

Heterogeneity distributions of willingness-to-pay in choice models

Garrett Sonnier · Andrew Ainslie · Thomas Otter

Received: 3 January 2006 / Accepted: 8 March 2007 /
Published online: 9 August 2007
© Springer Science + Business Media, LLC 2007

Abstract We investigate direct and indirect specification of the distribution of consumer willingness-to-pay (WTP) for changes in product attributes in a choice setting. Typically, choice models identify WTP for an attribute as a ratio of the estimated attribute and price coefficients. Previous research in marketing and economics has discussed the problems with allowing for random coefficients on both attribute and price, especially when the distribution of the price coefficient has mass near zero. These problems can be avoided by combining a parameterization of the likelihood function that directly identifies WTP with a normal prior for WTP. We show that the typical likelihood parameterization in combination with what are regarded as standard heterogeneity distributions for attribute and price coefficients results in poorly behaved posterior WTP distributions, especially in small sample settings. The implied prior for WTP readily allows for substantial mass in the tails of the distribution and extreme individual-level estimates of WTP. We also demonstrate the sensitivity of profit maximizing prices to parameterization and priors for WTP.

Keywords Bayesian analysis · Choice modeling · Willingness-to-pay

JEL classification C11 · M31

G. Sonnier
University of Texas at Austin, 1 University Station, Austin, TX 78712, USA
e-mail: garrett.sonnier@mcombs.utexas.edu

A. Ainslie (✉)
UCLA, 110 Westwood Plaza, Los Angeles, CA 90095, USA
e-mail: andrew.ainslie@anderson.ucla.edu

T. Otter
Ohio State University, 2100 Neil Avenue, Columbus, OH 43210, USA
e-mail: otter_2@cob.osu.edu

1 Introduction

Markov chain Monte Carlo (MCMC) techniques to simulate from a distribution have facilitated the exploration of any aspect of the distribution under study, including the distribution of any transformation of random variables. Once a sample of random variables from the distribution is available, change-of-variable calculus is no longer needed to derive the distribution of the transformation. The researcher is free to empirically explore the posterior distribution of the transformation by simply computing the transformation on each iteration of the sampler (e.g., Edwards and Allenby 2003). This technique is referred to as “post-processing” MCMC draws. Posterior summaries of any transformation of parameters are easily obtained from the MCMC output. This can be particularly advantageous when the estimation problem is difficult or intractable in the space of interest, but tractable in another space. Post processing readily allows the researcher to move between the two spaces. While the advantages of post-processing techniques are well-documented, less attention has been paid to the fact that the priors used for the parameters in one space necessarily form an implied prior on the transformed parameters.

It is well known that in a Bayesian model, a change in the likelihood parameterization must be reflected in the prior to leave the posterior predictive density unchanged. Hierarchical models introduce a prior distribution on the parameters across the observational units. Changes in the parameterization of the full conditional likelihood will alter the predictive density of the hierarchical model unless the prior distribution is adapted accordingly. In applied work, the choice of parameterization is often viewed in isolation of the prior distribution, which is typically chosen for analytic convenience (e.g., conjugacy). However, a convenient and diffuse prior in one space does not necessarily result in an equivalent implied prior in the transformed space (Rossi et al. 2005).

An interesting example of transforming model parameters with relevance to marketing and economics occurs when estimating the willingness-to-pay (WTP) for changes in product attributes using choice data. In this paper, we contrast two approaches to estimating the distribution of WTP with choice models. In the first approach WTP is defined as the ratio of attribute and price parameters and the implied prior distribution for WTP is a function of the priors for these parameters. The posterior of the WTP distribution is explored empirically via post-processing. This is consistent with the work of Meijer and Rouwendal (2006), which investigates the properties of WTP defined as a ratio for different distributions of the numerator and the denominator. The second approach re-parameterizes the full conditional likelihood to directly identify WTP. This allows the researcher to directly implement a prior for WTP. We demonstrate the sensitivity of inferences about WTP to different parameterizations in combination with what are regarded as standard assumptions about the hierarchical prior (i.e., the heterogeneity distribution). The sensitivity is particularly pronounced in small sample settings. We show how a normal prior directly specified for WTP results in better inferences. Moreover, not only is the posterior of WTP sensitive to the different parameterization and prior assumptions, but so are all marketing actions derived from the distribution of WTP, such as the setting of profit maximizing prices. The results illustrate the practical importance of paying attention to implied priors.

Implied priors can be problematic even if there is no interest in the transformed parameters themselves. In choice-based conjoint (CBC), for example, the analyst may not be interested in the model coefficients or WTP, per se, but rather in using the model to analyze demand and pricing policies. The reservation and equalization prices, which completely characterize incidence and switching behavior in response to price changes, are a function of attributes and attribute WTPs (Jedidi and Zhang 2002).¹ If the distributions of reservation and equalization prices are impacted by the implied priors on WTP, demand estimates and price policies will be dramatically affected, as we will demonstrate. While certain point estimates of the distribution may be less influenced by the implied prior (e.g., the median versus the mean), any investigation of the nature of consumer demand will require the researcher to consider more than just a particular statistic.

The organization of the remainder of the paper is as follows: Section 2 presents two parameterizations of choice models that result in equivalent full conditional likelihoods. It then discusses how the two parameterizations result in different prior predictive and posterior densities depending on the choice of the prior, particularly once one introduces heterogeneity. Section 3 illustrates the size of the effect using simulated data. Section 4 presents the results from two CBC studies. Section 5 summarizes and offers a brief discussion on the role of prior information in conjoint analysis.

2 Utility and surplus maximization

2.1 Equivalence of likelihood functions

Consider first consumers’ discrete choice problem as that of maximizing an indirect utility function. We have consumers choosing among alternatives on each of occasions. Let V_{ijt}^* denote consumer i ’s indirect utility for alternative j on choice occasion t . It is assumed that indirect utility can be expressed as a linear function of the alternative’s non-price attributes, x_{ijt} , income y_i and price p_{ijt} .

$$V_{ijt}^* = x'_{ijt}\varphi^* + \gamma^*(y_i - p_{ijt}) + \varepsilon^*_{ijt} \text{ with } V_{i0t}^* = \varepsilon^*_{i0t}. \tag{1}$$

We assume the error terms are independent and identically distributed according to a type I extreme value distribution. For exposition, we initially leave the scale parameter as unknown, $\varepsilon^*_{ijt} \sim EV(0, \mu)$. It is well-known that multiplying the indirect utility function for each choice by a constant does not change the utility maximizing alternative. Thus, V_{ijt}^* must be normalized, which is typically accomplished by standardizing the error distribution to be $EV(0, 1)$ such that

$$V_{ijt} = x'_{ijt}\varphi + \gamma(y_i - p_{ijt}) + \varepsilon_{ijt} = \frac{x'_{ijt}\varphi^*}{\mu} + \frac{\gamma^*}{\mu}(y_i - p_{ijt}) + \frac{\varepsilon^*_{ijt}}{\mu}. \tag{2}$$

¹ The reservation price is the price that induces indifference between purchase and non-purchase in the category. The equalization price (Swait et al.1993) is the price that induces indifference between two choice alternatives within the category.

