	.32	.20	.19
TropP96		1.00	.41	.36	.20	.30	.15	.19	.25	.25	.26
FNat64			1.00	.33	.25	.23	.29	.29	.56	.15	.12
Trop64				1.00	.30	.26	.43	.38	.41	.14	.12
MMaid64					1.00	.60	.39	.49	.36	.17	.13
MMaid96						1.00	.27	.38	.24	.25	.25
CHill64							1.00	.42	.49	.14	.07
TFrsh64								1.00	.38	.14	.10
FGol64									1.00	.11	.06
Dom64										1.00	.75
Dom128											1.00

Table 12. Posterior probability that a pair of products will be assigned to the same subcategory using a weak prior. The product abbreviations are given in Table 1.