3   Measurement of own- and cross-price effects
Qing Liu, Thomas Otter and Greg M. Allenby

Abstract

The accurate measurement of own- and cross-price effects is difficult when there exists a moderate to large number of offerings (e.g., greater than five) in a product category because the number of cross-effects increases geometrically. We discuss approaches that reduce the number of uniquely estimated effects through the use of economic theory, and approaches that increase the information contained in the data through data pooling and the use of informative prior distributions in a Bayesian analysis. We also discuss new developments in the use of supply-side models to aid in the accurate measurement of pricing effects.

Introduction

The measurement of price effects is difficult in marketing because of the many competitive offerings present in most product categories. For J brands, there are J^2 possible effects that characterize the relationship between prices and sales. The number of competitive brands in many product categories is large, taxing the ability of the data to provide reliable estimates of own- and cross-price effects. A recent study by Fennell et al. (2003), for example, reports the median number of brands in 50 grocery store product categories to be 15. This translates into 225 own- and cross-effects that require measurement in the demand system.

Structure-imposing assumptions are therefore required to successfully estimate price effects. At one end of the spectrum, a pricing analyst could simply identify subsets of brands that are thought to compete with each other, and 'zero-out' the cross-effects for brands that are assumed not to compete. While this provides a simple solution to the task of reducing the dimensionality of the measurement problem, it requires strong beliefs about the structure of demand in the marketplace. Moreover, this approach does not allow the data to express contrary evidence.

Alternatively, one might attempt to measure directly all J^2 own- and cross-price effects. However, it is quickly apparent that using a general rule of thumb that one should have n data points for each effect-size measured rules out the use of most commercially available data. Using weekly sales scanning data and the rule that $n = 5$ results in the need for 20 years of data in food product categories such as orange juice or brownies. One could also engage in the generation of data through experimental means, using surveys or field experiments. The data requirements, however, remain formidable.

We discuss approaches to measuring price effects that rely on modeling assumptions to (i) reduce the number of the effects being measured; and/or (ii) increase the information available for measurement. We begin with a brief review of economic theory relevant to price effects, and then discuss the use of economic models to measure them. We then turn our attention to approaches that increase the available information. These approaches are Bayesian in nature, with information being available either through prior information or from data pooled from other sources. We provide a brief review of modern Bayesian

methods for pooling data, including the use of hierarchical models, and models that incorporate the price-setting behavior of firms (i.e. supply-side models). We conclude with a discussion of measuring price effects in the presence of dynamic effects and other forms of interactions.

1.   Economic models for pricing

According to economic theory, own-price effects should be negative and cross-price effects should be positive for competitive goods. As the price of a brand increases, its own sales should decline. As the price of a competitive brand increases, sales should increase. A commonly encountered problem in the use of regression models for measuring price effects is that cross-effects are often estimated to have the wrong algebraic sign – i.e. they are estimated to be negative when they should be positive. Similarly, but less often, own-price effects are estimated to be positive when they should be negative.

When price effects estimates have erroneous signs and large standard errors, a pricing analyst may be tempted to zero them out and re-estimate the remaining effects as described above. However, doing so imposes strong assumptions about the competitive nature of demand – it means that price of one brand has no effect on another brand, for any price, including zero. While approaches such as Bayesian variable selection (George and McCulloch, 1993) help quantify uncertainty in specification searches (Leamer, 1978) such as this, they require the strong assumptions that some of the effect-sizes have a prior probability of being zero. The assumption of a zero effect is often untenable, especially when deriving estimates from aggregate sales data where at least some customers will react to the price change. So, while the practice of setting coefficients to zero solves the problem of incorrectly signed estimates, it does so by imposing somewhat unbelievable assumptions about the structure of demand.

An alternative approach is to employ economic theory to avoid the direct estimation of the J^2 price effects. As with any theory, the use of an economic model reduces the dimensionality of the effects through model parameters. Economic models of behavior are based on the idea of constrained utility maximization:

$$\max_{x} U(x_1, \ldots, x_J) = \sum_{j=1}^{J} \psi_j x_j$$
$$\text{subject to } \sum_{j=1}^{J} p_j x_j \leq E$$

(3.1)

where $U(x_1, \ldots, x_J)$ denotes the utility of x_1 units of brand 1, x_2 units of brand 2, . . . and x_J units of brand J. In the specification above, utility takes on an additive form that implies that the brands are perfect substitutes. Moreover, this model assumes that utility increases by a constant amount ψ_j as quantity (x_j) increases (i.e. marginal utility is constant). A consumer maximizes utility subject to the budget constraint where p_j is the unit price of brand j, and E is the budgetary allotment – the amount the consumer is willing to spend.

The solution to equation (3.1) can be shown to lead to a discrete choice model, where all expenditure E is allocated to the brand with the biggest bang-for-the-buck, ψ_j/p_j. Assuming that marginal utility has a stochastic component unobservable to the analyst, i.e. $\psi_j = \overline{\psi}_j \exp(\varepsilon_j)$, leads to the demand model:

$$\text{Pr}(x_k > 0) = \text{Pr}\left(\frac{\psi_k}{p_k} > \frac{\psi_j}{p_j} \text{ for all } j\right)$$

$$= \text{Pr}(\ln\psi_k - \ln p_k > \ln\psi_j - \ln p_j \text{ for all } j) \qquad (3.2)$$

$$= \text{Pr}(\ln\overline{\psi}_k - \ln p_k + \varepsilon_k > \ln\overline{\psi}_j - \ln p_j + \varepsilon_j \text{ for all } j)$$