The familiar MNL choice probabilities take the form

$$Pr_{ijt}^u = \frac{\exp \left[\frac{x'_{ijt} \varphi^* - \gamma^* p_{ijt}}{\mu} \right]}{1 + \sum_{m=1}^J \exp \left[\frac{x'_{imt} \varphi^* - \gamma^* p_{imt}}{\mu} \right]} = \frac{\exp [x'_{ijt} \varphi - \gamma p_{ijt}]}{1 + \sum_{m=1}^J \exp [x'_{imt} \varphi - \gamma p_{imt}]} \tag{3}$$

where the superscript u denotes the probability obtained using the utility model. WTP for an improvement in x_{ijkt} , the k th attribute of alternative j , is the price change that would leave the individual indifferent between the alternative with the new level and the alternative with the original level. For continuous x , we have $\partial V_{ijt} = \varphi_k \cdot \partial x_{ijkt} - \gamma \cdot \partial p_{ijt} = 0$ and the change in price that keeps utility constant given a change in attribute k is $\frac{\partial x_k}{\partial p_k} = \frac{\varphi_k}{\gamma}$ (Train 2003). Note that the scale parameter μ drops out of the WTP.

We can re-parameterize the indirect utility function in (1) by dividing through by γ^* .

$$\begin{aligned} \frac{V_{ijt}^*}{\gamma^*} &= x'_{ijt} \frac{\varphi^*}{\gamma^*} + (y_i - p_{ijt}) + \frac{\varepsilon_{ijt}^*}{\gamma^*} \\ C_{ijt} &= x'_{ijt} \beta + (y_i - p_{ijt}) + \eta_{ijt} \end{aligned} \tag{4}$$

In this reparameterization, C_{ijt} is consumer i 's surplus from good j on purchase occasion t (Jedidi et al. 2003). Surplus is determined in part by the attributes of the products in the set, x_{ijts} , and the WTP for the attributes, β . Consumers arrive at their choices by maximizing the surplus (i.e., the difference between the monetary value of the attribute bundle and the price to acquire the bundle) among the J alternatives in a set on occasion t . The MNL choice probability associated with the surplus model is

$$Pr_{ijt}^s = \frac{\exp \left[\frac{x'_{ijt} \beta - p_{ijt}}{\mu} \right]}{1 + \sum_{m=1}^J \exp \left[\frac{x'_{imt} \beta - p_{imt}}{\mu} \right]} \tag{5}$$

where the superscript s denotes the surplus model. The probability expressions in Eqs. (3) and (5) are equivalent over the range of parameters for which the transformations $\beta = \frac{\varphi}{\gamma}$ and $\mu = \frac{1}{\gamma}$ are well defined. In the case of maximum likelihood (ML) estimation, the Invariance Property of the ML estimator ensures that precisely the same point estimates of WTP will be achieved regardless of whether the likelihood is based on (3) or (5) (Cameron and James 1987).

2.2 Bayesian analysis, priors, and posterior distributions for WTP

The ML estimator of the WTP ratio, defined as the ratio of the ML estimates of φ_k and γ , does not possess finite moments and has infinite risk relative to quadratic and many other loss functions (Zellner 1978). In a Bayesian framework the problems associated with the ML estimator are alleviated by the introduction of informative prior distributions. The model thus consists of the full conditional likelihood for the data and the prior distribution for the model parameters. A hierarchical prior distribution defined on the positive real line for γ solves the problem of positive

WTP for a decrease in utility and ensures that the prior and posterior moments of the ratio are finite. The prior and posterior for the ratio is implied by the same for the numerator and denominator.

In the context of random coefficient models, Meijer and Rouwendal (2006) discuss the properties of the WTP ratio for a number of different distributions for φ_k and γ . Only in special cases (e.g., a log-normal distribution for both coefficients) does the ratio of coefficients follow the same distribution as the coefficients. This implies that, generally, the distributional form of the prior used with the likelihood in (5) will differ from that implied by the prior for φ_k and γ . Thus, unlike ML estimates of WTP from the homogenous model, the posterior distribution of WTP formed by mixing the likelihoods in (3) and (5) with priors for the respective coefficients will generally result in distinct posterior WTP distributions and distinct characterizations of demand as a function of price. What discrepancy can we expect from these two approaches? To the extent that the data overwhelm the prior, the posterior WTP distributions from the two approaches will converge despite the differences in the prior. This will generally happen for models that impose homogeneity on the coefficients. The more interesting case occurs with hierarchical models.

In hierarchical models, we typically encounter many units (e.g., consumers) and relatively few observations per unit. Thus, the full conditional likelihood of any one consumer is informed by a limited amount of data and the prior distribution will generally have much more influence on the posterior compared with a homogenous model. A hierarchical model may build on either parameterization, using either of the likelihood functions in Eqs. (3) or (5) as the full conditional likelihood.² The likelihood in Eq. (3) in combination with a (hierarchical) prior for γ_i that has positive density arbitrarily close to zero readily accommodates respondents that do not appear to be sensitive to price. Such respondents can in turn have a tremendous influence on the posterior WTP distribution implied by the model and thus on any characterization of demand as a function of price. Hierarchical models with likelihood functions built on Eq. (5), measure WTP directly by β_i . An advantage of this formulation is that a hierarchical prior for WTP can be specified directly.³ For example, a normal prior for β_i will place less mass on absolutely large WTP values.

The problems we outline with the WTP ratio are neither unique to choice models nor WTP. They apply to any quantity that can be defined as a ratio of model parameters. However, estimation of WTP (and the related concept of reservation price) is a particularly relevant problem in marketing and economics. Recently, the marketing literature has sharpened its focus on the study of WTP and reservation prices because of the direct implications for pricing strategy (Jedidi and Zhang 2002; Jedidi et al. 2003; Shaffer and Zhang 1995, 2000). The economics literature has recognized the potential problems with random coefficient ratio estimates of WTP (Meijer and Rouwendal 2006; Revelt and Train 1998). Marketing practitioners have

² We adopt a fully Bayesian approach to inference in this paper. However, the points made apply in the context of hierarchical models independent of the estimation technique.