The assumption that the error term, ε, is normally distributed leads to a probit model, and the assumption of extreme value errors leads to the logit model. More specifically, if ε is distributed extreme value with location zero and scale σ, then equation (3.2) can be expressed as (McFadden, 1981):

$$\text{Pr}(x_k > 0) = \frac{\exp\left[\dfrac{\ln\overline{\psi}_k - \ln p_k}{\sigma}\right]}{\sum\limits_{j=1}^{J}\exp\left[\dfrac{\ln\overline{\psi}_j - \ln p_j}{\sigma}\right]}$$

$$= \frac{\exp[V_k]}{\sum\limits_{j=1}^{J}\exp[V_j]} \qquad (3.3)$$

where V_k can be written as $\beta_{0k} - \beta_p \ln p_k$ with $\beta_p = 1/\sigma$ and the intercept β_{0k} equal to $\ln \overline{\psi}_k/\sigma$. Since the sum of all probabilities specified by (3.3) adds up to 1, one of the model intercepts is not identified, and it is customary to set one intercept to zero, leaving $J - 1$ free intercepts and one price coefficient.

Thus the use of an economic model (equations 3.1–3.3) requires J parameters to measure the J^2 own- and cross-price effects. This represents a large reduction in parameters (e.g. from 225 to 15 when $J = 15$) that greatly improves the accuracy of estimates. Given the estimated parameters in equation (3.3), own- and cross-price effects can be computed under the assumption that demand (x) takes on values of only zero or one. With this assumption, we can equate choice probability with expected demand, and we can compute own- and cross-effects as

$$\frac{\partial\ln\text{Pr}_j}{\partial\ln p_j} = -\beta_p(1 - \text{Pr}_j) \quad \text{and} \quad \frac{\partial\ln\text{Pr}_j}{\partial\ln p_k} = \beta_p\text{Pr}_k \qquad (3.4)$$

where the former is what economists call own elasticity, and the later is the cross-elasticity. It measures the percentage change in expected demand for a percentage change in price.

Economic models can be used to improve the measurement of own- and cross-price effects in either of two ways. The first is to use the model to suggest constraints for an otherwise purely descriptive model. The second is to directly estimate parameters of the micro economic model, and then use these to measure the price effects.

Using economic theory to constrain descriptive models
Most descriptive models of demand are of log-log or semi-log form. Researchers have extended descriptive models in various ways to achieve more flexible functional forms and to account for uncertainty in the functional form (Kalyanam, 1996; Kalyanam and Shively, 1998). For typical marketing data, where the effective unit of analysis usually

only supplies a limited amount of data, highly flexible descriptive models are especially likely to benefit from constraints derived from economic theory. As we will show, the use of economic theory to derive prior distributions for descriptive models is especially useful in this context. A strong signal in the data can override the implications of economic theory but economic theory will dominate data that are not informative to begin with.

Equation (3.4) suggests a number of constraints on price coefficients that can aid direct estimation of the J^2 own- and cross-price effects using descriptive models. Since β_p is simply the inverse of the scale of the error term, we have $\beta_p > 0$ as $\sigma^2 > 0$, implying that

$$\frac{\partial \ln \Pr_j}{\partial \ln p_j} < 0 \quad \text{and} \quad \frac{\partial \ln \Pr_j}{\partial \ln p_k} > 0 \tag{3.5}$$

Constraints of this type, which we call 'ordinal restrictions', occur frequently in the analysis of marketing data. In addition to demand system estimation, the analysis of survey data and use of conjoint analysis are settings in which it is desirable to constrain coefficients so that they are sensible. In addition to expecting that people would rather pay less than more for an offering, researchers also may want to estimate models where preference for a known brand is preferred to an unknown brand, or that respondents prefer better performance assuming all else is held constant.

Natter et al. (2007) describe a decision support system used by bauMax, an Austrian firm in the do-it-yourself home repair industry, which employs ordinal restrictions to derive own effects with correct (negative) algebraic signs. These effects are used by bauMax to derive optimal mark-down policies for the 60,000 stockkeeping units in its stores. Store profits are reported to have increased by 8.1 percent using the decision support system.

Bayesian statistical analysis (see Rossi et al., 2005) offers a convenient solution to incorporating ordinal constraints in models of demand. In a Bayesian analysis, the analyst specifies a prior distribution for the model parameters that reflects his or her beliefs before observing the data. The prior is combined with the data through the likelihood function to arrive at the posterior distribution:

$$\pi(\theta \,|\, Data) \propto \pi(Data \,|\, \theta)\pi(\theta) \tag{3.6}$$

where $\pi(\theta)$ denotes the prior distribution, $\pi(Data|\theta)$ denotes the likelihood function; and $\pi(\theta|Data)$ is the posterior distribution. In a regression model, for example, we have

$$y_i = x_i'\beta + \varepsilon_i; \quad \varepsilon_i \sim Normal(0, \sigma^2) \tag{3.7}$$

and assuming the error terms are normally distributed, the likelihood of the observed data is

$$\pi(Data \,|\, \theta = (\beta, \sigma^2)) = \prod_{i=1}^{n} \pi(y_i | x_i, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_i - x_i'\beta)^2}{2\sigma^2}\right] \tag{3.8}$$

where x_i is treated as an independent variable and used as a conditioning argument in the likelihood, and the observations are assumed to be independent given the independent variables x and model parameters, $\theta = (\beta, \sigma^2)$. A prior distribution for the regression coefficients β typically also takes on the form of a normal distribution:

$$\pi(\beta \,|\, b, s^2) = \frac{1}{\sqrt{2\pi s^2}} \exp\left[\frac{-(\beta - b)^2}{2s^2}\right] \tag{3.9}$$

where the prior mean, b, and prior variance, s^2, are specified by the analyst. The prior for σ^2 is typically taken to be inverted chi-squared.

Allenby et al. (1995) demonstrate that ordinal constraints can be incorporated into the analysis by specifying a truncated normal prior distribution in (3.9) instead of a normal distribution:

$$\pi(\beta \,|\, b, s^2, \textit{ordinal restrictions}) = k \exp\left[\sum_{i=1}^{n} \frac{-(\beta - b)^2}{2s^2}\right] I_{\textit{ordinal restrictions}} \tag{3.10}$$

where k is an integrating constant that replaces the factor $1/\sqrt{2\pi s^2}$ in equation (3.9), I is an indicator function equal to one when the ordinal constraints are satisfied, and the parameters b and s^2 are specified by the analyst. Examples of ordinal constraints are that an own-price coefficient should be negative, or that a cross-price coefficient should be positive.

From (3.6), the posterior distribution obtained from the likelihood (equation 3.8) and truncated prior (equation 3.10) is:

$$\pi(\theta \,|\, Data) \propto \pi(Data \,|\, \theta)\pi(\theta)I_{\textit{ordinal restrictions}} \tag{3.11}$$

which is the truncated version of the unconstrained posterior. Thus the incorporation of ordinal constraints in an analysis is conceptually simple. The difficulty, until recently, has been in making equation (3.11) operational to the analyst. Analytical expressions for the posterior mean and associated confidence, or credible intervals for the posterior distribution, are generally not available.

Markov chain Monte Carlo (MCMC) estimation offers a tractable approach to working with the truncated posterior distribution in (3.11). The idea is to replace difficult analytic expressions with a series of simple, iterative calculations that result in Monte Carlo draws from the posterior. A Markov chain is constructed with stationary distribution equal to the posterior distribution, allowing the analyst to simulate draws from the posterior. These draws are then used to characterize the posterior distribution. For example, the posterior mean is estimated by taking the mean of the simulated draws from the posterior. Confidence intervals and standard deviations are evaluated similarly.

An important insight about simulation-based methods of estimation (e.g. MCMC) is that once a simulator is developed for sampling from the unconstrained parameter distribution (equation 3.6), it is straightforward to sample from the constrained distribution (equation 3.11) by simply ignoring the simulated draws that do not conform to the restrictions. This is a form of rejection sampling, one of many tools available for generating draws from non-standard distributions.

Economic theory can also be used to impose exact restrictions on own- and cross-price effects. Consider, for example, the constraints implied by equation (3.4). A total of $J^2 - J$ constraints is implied by equation (3.4) because there are J^2 own- and cross-price effects and just J parameters in the logit model in (3.3). One set of constraints is related to the well-known independence of irrelevant alternative (IIA) constraints of logit models. The IIA constraint is typically derived from the logit form in (3.3), where the ratio of choice probabilities of any two brands (e.g. i and j) is unaffected by other brands (e.g. k). Thus

changes in the price of brand k must draw proportionally equal choice probability share from brands i and j.

The IIA property is also expressed in equation (3.4) by realizing that the elasticity of demand for brand j with respect to the price of brand k (i.e. η_{jk}) takes the form:

$$\eta_{jk} = \frac{\partial \ln \Pr_j}{\partial \ln p_k} = \beta_p \Pr_k \quad \text{implying} \quad \eta_{jk} = \eta_{ik} = \ldots = \eta_{Jk} | j \neq k \qquad (3.12)$$

Thus the change in the price of brand k has a proportionately equal effect on all other choice probabilities. Equation (3.12) implies a 'proportional draw' property for cross-price effects. In a similar manner it can be shown (see Allenby, 1989) that

$$\frac{\eta_{jk}}{\eta_{ji}} = \frac{\Pr_k}{\Pr_i} \qquad (3.13)$$

indicating that the magnitude of price elasticity is proportional to the choice probability. Equation (3.13) implies a 'proportional influence' property where an individual's choice probability is influenced more by price changes of the brands they prefer. At an aggregate level, this implies that brands with greater market share have greater influence.

The constraints implied by equations (3.12) and (3.13) can be incorporated into descriptive regression models either by direct substitution or through the use of a prior distribution. Direct substitution imposes the constraints exactly, and a prior distribution provides a mechanism for stochastically imposing the constraints. For example, in analysis of aggregate data, one could substitute a brand's average market share (m) for the choice probability, and reduce the number of parameters in a regression model by using equation (3.13):

$$\ln m_{jt} = \beta_{0j} + \eta_{jj} \ln p_{jt} + \eta_{jk} \ln p_{kt} + \eta_{ji} \ln p_{it} + \cdots$$
$$= \beta_{0j} + \eta_{jj} \ln p_{jt} + \eta_{jk} \left(\ln p_{kt} + \frac{m_i}{m_k} \ln p_{it} + \cdots \right) \qquad (3.14)$$

where t is an index for time. A more formal and flexible approach is to employ a prior distribution that stochastically constrains model parameters to lie close to the subspace implied by the restrictions. Restrictions on the own- and cross-price effects can be expressed as functions of parameters, and priors can be placed on their functional values. To express the equality in equation (3.12), which is equivalent to $\eta_{1k} - \eta_{2k} = \ldots = \eta_{1k} - \eta_{Jk}$, a contrast matrix, R, is used:

$$R = \begin{bmatrix} 1 & -1 & 0 & \cdots \\ 1 & 0 & -1 & \cdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & -1 \end{bmatrix} \qquad (3.15)$$

If equation (3.12) holds exactly, the product $R\eta$ with $\eta = (\eta_{1k}, \ldots \eta_{Jk})$ is a vector of zeros and a prior centered on this belief can be expressed using a normal distribution with mean zero:

$$\pi(R\eta) = (2\pi)^{-(J-1)/2} |S|^{-1/2} \exp\left[-\frac{1}{2}(R\eta)S^{-1}(R\eta)' \right] \qquad (3.16)$$

   An advantage of this approach is that the prior distribution can be used to control the precision of the restriction through the variance–covariance matrix S.