³ Another motivation for directly parameterizing the model in WTP terms is that the researcher often has variables such as demographics that may be useful in the characterization of consumer heterogeneity. In such instances it seems more reasonable to build hierarchical regression structures for WTP instead of parameters with less clear interpretation. We do not explore this issue here.

also recognized the problems, advocating use of the median as a summary of the posterior WTP distribution (Orme 2001). While the median will likely be a more robust statistic, Bayesian decision theoretic analyses of demand as a function of price aimed at identifying optimal actions rely on the entire posterior distribution of WTP. To the extent that the posterior distribution of WTP is sensitive to the prior assumptions, so will the optimal action.

2.3 Optimal pricing

Ignoring the implied prior on WTP can adversely impact demand and price analyses. Firms often use the model coefficients estimated from CBC data to build market share simulators, which are useful for assessing response to price changes and optimal pricing. Given a set of non-price attributes, market share (and demand) can be completely characterized by consumer surplus. Consumers choose the inside alternative that yields the maximum surplus and forgo a category purchase if the surplus from the best alternative is less than the surplus generated by the outside alternative. The price that determines the incidence and choice decisions is the reservation price, \tilde{p}_{ijt} , which induces indifference between buying alternative j and forgoing a category purchase. For the surplus model, $\tilde{p}_{ijt} = x_{ijt}\beta_i + (\eta_{ijt} - \eta_{i0t})$. Importantly, any proper indirect utility function implies a function for \tilde{p}_{ijt} . In our case, the reservation price from the utility model is $\tilde{p}_{ijt} = \frac{x_{ijt}\varphi_i + (\varepsilon_{ijt} - \varepsilon_{i0t})}{\gamma_i}$. From this equation, we can see that the change in the reservation price given a change in an attribute is given by the WTP for that attribute.

If the no-buy option is not included in the CBC experiments, the reservation price is not identified. In this case, what we can identify is the equalization price \tilde{p}_{ijt} which is the price for good j that equalizes the surplus generated by goods j and j' . For the surplus model, $\tilde{p}_{ijt} = (x_{ijt} - x_{ij't})\beta_i + \tilde{p}_{ij't}(\eta_{ijt} - \eta_{ij't})$. Again, any proper indirect utility function implies a function for \tilde{p}_{ijt} . In our case, $\tilde{p}_{ijt} = \frac{(x_{ijt} - x_{ij't})\varphi_i + \gamma_i p_{ij't} + (\varepsilon_{ijt} - \varepsilon_{ij't})}{\gamma_i}$.

Consider now using the utility or surplus models to find the profit maximizing price for firm j , taking the competing firms' prices as given. To the extent that the utility model a priori puts greater mass on extreme WTP values, the posterior distribution of reservation and equalization prices will also be thick tailed. This is especially so in sparse data environments, and implies that the firm could continue to raise prices and still find consumers willing to purchase. Thus, inference about the profit maximizing prices based on the posterior distributions of the parameters will depend on the model.

3 A simulation study

We more closely investigate the properties of the two approaches to WTP estimation in the following simulation study. We generate four data sets, two each from the utility and surplus models, which we will refer to as D1, D2, D3, and D4. For the utility model data sets, D1 and D2, we assume the following population distribution, $\Phi_i \sim N(\bar{\Phi}, \Sigma_\Phi)$, where $\Phi_i = [\varphi'_i \ \log(\gamma_i)]'$. For both D1 and D2, the covariance matrix Σ_θ is assumed to be diagonal and we choose parameters such that the distribution of γ_i is centered near 1. For D1, we allow for some mass of the

Table 1 Data generating parameters, utility model data sets

Data set	D1		D2	
	Mean	Variance	Mean	Variance
φ_{i1}	-0.5	1	-0.5	1
φ_{i2}	1	1	1	1
φ_{i3}	1.5	1	1.5	1
φ_{i4}	0.5	1	0.5	1
φ_{i5}	0.75	1	0.75	1
φ_{i6}	1	1	1	1
Log (γ_i)	-1	2	0	0.2

distribution of γ_i to be near zero by choosing a large value for the variance of log (γ_i). For D2, we choose the variance of log (γ_i) such that γ_i is tightly distributed around one, with little to no mass near zero. In the case of the former, some individual-level WTPs will be extremely large for values of $\gamma_i \rightarrow 0$ while in the latter, the distribution of WTP should be closer to normal. For the datasets generated by the surplus model, D3 and D4, we assume $\theta_i \sim N(\bar{\theta}, \Sigma_\theta)$ where $\theta_i = [\beta'_i \ \log(\mu_i)]'$. In this case, the distribution of WTP is specified directly. For D3, we choose parameters such that μ_i is, on average, larger and the deterministic component of surplus has relatively lower explanatory power. For D4, we choose parameters such that μ_i is, on average, smaller, translating into more extreme choice probabilities.

For all models, we assume 300 individuals choosing amongst three alternatives and an outside good on each of 15 choice occasions. The covariates include three alternative specific constants, a discrete attribute with four levels, and a price. Each alternative is created by randomly choosing a level of the discrete attribute and a price from the range [1.5–2.5] (in increments of 0.1). Tables 1 and 2 contain the parameters of the distributions used to generate the four data sets. We retain the last choice of each simulated respondent to create a holdout sample. Using MCMC methods, we estimate the utility and surplus models on each of the four data sets, for a total of eight sets of results. The details of the sampler have been reported elsewhere (e.g., Allenby and Lenk 1994; Arora et al. 1998; Train 2003). We use a normal-inverted Wishart hyper-prior structure for the population distribution parameters $\bar{\theta}$ and Σ_θ . The prior on $\bar{\theta}$ is set to $N(0_{(K)}, 10^6 \times I_{(K \times K)})$. The prior on Σ_θ is set to $IW(K + 1, I_{(K \times K)})$. These are proper but diffuse priors. We use identical priors for $\bar{\Phi}$ and Σ_Φ in the utility model.