   Montgomery and Rossi (1999) use such an approach to impose restrictions on price elasticities in a descriptive model of demand. This approach assumes that the prior distribution can be constructed with measures that are (nearly) exogenous to the system of study. This assumption is also present in equation (3.14) when employing average market shares, m_i, to impose restrictions. It is reasonable when there are many brands in a category, such that any one brand has little effect on the aggregate expenditure elasticity for the category, when there are sufficient time periods so that the average market share for a brand is reliably measured and when there are no systematic movements in the shares across time.

Formal approaches to demand estimation

The use of linear models to estimate own- and cross-price effects has a long history in economics. Linearity, however, has been limiting research to a restricted number of utility functions. Demand functions, in general, are derived by solving for the demand that maximizes utility subject to the budget (i.e. income) constraint. For the Cobb–Douglas utility function, the demand function can be shown to be of log-log form where the logarithm of quantity is a linear function of logarithm of income and logarithm of price (Simon and Blume, 1994, Example 22.1). Other utility functions do not result in demand functions that are easily estimable with OLS (ordinary least squares).

   Some analysts elect to start with the indirect utility function rather than the utility function. The indirect utility function is defined as the maximum utility attainable for a given set of prices and expenditure. It can be shown that differentiating the indirect utility function using Roy's identify (see Simon and Blume, 1994, Theorem 22.5) leads to the demand equation in which demand is expressed as a function of price and income. Varian (1984, ch. 4) demonstrates that this approach usually leads to demand functions that are nonlinear. Some indirect utility functions, such as the translog function of Christensen et al. (1975), lead to linear systems for estimation if a representative economic agent is assumed and consumer heterogeneity is thus ignored. Integrating over a distribution of heterogeneity results in a nonlinear specification that requires the use of alternative methods of estimation (see Allenby and Rossi, 1991 for an exception).

   A direct approach to demand estimation is to derive the likelihood of the data corresponding to constrained utility maximization. Distributional assumptions are made about stochastic errors that enter the utility function, understood as information known to the consumer but not observed by the analyst, and from these primitive assumptions the likelihood is derived. Kim et al. (2002) provide an example of such an approach, where utility is specified with diminishing marginal returns:

$$\max_x U(x_1, \ldots, x_J) = \sum_{j=1}^{J} \psi_j (x_j + \gamma_j)^{\alpha_j}$$

$$\text{subject to } \sum_{j=1}^{J} p_j x_j \leq E \tag{3.17}$$

Here, γ_j translates the utility function to allow for corner and interior solutions. Diminishing marginal returns occur if α_j is positive and less than one. The likelihood is

obtained by differentiating the Lagrangian $U(x) - \lambda(p'x - E)$ to obtain the Kuhn–Tucker (KT) conditions as follows:

$$\frac{\partial U}{\partial x_1} - \lambda p_1 = \ldots = \frac{\partial U}{\partial x_J} - \lambda p_J = 0, \text{ that is, } \frac{\partial U}{\partial x_1}\frac{1}{p_1} = \ldots = \frac{\partial U}{\partial x_J}\frac{1}{p_J} = \lambda$$

where $\partial U/\partial x_j = \psi_j \alpha_j (x_j + \gamma_j)^{\alpha_j - 1}, j = 1, \ldots J$. Assuming that log marginal utility can only be measured up to additive error, i.e. $\ln\psi_j = \ln\overline{\psi}_j + \varepsilon_j$, and that the observed data conform to the KT conditions, we have for both x_i and x_j positive:

$$\ln(\overline{\psi}_i \alpha_i (x_i + \gamma_i)^{\alpha_i - 1}) - \ln p_i + \varepsilon_i = \ln(\overline{\psi}_j \alpha_j (x_j + \gamma_j)^{\alpha_j - 1}) - \ln p_j + \varepsilon_j \quad (3.18)$$

or

$$(\ln(\overline{\psi}_i \alpha_i (x_i + \gamma_i)^{\alpha_i - 1}) - \ln p_i) - (\ln(\overline{\psi}_j \alpha_j (x_j + \gamma_j)^{\alpha_j - 1}) - \ln p_j) = \varepsilon_j - \varepsilon_i \quad (3.19)$$

Equation (3.19) provides a basis for deriving the likelihood of the data, $\pi(Data|\theta = (\overline{\psi}, \alpha, \gamma))$ through the distribution of $(\varepsilon_j - \varepsilon_i)$. The distribution of the observed data $\{x_i, x_j\}$ is obtained as the distribution of the calculated errors $\{\varepsilon_i, \varepsilon_j\}$ multiplied by the Jacobian of the transformation from ε to x. Modern Bayesian (MCMC) methods are well suited to estimate such models because they require the evaluation of the likelihood only at specific values of the parameters, and do not require the evaluation of gradients or Hessians of the likelihood. Once the parameters of the utility function are available, estimates of own- and cross-effects can be obtained by solving equation (3.17) numerically for various price vectors and computing numeric derivatives.

Standard discrete choice models such as multinomial logit and probit models are the simplest examples of the direct approach. Utility is assumed to take a linear form with constant marginal utility (equation 3.1), and random error is introduced as shown in equation (3.2). Constant marginal utility implies that as income increases consumers simply consume more of the same brand rather than switching to a higher-quality brand. Allenby and Rossi (1991) use a non-constant marginal utility (non-homothetic), which motivates switching from inferior goods to superior goods as income increases. As a consequence, price responses are asymmetric. Price changes of high-quality brands have a higher impact on low-quality brands than vice versa (see Blattberg and Wisniewski, 1989 for a motivation of asymmetric price response based on heterogeneity).

Chiang (1991) and Chintagunta (1993) remove the 'given purchase' condition inherent to discrete choice models and model purchase incidence, brand choice and purchase quantity simultaneously through a bivariate utility function. A generalized extreme value distribution implies both a probability to purchase and a brand choice probability. A flexible translog indirect utility function is maximized with respect to quantity given a brand is purchased. Variants of this approach have been used by Arora et al. (1998), Bell et al. (1999), and Nair et al. (2005).

The translog approach results in price effects that can be decomposed into three parts: changes in purchase probability, changes in brand choice given purchase occurrence; changes in purchase quantity given purchase occurrence and brand selection. Bell et al. (1999) show that these three components are influenced in different ways by exogenous consumer-, brand- and category-specific variables.

The linear additive utility specification popular in marketing implies that all brands are perfect substitutes, so that only one brand is chosen as the utility-maximizing solution. Nonlinear utility functions such as (3.17) allow for both corner and interior solutions. That is, a consumer chooses one alternative or a combination of different alternatives as the result of utility maximization. Thus the model quantifies the tradeoff between price and the variety of the product assortment (see Kim et al., 2002, 2007 for details). A different form of nonlinear utility function is used by Dubé (2004), who motivates the choice of more than one brand by multiple consumption occasions that are considered during a customer's shopping trip.

2.   Improving measurement with additional information

An alternative to constraining and/or reducing the parameter space through the use of economic models is to use approaches that attempt to increase the available information for estimation. We investigate two approaches to data pooling. The first is with the use of random-effects models that effectively borrow information from other similar units through the random-effects distribution. The second approach pools information from the supply side. This approach views the prices themselves as endogenous to the system of study, and models are specified as a system of demand and supply equations. Both approaches have become practical in applications with the advent of modern Bayesian methods.

Pooling across units

Random-effects models add another layer to the Bayesian prior distribution. Equation (3.9) is the prior associated with one unit of analysis, where the unit might be sales at a specific retailer or in a specific geographic region. When multiple units of analysis are available, it is possible to pool the data by specifying a relationship among the model parameters:

$$\pi(Data_i | \theta_i) \quad \text{for } i = 1, \ldots, N$$
$$\pi(\theta_i | \zeta)$$
$$\pi(\zeta) \tag{3.21}$$

where ζ are known as hyper-parameters – i.e. parameters that describe the distribution of other parameters. For example, $\pi(Data_i | \theta_i)$ could represent a time-series regression model for sales of a specific brand in region i, with own- and cross-effects coefficients θ_i. The second layer of the model, $\pi(\theta_i | \zeta)$, is the random-effects model. A commonly assumed distribution is multivariate normal. Finally, the third layer, $\pi(\zeta)$, is the prior distribution for the hyper-parameters.

Pooling occurs in equation (3.21) because θ_i is present in both the first and second equations of the model, not just the first. The data from all units are used to inform the hyper-parameters, and as the accuracy of the hyper-parameter estimates increases, so does that of the estimates of the individual-level parameters, θ_i. The posterior distribution of the hierarchical model in (3.21) is

$$\pi(\{\theta_i\}, \zeta | \{Data_i\}) \propto \prod_{i=1}^{N} \left(\prod_{t=1}^{T_i} \pi(Data_{it} | \theta_i) \right) \pi(\theta_i | \zeta) \pi(\zeta) \tag{3.22}$$

which highlights a key difference between the Bayeisan and non-Bayesian treatment of random-effects models. In a Bayesian treatment, the posterior comprises the hyper-parameters and all individual-level parameters. In a non-Bayesian treatment, parameters are viewed as fixed but unknown constants, the analysis proceeds by forming the marginalized likelihood of the data:

$$\pi(\{Data_i\}|\zeta) = \prod_{i=1}^{N} \int \left(\prod_{t=1}^{T_i} \pi(Data_{it}|\theta_i) \right) \pi(\theta_i|\zeta) d\theta_i \qquad (3.23)$$

The Bayesian treatment does not remove the individual-level parameters from analysis, and inferences about unit-specific parameters are made by marginalizing the posterior distribution in equation (3.22):

$$\pi(\theta_i|\{Data_i\}) = \int \pi(\{\theta_i\}, \zeta|\{Data_i\}) d\{\theta_{-i}, \zeta\} \qquad (3.24)$$

Modern Bayesian methods deliver the marginal posterior distribution of model parameters at no additional computational cost. The MCMC algorithm simulates draws from the full posterior distribution of model parameters in (3.22). Analysis for a particular unit, θ_i, proceeds by simply ignoring the simulated draws of the other model parameters, θ_{-i} and ζ. Thus the hierarchical model, coupled with modern Bayesian statistical methods, offers a powerful and practical approach to data pooling to improve parameter estimates.

Allenby and Ginter (1995), and Lenk et al. (1996) demonstrate the efficiency of the estimates obtained from the hierarchical Bayes approach in comparison with the traditional estimation methods. The number of erratic signs on price-elasticity estimates is significantly reduced as more information becomes available via pooling. Montgomery (1997) uses this methodology to estimate store-level parameters from a panel of retailers. Ainslie and Rossi (1998) employ a hierarchical model to measure similarities in demand across categories. Arora et al. (1998) jointly model individual-level brand choice and purchase quantity, and Bradlow and Rao (2000) model assortment choice using hierarchical models.