Table 2 Data generating parameters, surplus model data sets

Data set	D3		D4	
	Mean	Variance	Mean	Variance
β_{i1}	-0.5	1	-0.5	1
β_{i2}	1	1	1	1
β_{i3}	1.5	1	1.5	1
β_{i4}	0.5	1	0.5	1
β_{i5}	0.75	1	0.75	1
β_{i6}	1	1	1	1
log (μ_i)	1	0.1	-1	0.5

Table 3 Root mean squared error

Data set	D1		D2		D3		D4	
Data generation	Utility		Utility		Surplus		Surplus	
Model	Utility	Surplus	Utility	Surplus	Utility	Surplus	Utility	Surplus
WTP ₁	48.54	45.68	2.43	1.55	26.06	1.60	1.52	1.19
WTP ₂	67.61	64.82	1.53	1.30	16.98	1.36	1.04	0.85
WTP ₃	91.83	91.25	1.42	1.19	16.31	1.36	1.06	0.72
WTP ₄	26.78	24.82	1.74	1.28	16.90	1.60	1.33	0.95
WTP ₅	41.58	38.96	1.68	1.33	16.97	1.46	1.37	0.95
WTP ₆	33.42	29.87	1.65	1.27	19.16	1.73	1.13	0.83
Average RMSE	51.63	49.23	1.74	1.32	18.73	1.52	1.24	0.92

For the utility models, we compute the individual-level WTPs as $\frac{\alpha_i}{\gamma_i}$ on each iteration of the sampler. For the surplus models, draws of the individual-level WTPs are directly available. We compute the mean absolute error (MAE) and the root mean-squared error (RMSE) between the true and estimated WTPs on each iteration of the sampler and report the means over iterations. Using the harmonic mean estimator (Newton and Raftery 1994), we compute the log marginal density (LMD) statistic for each model. We also report the deviance information criteria (DIC) (Spiegelhalter et al. 2002) and the log predictive density (LPD) of the holdout data.

Tables 3, 4 and 5 presents the results of our simulation study. D1 and D2 are generated according to the heterogeneous utility model. D1 contains individuals with price coefficients near zero and thus extremely large WTPs. Relative to the other conditions, the error statistics are quite high in this setting. As evidenced by smaller RMSE and MAE, note that the surplus model has more accurate recovery of the true WTPs, even though the utility model is consistent with the data generating process. In terms of fit statistics, the LMD, DIC and LPD all favor the utility model. In D2, the distribution of the price coefficient has most of its mass away from zero. Again, the surplus model has lower RMSE and MAE. The LMD and LPD favor the utility model, while the DIC favors the surplus model. Thus, even when the true WTPs are a ratio of random coefficients, the surplus model more accurately recovers the true WTPs compared with the utility model under a range of population distribution parameters.

Table 4 Mean absolute error

Data Set	D1		D2		D3		D4	
Data generation	Utility		Utility		Surplus		Surplus	
Model	Utility	Surplus	Utility	Surplus	Utility	Surplus	Utility	Surplus
WTP ₁	11.88	10.25	1.52	1.21	14.64	1.29	1.10	0.94
WTP ₂	14.85	12.32	1.05	1.01	9.15	1.08	0.75	0.66
WTP ₃	18.01	15.92	0.97	0.89	8.44	1.08	0.73	0.56
WTP ₄	8.58	7.10	1.17	1.00	9.05	1.27	0.94	0.74
WTP ₅	10.55	8.64	1.16	1.03	9.14	1.16	0.97	0.75
WTP ₆	10.91	8.95	1.13	0.95	10.49	1.38	0.82	0.66
Average MAE	12.46	10.53	1.17	1.01	10.15	1.21	0.89	0.72

Table 5 Model fit statistics

Data set	D1		D2		D3		D4	
Data generation	Utility		Utility		Surplus		Surplus	
Model	Utility	Surplus	Utility	Surplus	Utility	Surplus	Utility	Surplus
LMD	-3532.20	-3553.00	-3942.80	-3946.30	-5309.60	-5428.90	-2493.40	-2340.50
DIC	3859.20	3871.10	4324.80	4318.30	5570.70	5497.90	2913.60	2751.50
LPD	-287.47	-296.64	-322.79	-331.15	-410.41	-403.82	-236.34	-231.89

D3 and D4 are generated with the heterogeneous surplus model. In D3, the true scale parameter μ_i is, on average, larger. In this setting, the utility model estimates of WTP are particularly error-prone. Once more, the surplus model is better at recovering the true WTP parameters. Interestingly, the LMD statistic favors the utility model, despite the lack of recovery of the true WTPs.⁴ The DIC and LPD favor the surplus model. In D4, the scale parameter is, on average, smaller. Relative to D3, the utility model does a better job of recovering the WTPs here, but again, the surplus model has more accurate WTP recovery. All three of the fit statistics favor the surplus model.

In summary, the surplus models always recover the true WTPs with more accuracy, regardless of the data generating mechanism. Even when the true WTPs are distributed as a ratio of random coefficients, the ratio estimator does not recover the true WTPs as accurately as simply directly specifying a prior on WTP. We attribute this to the fact that the surplus model employs a more sensible prior distribution for WTP.

4 Two CBC studies

Using CBC data sets provided to us by firms in the camera and automotive categories, we replicate the findings from our simulation study in the sense that the posterior of WTP from the utility model is rather different from the posterior obtained from the surplus model. Moreover, inferences obtained with the utility model lack face validity. Table 6 presents the attributes and levels involved in the design of each study.

4.1 Data and models

The first data set is CBC data on midsize sedans. The data were provided by a major automotive manufacturer. Respondents qualified for participation in the study on the basis of the vehicle they currently own, their intention to purchase a midsize sedan, and other socio-economic information. A total of 333 respondents participated in the study. Each respondent completed 15 choice tasks, with each task consisting of three sedans. The no-buy option was not included in this study. The second data set is

⁴ This is similar to previous simulation studies in the literature that find the harmonic mean estimator of the LMD sometimes favors models with relatively poor parameter recovery (Andrews et al. 2002; Liechty et al. 2005).

Table 6 Attributes and levels

Camera data		Sedan data	
Attribute	Levels	Attribute	Levels
Body Style	Low Medium High	Make/Model	Ford Taurus Toyota Camry Nissan Maxima Honda Accord VW Passat
Mid-roll change ^a	None Manual Automatic	Engine	4 cylinder; 1.8 L; 150 HP 4 cylinder; 2.4 L; 160 HP 6 cylinder; 3.0 L; 155 HP 6 cylinder; 3.0 L; 222 HP
Annotation ^a	None Pre-set List Customized List Custom Input Method 1 Custom Input Method 2 Custom Input Method 3	Audio and navigation	Standard Audio Premium Audio Premium Audio with Navigation
Camera operation feedback ^a	No Yes	Antilock Brakes (ABS)	No Yes
Zoom	None 2X 4X	Side Door/Window Curtain Airbags(CAB)	No Yes
Viewfinder	Regular Large	Vehicle Skid Control (VSC)	No Yes
Camera settings feedback	None LCD Viewfinder LCD and Viewfinder		
Price (nested within camera type)	from \$41 to \$499	Price	\$17,400 \$18,900 \$20,400 \$21,900 \$23,400 \$24,900 \$26,400

^a feature only available on APS

CBC data on cameras. The study was conducted by the Eastman Kodak Company to assess the market for a new camera format, the Advanced Photo System (APS). A detailed description of the data is given by Gilbride and Allenby (2004). A total of 302 respondents participated in the study. Each respondent completed 14 choice tasks, with each task consisting of three 35 mm cameras, three APS cameras, and a no-buy option. Some attributes were available only on the APS camera, and price was nested within camera type.