Bayesian pooling techniques have found their way into practice through firms such as DemandTec (demandtec.com), who specialize in retail price optimization. Current customers of DemandTec include Target, WalMart and leading grocers such as Safeway and Giant Eagle. A major challenge in setting optimal prices at the stockkeeping unit level is the development of demand models that accurately predict the effects of price changes on own sales and competitive sales. Retailers want to set prices to optimize profits in a product category, and a critical element involves estimating coefficients with correct algebraic signs (i.e. own-effects are negative, cross-effects are positive) so that an optimal solution exists. For example, if an own-effect is estimated to be positive, it implies that an increase in price is associated with an increase in demand, and the optimal price is therefore equal to positive infinity. This solution is neither reasonable nor believable. DemandTec uses hierarchical Bayesian models such as equation (3.21) to pool data across similar stockkeeping units to help obtain more accurate price effects with reasonable algebraic signs.

Another industry example of the use of hierarchical Bayesian analysis is Sawtooth Software (sawtoothsoftware.com), the leading supplier of conjoint software. Conjoint

analysis is a popular quantitative technique used to evaluate consumer utility for attribute levels, and express them in terms of a common metric. For example, consumer preference for different credit cards can be viewed in terms of utility for different interest rates, grace periods, annual fees, etc. Conjoint analysis estimates the part-worths of the levels of these attributes. In most studies, price is specified as an attribute, and consumer price-sensitivity (β_p) is measured at the individual-respondent level using a hierarchical model. The individual-level estimates are then used to predict changes in demand for all products in a category in response to changes in product attributes, including price. Data pooling via a hierarchical model structure is critical for obtaining individual-level part-worths because of the limited number of conjoint questions that can be asked of a respondent in an interview. Sales for the hierarchical Bayes version of Sawtooth's conjoint software now dominates their non-Bayesian version.

Incorporating supply-side data

Up to this point we have considered models where prices are viewed as explanatory of sales, and also independently determined. This assumption is acceptable when analyzing survey and experimental data because prices are set by the analyst. However, when data are from the marketplace, prices are set in anticipation of demand and profits. Observed prices are influenced by the preferences and sensitivities of consumers, the same factors (e.g. utility function parameters) that influence the magnitude of the own- and cross-price effects.

When explanatory variables are endogenously determined, the likelihood will comprise multiple equations that form a system of equations. Exceptions to this general rule are discussed by Liu et al. (2007). As discussed in the use of formal economic models above, the key in conducting analysis of simultaneous equation systems is to relate primitive assumptions about how errors enter the model to the likelihood for the observed data.

Consider, for example, a monopolist pricing problem using a constant elasticity model, where it is assumed that the variation in prices over time is due to stochastic departures from optimal price-setting behavior. The likelihood for the data is a combination of a traditional demand model:

$$\ln y_t = \beta_0 + \beta_1 \ln p_t + \varepsilon_t; \quad \varepsilon_t \sim Normal(0, \sigma_\varepsilon^2) \tag{3.25}$$

and a factor for the endogenous price variable. Optimal pricing for the monopolist can be shown to be (see for example, Pashigian, 1998, p. 333):

$$p_t = mc\left(\frac{\beta_1}{1 + \beta_1}\right)e^{v_t}; \quad v_t \sim Normal(0, \sigma_v^2) \tag{3.26}$$

where *mc* denotes the marginal cost of the brand, and a supply-side error term has been added to account for temporal variation of observed prices from the optimal price. Taking logs of equation (3.26) yields

$$\ln p_t = \ln mc + \ln\left(\frac{\beta_1}{1 + \beta_1}\right) + v_t; \quad v_t \sim Normal(0, \sigma_v^2) \tag{3.27}$$

Equations (3.25) and (3.27) form a system of equations that effectively pools supply-side information and improves the estimation of the own-price effect, β_1, if the marginal cost

of the brand is known. That is, the average level of price is informative about β_1 given marginal cost. The likelihood for equations (3.25) and (3.27) is obtained by solving for error terms:

$$\varepsilon_t = \ln y_t - \beta_0 - \beta_1 \ln p_t \sim Normal(0, \sigma_\varepsilon^2)$$

$$v_t = \ln p_t - \ln mc - \ln\left(\frac{\beta_1}{1 + \beta_1}\right) \sim Normal(0, \sigma_v^2) \tag{3.28}$$

and computing:

$$\pi(Data|\theta) = \prod_{t=1}^{T} \pi(y_t, p_t | \beta_0, \beta_1, \sigma_\varepsilon^2, \sigma_v^2)$$

$$= \prod_{t=1}^{T} \pi(\varepsilon_t, v_t | \beta_0, \beta_1, \sigma_\varepsilon^2, \sigma_v^2) \times J_{(\varepsilon_t, v_t) \rightarrow (y_t, p_t)} \tag{3.29}$$

$$= \prod_{t=1}^{T} \pi(\varepsilon_t, v_t | \beta_0, \beta_1, \sigma_\varepsilon^2, \sigma_v^2) \times \frac{1}{y_t p_t}$$

In this example, the supply-side equation offers additional information that is useful for estimating the own-price effect in two ways. The first way, as mentioned above, is to help locate the value of β_1 if marginal cost is known. The second way is through an ordinal constraint imposed by the supply-side model – i.e. $\beta_1 < -1$ for the supply equation to be valid. If $-1 \leq \beta_1 < 0$, $\beta_1/(1+\beta_1)$ is negative, equation (3.26) no longer yields the price that maximizes profits and thus the logarithm in equation (3.27) is not defined. Optimal pricing behavior with positive, finite prices exists only when own-price effects are elastic. Thus the supply-side equation constrains the estimates of price effects by merely ascertaining that optimal pricing with positive, finite prices is possible. This aspect of supply-side analysis is investigated in more detail by Otter et al. (2007).

When the error terms, ε_t and v_t, are correlated, analysis without the supply side leads to inconsistent estimates (Besanko et al. 1998; Villas-Boas and Winer, 1999). The typical rational for correlated demand- and supply-side shocks is the presence of a common omitted variable that raises prices and demand at the same time – e.g. a retailer correctly anticipates a demand shock and simultaneously raises prices. Thus the presence of endogenous price variation requires joint estimation of demand- and supply-side equations to obtain consistent estimates of own- and cross-price effects.

Supply-side equations may be reduced-form linear models (Villas-Boas and Winer, 1999), or structural models where the supply-side equations are obtained through maximizing objective functions of firms and/or retailers. For example, Sudhir (2001a) obtains the supply-side pricing equations by assuming that the firm maximizes the sum of own profits and weighted competitor profits, where the weight on competitor profits characterizes cooperative (positive weight) or aggressive (negative weight) competitive behavior. Chintagunta (2002) obtains the supply-side pricing equations by assuming that retailers set prices to maximize a weighted sum of category profits and store brand share while accounting for manufacturers' actions, store traffic effects and retail competition. Chintagunta and Desiraju (2005) obtain supply-side equations by maximizing a profit function that accounts for firm interactions within a geographic market as well as interactions across all geographic markets. Other examples of structural supply-side models

include Besanko et al. (1998), Sudhir (2001b), Draganska and Jain (2004) and Villas-Boas and Zhao (2005).

Techniques to obtain parameter estimates in demand- and supply-side equations include generalized method of moments (GMM) estimation using instrumental variables (see Berry, 1994; Berry et al., 1995; and Nevo, 2001), maximum likelihood estimation (see Villas-Boas and Winer, 1999; Villas-Boas and Zhao, 2005; and Draganska and Jain, 2004), and the Bayesian approach (see Yang et al., 2003).

3.   Concluding comments

The measurement of own- and cross-price effects in marketing is complicated by many factors, including a potentially large number of effects requiring measurement, heterogeneity in consumer response to prices, the presence of nonlinear models of behavior, and the fact that prices are set strategically in anticipation of profits by manufacturers and retailers. Over the course of the past 20 years, improvements in statistical computing have allowed researchers to develop new models that improve the measurement of price effects.

The measurement of price effects is inextricably linked to choice and demand models, and more generally consumer decision-making. These are very active research areas, and the implications of many of the more recently published choice models for the measurement of price effects and price setting have yet to be explored. In this chapter we focused on static models that imply (only) an immediate and continuous price response. There is active research on dynamic price effects. Dynamic price effects refer to the effects of price change on future sales as mediated by stockpiling and/or increased consumption. Effects to be measured include immediate, future and cumulative (immediate + future) effects of promotional and/or regular price changes, which may differ in sign and magnitude. For example, as shown by Kopalle et al. (1999), promotions have positive immediate effects but negative future effects on baseline sales. Autoregressive descriptive demand models (see, e.g., Kopalle et al., 1999; Fok et al., 2006) and utility-based demand models (Erdem et al., 2003) have recently been used to account for carry-over effects from past discounts, forward-looking consumer behavior and competitive price reactions. The same approaches are taken to dealing with measurement difficulties – using theory to impose restrictions on parameters, Bayesian pooling, and adding supply-side information.

Finally, there is a large behavioral literature documenting the influence of consumer cognitive capacity, memory, perceptions and attitudes in reaction to price (see Monroe, 2002 for a review). An active area of current research develops demand models that incorporate such behavioral decision theory for an improved measurement of price effects (Gilbride and Allenby, 2004, 2006).

References

Ainslie, A. and P.E. Rossi (1998), 'Similarities in choice behavior across product categories', *Marketing Science*, **17** (2), 91–106.
Allenby, G.M. (1989), 'A unified approach to identifying, estimating and testing demand structures with aggregate scanner data', *Marketing Science*, **8** (3), 265–80.
Allenby, G.M. and J.L. Ginter (1995), 'Using extremes to design products and segment markets', *Journal of Marketing Research*, **32** (4), 392–403.
Allenby, G.M. and P.E. Rossi (1991), 'Quality perceptions and asymmetric switching behavior between brands', *Marketing Science*, **10** (3), 185–204.
Allenby, G.M., N. Arora and J.L. Ginter (1995), 'Incorporating prior knowledge into the analysis of conjoint studies', *Journal of Marketing Research*, **32** (2), 152–62.