For both data sets, we model consumer i 's surplus for alternative j at choice occasion t as a linear function of non-price attributes, attribute WTPs, and price

$$\begin{aligned}
 C_{ijt} &= x'_{ijt}\beta_i - p_{ijt} + \varepsilon_{ijt} \quad \varepsilon_{ijt} \sim EV(0, \mu_i) \\
 \theta_i &\sim N(\bar{\theta}, \Sigma_\theta) \\
 \theta_i &= [\beta'_i \quad \log(\mu_i)]'
 \end{aligned}
 \tag{6}$$

For the camera data, we set the deterministic component of the surplus for the no-buy option to zero. For identification, the lowest level of each attribute is dropped (with the exception of the body type attribute since the baseline is the “no-buy” option). We use the negative of price (in \$100s) in the likelihood. The coding scheme is the same as that employed by Gilbride and Allenby (2004), and results in a total of $K=18$ parameters. For the sedan data, the make/model “VW Passat” is dropped, as are the lowest level of each of the remaining non-price attributes. This results in a total of $K=13$ parameters. We use the negative of price (in \$1,000s) in the likelihood.

For both data sets, we compare estimates of the distribution of WTP from the surplus model with that of the linear utility model, where θ_i is replaced with $\Phi_i = [\varphi'_i \quad \log(\gamma_i)]'$. Here, the choice probabilities are based on (3). The same normal-inverted Wishart hyper-prior structure used for $\bar{\theta}$ and Σ_θ is used for hyper-priors on the population parameters $\bar{\Phi}$ and Σ_Φ . The linear utility model requires we calculate the WTP from the model parameters using the ratio transformation. On each iteration of the sampler, we compute the ratio $\frac{\varphi_i}{\gamma_i}$ using the draws of the individual level parameters. We then compute the mean, median, and standard deviation over individuals, and report the mean of these quantities over iterations of the sampler. For both data sets, the samplers are run for 20,000 iterations. We keep the last 5,000 iterations for posterior inference. Parameter estimates are calculated with $T-1$ choice tasks. We keep the last task for each individual to assess holdout performance via LPD. To assess in-sample performance, we compute the LMD and DIC statistic for each model.

4.2 Results

In Tables 7 and 8 we report the mean and standard deviation of the distribution of WTP for the utility and surplus models. Posterior standard deviations of the reported statistics are in parentheses. Table 7 contains the results from the sedan data while Table 8 contains the results from the camera data. The mean and standard deviation of the population distribution of WTP are dramatically affected by the priors. For the sedan data, the means of the utility model estimates are two to three times the magnitude of the surplus model. The WTP distributions are also far more dispersed, with standard deviations that are five to six times larger. For the camera data, the means are also much larger for the utility model. However, most of the standard deviation estimates for the utility model have large posterior standard deviations. For both data sets, the median of the population distribution is much less sensitive to the prior than the mean or standard deviation.

The in-sample fit statistics are somewhat mixed. In both data sets, the LMD strongly favors the utility model. This result echoes that of the third synthetic data

Table 7 WTP estimates for sedan data (standard errors in parentheses)

Attribute	WTP (\$1,000's)	Utility model			Surplus model		
		$\frac{\psi_i}{\gamma_i}$			β_i		
	Level	Mean	Median	Std dev	Mean	Median	Std dev
Make/Model	Ford Taurus	-5.02 (2.76)	-2.20 (0.52)	46.54 (12.80)	-2.23 (0.32)	-2.27 (0.43)	10.06 (0.29)
	Toyota Camry	16.23 (3.43)	6.76 (0.58)	47.87 (15.86)	7.00 (0.23)	6.74 (0.36)	9.73 (0.37)
	Nissan Maxima	12.42 (3.46)	4.82 (0.52)	46.88 (15.92)	5.13 (0.27)	5.12 (0.42)	9.35 (0.40)
	Honda Accord	9.25 (3.06)	3.70 (0.48)	42.10 (13.68)	3.96 (0.26)	3.81 (0.31)	8.48 (0.25)
Engine	4-cyl; 2.4L; 160HP	4.88 (1.12)	1.98 (0.39)	14.27 (5.50)	1.82 (0.37)	1.82 (0.37)	2.00 (0.25)
	6-cyl; 3.0L; 155HP	7.89 (1.33)	3.78 (0.49)	16.61 (5.65)	3.67 (0.34)	3.60 (0.33)	3.48 (0.36)
	6-cyl; 3.0L; 222HP	9.89 (1.39)	5.21 (0.51)	20.25 (6.85)	4.87 (0.23)	4.78 (0.24)	3.95 (0.38)
Audio	Premium Audio	2.52 (1.02)	1.10 (0.26)	9.29 (3.61)	1.34 (0.21)	1.31 (0.20)	1.56 (0.21)
	Premium Audio w/Nav	3.54 (1.12)	1.57 (0.26)	10.54 (4.20)	1.61 (0.24)	1.62 (0.25)	2.02 (0.21)
Safety features	Antilock Brakes	4.78 (1.05)	2.11 (0.26)	12.23 (4.20)	2.08 (0.20)	2.07 (0.20)	2.21 (0.18)
	Side Curtain Airbags	2.89 (0.88)	1.22 (0.26)	9.36 (3.46)	1.18 (0.19)	1.17 (0.21)	1.52 (0.27)
	Vehicle Skid Control	2.83 (0.82)	1.28 (0.23)	9.64 (3.59)	1.29 (0.26)	1.28 (0.26)	1.42 (0.24)
	Fit statistics						
	LMD	-2220.4			-2665.8		
	DIC	2,966.4			3,044.3		
	LPD	-340.7			-333.5		

set, D3, which was generated by the surplus model. In this case, the LMD strongly favored the utility model despite its inconsistency with the data generating process and its extremely poor parameter recovery. The DIC favors the utility model in the sedan data and the surplus model in the camera data. In contrast to the in-sample fit measures, the LPD favors the surplus model in both the sedan data and the camera data, indicating the surplus model has superior out-of-sample performance. We will now examine more closely the distribution of WTP and optimal prices implied by the two models. From this vantage point, the differences across the two models are less ambiguous.