Arora, N., G.M. Allenby and J.L. Ginter (1998), 'A hierarchical Bayes model of primary and secondary demands', *Marketing Science*, **17** (1), 29–44.
Bell, D.R., J. Chiang and V. Padmanabhan (1999), 'The decomposition of promotional response: an empirical generalization', *Marketing Science*, **18** (4), 504–26.
Berry, S.T. (1994), 'Estimating discrete-choice models of product differentiation', *Rand Journal of Economics*, **25** (2), 242–53.
Berry, S., J. Levinsohn and A. Pakes (1995), 'Automobile prices in market equilibrium', *Econometrica*, **63** (4), 841–90.
Besanko, D., S. Gupta and D. Jain (1998), 'Logit demand estimation under competitive pricing behavior: an equilibrium framework', *Management Science*, **44** (11), 1533–47.
Blattberg, R.C. and K.J. Wisniewski (1989), 'Price-induced pattern of competition', *Marketing Science*, **8** (4), 291–309.
Bradlow, E.T. and V.R. Rao (2000), 'A hierarchical Bayes model for assortment choice', *Journal of Marketing Research*, **37** (2), 259–68.
Chiang, J. (1991), 'A simultaneous approach to the whether, what and how much to buy questions', *Marketing Science*, **10** (4), 297–315.
Chintagunta, P.K. (1993), 'Investigating purchase incidence, brand choice and purchase quantity decisions of households', *Marketing Science*, **12** (2), 184–208.
Chintagunta, P.K. (2002), 'Investigating category pricing behavior at a retail chain', *Journal of Marketing Research*, **39** (2), 141–54.
Chintagunta, P.K. and R. Desiraju (2005), 'Strategic pricing and detailing behavior in international markets', *Marketing Science*, **24** (1), 67–80.
Christensen, L., D. Jorgenson and L. Lau (1975), 'Transcendental logarithmic utility functions', *American Economic Review*, **3**, 367–83.
Draganska, M. and D. Jain (2004), 'A likelihood approach to estimating market equilibrium models', *Management Science*, **50** (5), 605–16.
Dubé, J.-P. (2004), 'Multiple discreteness and product differentiation: demand for carbonated soft drinks', *Marketing Science*, **23** (1), 66–81.
Erdem, T., S. Imai and M. Keane (2003), 'Brand and quantity choice dynamics under price uncertainty', *Quantitative Marketing and Economics*, **1** (1), 5–64.
Fennell, G., G.M. Allenby, S. Yang and Y. Edwards (2003), 'The effectiveness of demographic and psychographic variables for explaining brand and product use', *Quantitative Marketing and Economics*, **1**, 223–44.
Fok, D., C. Horvath, R. Paap and P.H. Franses (2006), 'A hierarchical Bayes error correction model to explain dynamic effects of price changes', *Journal of Marketing Research*, **43** (3), 443–61.
George, E. and R. McCulloch (1993), 'Variable selection via Gibbs sampling', *Journal of American Statistical Association*, **88**, 881–9.
Gilbride, T.J. and G.M. Allenby (2004), 'A choice model with conjunctive, disjunctive, and compensatory screening rules', *Marketing Science*, **23** (3), 391–406.
Gilbride, T.J. and G.M. Allenby (2006), 'Estimating heterogeneous EBA and economic screening rule choice models', *Marketing Science*, **25** (5), 494–509.
Kalyanam, K. (1996), 'Pricing decisions under demand uncertainty: a Bayesian mixture model approach', *Marketing Science*, **15** (3), 207–21.
Kalyanam, K. and T.S. Shively (1998), 'Estimating irregular price effects: a stochastic spline regression approach', *Journal of Marketing Research*, **35** (1), 16–29.
Kim, J., G.M. Allenby and P.E. Rossi (2002), 'Modeling consumer demand for variety', *Marketing Science*, **21** (3), 229–50.
Kim, J., G.M. Allenby and P.E. Rossi (2007), 'Product attributes and models of multiple discreteness', *Journal of Econometrics*, **138** (1), 208–30.
Kopalle, P.K., C.F. Mela and L. Marsh (1999), 'The dynamic effect of discounting on sales: empirical analysis and normative pricing implications', *Marketing Science*, **18** (3), 317–32.
Leamer, E. (1978), *Specification Searches: Ad-hoc Inference with Non-Experimental Data*, New York: John Wiley & Sons.
Lenk, P.J., W.S. DeSarbo, P.E. Green and M.R. Young (1996), 'Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental design', *Marketing Science*, **15** (2), 173–91.
Liu, Q., T. Otter and G.M. Allenby (2007), 'Investigating endogeneity bias in marketing', *Marketing Science*, **26** (5), 642–50.
McFadden, D. (1981), 'Econometric models of probabilistic choice', in C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press, pp. 198–272.
Monroe K.B. (2002), *Pricing: Making Profitable Decisions*, 3rd edn, Boston, MA: McGraw-Hill/Irwin.

Montgomery, A. (1997), 'Creating micro-marketing pricing strategies using supermarket scanner data', *Marketing Science*, **16** (4), 315–37.

Montgomery, A.L. and P.E. Rossi (1999), 'Estimating price elasticities with theory-based priors', *Journal of Marketing Research*, **36** (4), 413–23.

Nair, H., J.-P. Dubé and P.K. Chintagunta (2005), 'Accounting for primary and secondary demand effects with aggregate data', *Marketing Science*, **24** (3), 444–60.

Natter, M., T. Reutterer, A. Mild and A. Taudes (2007), 'An assortment-wide decision-support system for dynamic pricing and promotion planning in DIY retailing', Working Paper, Univeristy of Frankfurt.

Nevo, A. (2001), 'Measuring market power in the ready-to-eat cereal industry', *Econometrica*, **69** (2), 307–42.

Otter, T., L.-J. Kao, C.-C. Chiu, T. Gilbride and G.M. Allenby (2007), 'Evaluating the effectiveness of marketing expenditures', Working Paper, University of Frankfurt.

Pashigian, P.B. (1998), *Price Theory and Applications*, 2nd edn, New York: McGraw-Hill.

Rossi, P.E., G.M. Allenby and R. McCulloch (2005), *Bayesian Statistics and Marketing*, Hoboken, NJ: John Wiley & Sons.

Simon, C.P. and L. Blume (1994), *Mathematics for Economists*, New York: Norton.

Sudhir, K. (2001a), 'Competitive pricing behavior in the auto market: a structural analysis', *Marketing Science*, **20** (1), 42–60.

Sudhir, K. (2001b), 'Structural analysis of manufacturer pricing in the presence of a strategic retailer', *Marketing Science*, **20** (3), 244–64.

Varian H.R. (1984), *Microeconomic Analysis*, 2nd edn, New York: Norton.

Villas-Boas, J.M. and R.S. Winer (1999), 'Endogeneity in brand choice models', *Management Science*, **45** (10), 1324–38.

Villas-Boas, J.M. and Y. Zhao (2005), 'Retailer, manufacturers, and individual consumers: modeling the supply side in the ketchup marketplace', *Journal of Marketing Research*, **42** (1), 83–95.

Yang, S., Y. Chen and G.M. Allenby (2003), 'Bayesian analysis of simultaneous demand and supply', *Quantitative Marketing and Economics*, **1**, 251–75.