It is evident that the utility and surplus models result in dramatically different estimates of the distribution of WTP. The utility model estimates seem to be implausible and not reflective of consumers' monetary valuation of product attributes. Figure 1 presents boxplots of the individual-level make/model WTP estimates for both the utility and surplus models. These are measured relative to the VW Passat and can be interpreted as equalization prices; the relative price difference that equalizes the utility of comparably equipped competitive sedans and the Passat. The median of the utility model's individual level WTP estimates for the three

Table 8 WTP estimates for camera data (standard errors in parentheses)

Attribute	WTP (\$100's)	Utility model			Surplus model		
		$\frac{\varphi_i}{\gamma_i}$			β_i		
	Level	Mean	Median	Std dev	Mean	Median	Std dev
Body style	Low	-25.78 (6.74)	-4.67 (0.82)	84.40 (54.41)	-5.47 (0.38)	-5.51 (0.43)	4.90 (0.36)
	Medium	-10.15 (4.52)	0.49 (0.28)	56.60 (44.88)	1.07 (0.22)	1.14 (0.25)	3.59 (0.38)
	High	-7.82 (3.83)	0.33 (0.27)	48.09 (36.43)	0.85 (0.27)	0.80 (0.31)	3.78 (0.64)
Mid-roll change	Manual	1.96 (1.55)	0.44 (0.24)	14.02 (6.12)	-0.36 (0.50)	-0.36 (0.48)	2.84 (0.26)
	Automatic	3.93 (1.21)	0.52 (0.12)	16.99 (10.99)	0.80 (0.12)	0.71 (0.14)	2.40 (0.14)
Annotation	Pre-set list	1.51 (0.60)	0.38 (0.13)	7.32 (3.39)	0.34 (0.08)	0.33 (0.09)	1.01 (0.12)
	Customized list	3.25 (0.76)	0.98 (0.16)	9.47 (4.80)	1.21 (0.11)	1.19 (0.12)	1.09 (0.14)
	Custom Input Method 1	4.10 (1.62)	-0.28 (0.17)	22.63 (16.02)	-1.44 (0.24)	-1.48 (0.28)	3.37 (0.24)
	Custom Input Method 2	4.64 (1.12)	1.06 (0.16)	14.71 (8.49)	1.23 (0.11)	1.22 (0.13)	1.73 (0.18)
	Custom Input Method 3	2.09 (1.35)	-0.78 (0.15)	19.75 (11.85)	-0.99 (0.31)	-1.03 (0.32)	2.58 (0.28)
Operation feedback	Feedback	2.65 (0.86)	0.58 (0.12)	10.50 (4.85)	0.85 (0.25)	0.82 (0.25)	1.72 (0.10)
Zoom	×2 Zoom	5.47 (1.52)	1.95 (0.25)	14.96 (8.04)	2.70 (0.18)	2.70 (0.19)	2.09 (0.15)
	×4 Zoom	7.66 (2.12)	2.50 (0.41)	20.70 (10.42)	3.11 (0.24)	3.13 (0.26)	3.13 (0.33)
Viewfinder	Large viewfinder	0.40 (0.75)	-0.18 (0.10)	8.91 (4.23)	-0.47 (0.17)	-0.46 (0.16)	1.76 (0.22)
Settings feedback	LCD	4.30 (1.19)	1.12 (0.20)	13.27 (8.30)	-0.23 (0.16)	-0.23 (0.16)	1.30 (0.16)
	Viewfinder	4.49 (1.07)	1.20 (0.19)	13.42 (7.99)	-0.16 (0.16)	-0.15 (0.16)	1.59 (0.19)
	LCD and Viewfinder	5.33 (1.24)	1.51 (0.24)	14.92 (9.23)	0.21 (0.17)	0.21 (0.18)	1.53 (0.16)
	Fit statistics						
	LMD	-3879.7			-4100.4		
	DIC	4,633.6			4,620.4		
	LPD	-438.5			-434.7		

Japanese make/models are near or in excess of the range of prices shown to respondents.⁵ This is not being caused by a just a handful of respondents with estimates of γ_i near zero. The 75th percentiles for the individual-level equalization price between Toyota Camry vs. VW Passat and Nissan Maxima vs. VW Passat are approximately \$30,428 and \$24,684, respectively. The retail price of the Passat is about \$23,000. This implies that a quarter of the respondents would require Passat to

⁵ Note that the median of the posterior mean of the individual-level estimates will not be the same as the posterior mean of the median of the population distribution.

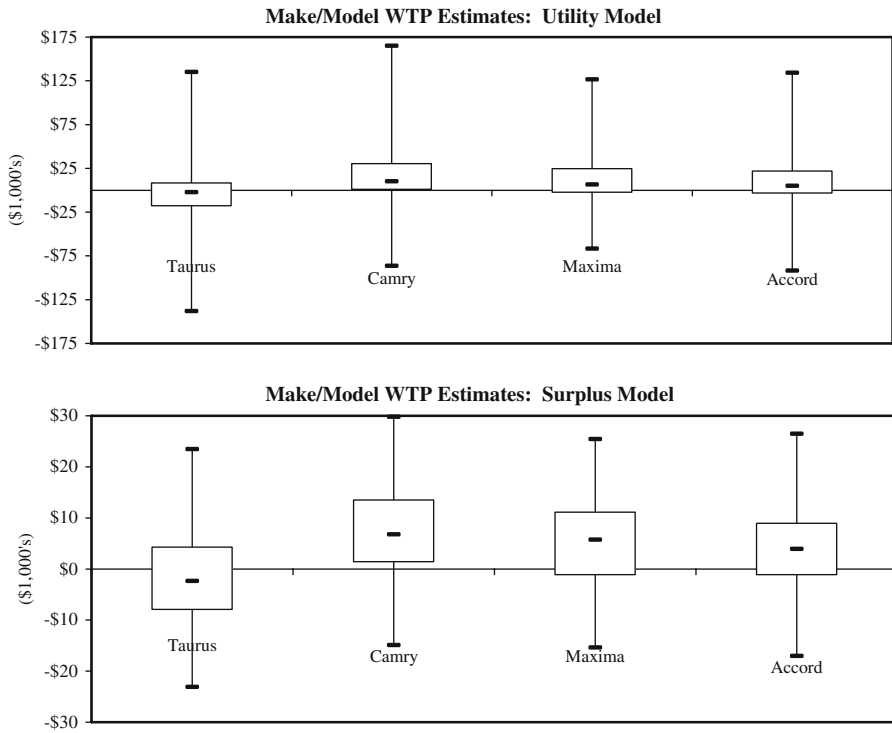


Fig. 1 Boxplots of individual-level make/model WTP estimates

Table 9 Attributes and levels for optimal pricing, sedan data

Attribute	Levels	Attributes and levels for alternatives				
		1	2	3	4	5
Make/Model	Ford Taurus	X				
	Toyota Camry		X			
	Nissan Maxima			X		
	Honda Accord				X	
	VW Passat					X
Engine	4 cylinder; 1.8 L; 150 HP					X
	4 cylinder; 2.4 L; 160 HP		X		X	
	6 cylinder; 3.0 L; 155 HP	X				
	6 cylinder; 3.0 L; 222 HP			X		
Audio and navigation	Standard Audio	X	X	X	X	X
	Premium Audio					
	Premium Audio with Navigation					
Antilock brakes	No			X		
	Yes	X	X		X	X
Side door/Window curtain airbags	No					
	Yes	X	X	X	X	X
Vehicle skid control	No			X		
	Yes	X	X		X	X
Price	(\$1,000)	\$20.1	\$20.7	\$24	\$20.5	\$22.5

Table 10 Attributes and levels for optimal pricing, camera data

Attribute	Levels	Attributes and levels for alternatives		
		1	2	3
Body style	Low	X		
	Medium		X	
	High			X
Mid-roll change	None			
	Manual	X		
Annotation	Automatic		X	X
	None			
	Pre-Set List	X		
	Customized List			
	Custom Input Method 1		X	
Camera operation feedback	Custom Input Method 2			X
	Custom Input Method 3			
	No	X		
	Yes		X	X
Zoom	None			
	2×	X		
	4×		X	X
Viewfinder	Regular	X	X	
	Large			X
Camera settings feedback	None			
	LCD			
	Viewfinder			
	LCD & Viewfinder	X	X	X
Price (nested within camera type)	from \$41 to \$499	\$100	\$225	\$400

have a zero price as well as a cash subsidy to induce indifference with a similarly equipped Camry or Maxima, which does not seem credible.

In the camera study, the individual-level estimates of WTP implied by the utility models also seem lacking in face validity. For example, the surplus model estimate of the median of the individual-level WTP estimates for a 2× zoom lens is \$295, with demand essentially zero at prices exceeding \$550. According to the utility model, the median of the individual level WTPs is \$322. At a price of \$550, 32% of respondents are still in the market. A quarter of respondents have WTP estimates in excess of \$750. Demand does not reach zero until prices exceed \$3,000. These estimates of WTP seem unreasonably high. Furthermore, any analysis of demand should take into account the uncertainty in the individual-level estimates. We now turn our attention to such an analysis.

Table 11 Market shares, sedan scenario

	Taurus	Camry	Maxima	Accord	Passat
Price	\$20.1	\$20.7	\$24	\$20.5	\$22.5
Utility (%)	16	36	19	21	8
Surplus (%)	16	36	18	22	9

Table 12 Market shares, camera scenario

	Camera 1	Camera 2	Camera 3	No Buy
Price	\$100	\$225	\$400	\$0
Utility (%)	17	35	32	16
Surplus (%)	14	28	35	23

4.3 An optimal pricing exercise

For the utility model, profits from alternative j in scenario z can be written as

$$\pi_j^{uz} \mid \Phi_i, x_j^z, p_j = \left(P_j^{uz} \mid \Phi_i, x_j^z, p_j \right) \times (p_j - c_j) \tag{7}$$

We seek the price p_j^{uz*} that maximizes the firm’s expected profit, $E_{\Phi} \left[\pi_j^{uz} \mid \Phi_i, x_j^z, p_j \right]$. The expected profit in scenario z is easily calculated with the output of the Gibbs sampler. For a given price, we simply average the profits calculated over the draws of Φ_i . Using routine optimization procedures, it is straightforward to find the optimal price. For the surplus model, profits from alternative j in scenario z can be written as

$$\pi_j^{sz} \mid \theta_i, x_j^z, p_j = \left(P_j^{sz} \mid \theta_i, x_j^z, p_j \right) \times (p_j - c_j). \tag{8}$$

As with the utility model, we seek the price p_j^{sz*} that maximizes the firm’s expected profit, $E_{\theta} \left[\pi_j^{sz} \mid \theta_i, x_j^z, p_j \right]$.

Our goal is to compare p_j^{uz*} and p_j^{sz*} . Tables 9 and 10 present the attributes and levels used to construct the competitive scenarios for our pricing exercise. In the sedan data, we consider a competitive set consisting of five sedans. In the camera data, we consider a competitive set consisting of three cameras. Tables 11 and 12 present the prices and market shares for each alternative for the sedan and camera scenarios. On each iteration of the sampler, we compute P_j and report the mean over iterations. For the sedan data, the two models predict practically the same shares. For the camera data, there is some disagreement, with the utility model predicting higher shares for Cameras 1 and 2 and lower shares for Camera 3 and the No-Buy alternative. For the sedan data, we will find the optimal price for Ford Taurus, assuming the competitive vehicle prices remain at their current levels. For the camera data, we will

Table 13 Taurus optimal price, sedan scenario

	Utility model				
	Taurus*	Camry	Maxima	Accord	Passat
Price(\$1,000)	\$33.2	\$20.7	\$24	\$20.5	\$22.5
Share (%)	3	41	22	25	9
	Surplus model				
	Taurus*	Camry	Maxima	Accord	Passat
Price	\$25.8	\$20.7	\$24	\$20.5	\$22.5
Share (%)	6	40	20	24	10

* denotes optimized product

Table 14 Camera 3 optimal price, camera scenario

	Utility model			
	Camera 1	Camera 2	Camera 3*	No Buy
Price	\$100	\$225	\$1,525	\$0
Share (%)	19	35	10	21
	Surplus model			
	Camera 1	Camera 2	Camera 3*	No Buy
Price	\$100	\$200	\$522	\$0
Share	15%	33%	27%	25%

* denotes optimized product

find the optimal price for Camera 3, assuming the competitive camera prices remain at their current levels. To conduct the exercise, we need to make some assumptions on costs. For simplicity, we assume the sedans are all built at a variable cost of \$18,000. For the cameras, we assume variable costs of \$50, \$60, and \$70 for Cameras 1, 2, and 3, respectively. Similar results were obtained using other sedans and cameras in the competitive scenarios, as well as other cost assumptions.

Tables 13 and 14 present the findings from the optimal pricing exercise. We present the optimal price for Ford Taurus and Camera 3 along with the new market shares. For the sedan data, using the utility model coefficients in the optimization results in an optimal price for Taurus of \$33,200. At this price, the largest relative price difference is \$12,500, observed between Taurus and Camry. The largest relative price difference shown in the experiments is \$9,000. The prior implied by the utility model supports excessive equalization prices, leading to optimized prices beyond the empirical range of prices in the data. In contrast, optimization based on the surplus model leads to an optimal price for Taurus of \$25,800. The largest relative price difference is well within the range of experimental prices. We obtain similar results from the camera data. Using the utility model, the optimal price for Camera 3 is over \$1,500. The maximum price shown to respondents in the study was \$499. For the camera data, using the surplus model results in an optimal price of about \$520. While this is slightly in excess of the maximum price, it is much more reasonable.

5 Summary and conclusions

Researchers in marketing and economics have recognized the problems associated with using random coefficient choice models derived from linear indirect utility functions to estimate WTP for product attributes. In this setting, WTP is estimated via the ratio of attribute and price coefficients. We illustrate that the prior implied for WTP by seemingly reasonable priors for the attribute and price coefficients results in posterior WTP distributions with extremely fat tails. This also affects the model's characterization of demand which has implications for pricing analyses. A number of ad-hoc solutions have been proposed, including constraining the price coefficient to be homogenous, or using the median as a measure of central tendency of the WTP

distribution. In this paper, we present a straightforward solution to the problems caused by the implied prior for WTP. Parameterizing the choice model in the space of consumer surplus allows for direct specification of a prior distribution for WTP. Such a direct specification is especially advantageous in the context of hierarchical models where the aforementioned solutions conflict with the purpose and value of quantifying consumer heterogeneity.

Using both simulated data and CBC data sets from the automotive and camera categories, we document the influence of the implied prior for WTP. Commonly employed diffuse priors for the attribute and price coefficients put too much prior mass on extreme WTP values to render reasonable posterior WTP distributions in small sample settings. Some posterior summaries are less sensitive to the assumed prior (e.g. median versus mean). However, marketing actions, such as setting profit maximizing prices, depend on the entire posterior distribution of WTP and thus will be sensitive to the implied prior. In the surplus parameterization a hierarchical prior for WTP can be directly specified. We found a hierarchical normal prior to be useful in controlling the tails of the WTP distribution. The relatively thinner tails of the normal result in more reasonable estimates of the WTP distribution and, in turn, profit-maximizing prices.

The surplus model results in more reasonable estimates of the distribution of WTP and profit-maximizing prices as well as superior out-of-sample performance. However, the in-sample fit statistics across the two parameterizations are ambiguous, even with simulated data. We leave this issue, specifically the performance of the Newton-Raftery estimator of the LMD and the DIC statistic as criteria for model choice, to future research. We acknowledge the existence of data generating mechanisms that leave respondent WTP for a particular attribute level inestimable. Among these are non-compensatory processing, price based quality inferences or the simple ignorance of the price attribute in the conjoint exercise. The utility model with standard priors will readily accommodate respondents who are, for whatever reason, insensitive to price in the conjoint task. The modeling question then becomes one of *how* and *whether* to implement prior knowledge about the range of likely WTP values. We have demonstrated that the surplus model is very effective in terms of *how* to implement such prior knowledge because it allows the researcher to put a prior directly on WTP.

Whether to implement prior knowledge about WTP in conjoint studies, especially when the data are better fit with arbitrarily large WTP values, touches upon the core of the inferential problems associated with conjoint experiments in marketing. Conjoint data are collected with the implicit goal of characterizing market demand. To the extent that the conjoint likelihood differs from the likelihood that generates choices in the market place, this generalization calls for the diligent use of prior knowledge held by the researcher about market behavior. That is, the prior should preserve certain well-known aspects of the target environment in the posterior and still be informed by the conjoint likelihood in other respects. We acknowledge that our argument here is limited to forming prior-predictive distributions given the conjoint data and other prior knowledge. In the long run, only a better understanding of the actual data generating mechanism underlying the conjoint data will enable researchers to develop the necessary procedural modifications to move it closer to the likelihood that generates choices in the market.

Acknowledgement The authors would like to thank Peter Rossi, JP Dubé, Jordan Louviere, Kenneth Train, and Greg Allenby for helpful insights. We also thank seminar participants at The Ohio State University, Duke University, University of Michigan and the University of Chicago for providing useful comments.

References

- Allenby, G., & Lenk, P. (1994). Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association*, 89, 1218–1231.
- Andrews, R., Ainslie, A., & Currim, I. (2002). An empirical comparison of logit choice models with discrete vs. continuous representations of heterogeneity. *Journal of Marketing Research*, 39, 479–487.
- Arora, N., Allenby, G. M., & Ginter, J. L. (1998). A disaggregate model of primary and secondary demand. *Marketing Science*, 17(1), 29–44.
- Cameron, T., & James, M. (1987). Estimating willingness-to-pay from survey data: An alternative pre-test-market evaluation procedure. *Journal of Marketing Research*, 24, 389–395.
- Edwards, Y., & Allenby, G. (2003). Multivariate analysis of multiple response data. *Journal of Marketing Research*, 40, 321–334.
- Gilbride, T., & Allenby, G. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, 23, 391–406.
- Jedidi, K., Jagpal, S., & Manchanda, P. (2003). Measuring heterogeneous reservation prices for product bundles. *Marketing Science*, 22, 107–130.
- Jedidi, K., & Zhang, J. (2002). Augmenting conjoint analysis to estimate consumer reservation prices. *Management Science*, 48, 1350–1368.
- Liechty, J., Fong, D., & DeSarbo, W. (2005). Dynamic models incorporating individual heterogeneity: Utility evolution in conjoint analysis. *Marketing Science*, 24, 285–293.
- Meijer, E., & Rouwendal, J. (2006). Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics*, 21, 227–244.
- Newton, M. A., & Raftery, A. E. (1994). Approximate bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B*, 56, 43–48.
- Orme, B. (2001). *Assessing the monetary value of attribute levels with conjoint: Warnings and suggestions*. Sawtooth Solutions Customer Newsletter (Spring), Sequim, WA: Sawtooth Software, Inc.
- Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics*, 4, 647–657.
- Rossi, P., Allenby, G., & McCulloch, R. (2005). *Bayesian Statistics and Marketing*. England: Wiley.
- Shaffer, G., & Zhang, J. (1995). Competitive coupon targeting. *Marketing Science*, 14, 395–416.
- Shaffer, G., & Zhang, J. (2000). Pay to switch or pay not to switch: Third degree price discrimination in markets with switching costs. *Journal of Economics & Management Strategy*, 9, 397–424.
- Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, 64, 583–639.
- Swait, J., Erdem, T., Louviere, J., & Dubelaar, C. (1993). The equalization price: A measure of consumer-perceived brand equity. *International Journal of Research in Marketing*, 10, 23–45, (March).
- Train, K. (2003). *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.
- Zellner, A. (1978). Estimation of functions of population means and regression coefficients including structural coefficients: A minimum expected loss (MELO) approach. *Journal of Econometrics*, 8, 127–158.