Bayesian Statistics and Marketing

Greg M. Allenby
Fisher College of Business
Ohio State University

Institute of Statistical Mathematics
December, 2002

Marketing

- An applied field.
  - What should we offer? To whom? At what price? Through which channel? How should we inform consumers?
- Heterogeneity in preferences, sensitivities, and motivations.
- Complex processes and limited information.

Bayesian Statistics

MCMC methods can exploit the structure of hierarchical models to:

- Estimate high-dimensional models
- Estimate models for non-standard (e.g., lumpy) data.
- Efficiently pool information to address Marketing's information poor environment.
  - Priors
  - Other units of analysis

Statistical Analysis in Marketing

Within-unit behavior (the conditional likelihood).

Across-unit behavior (the distribution of heterogeneity).

Action (the solution to a decision problem involving a loss function).

Bayesian Methods offer a Unified Framework

Within-Unit Analysis

Latent variable models

$$z = Xb + e \qquad e \sim N(0,S)$$

Tobit Model: $y=0,\ z<0;\ y=z,\ z\geq 0$

Ordered Probit Model: $y=r,\ c_{r-1} < z \leq c_r$

MNP: $y=j,\ z_j=\max(z_1,\ldots,z_m)$

Multivariate Probit: $y_j=1,\ z_j>0$; else $y_j=0$

Multivariate Probit Model

$$\Pr(i) = \Pr(z_i > z_j \text{ for all } j)$$

$$= \Pr(x_i'\beta + \varepsilon_i > x_j'\beta + \varepsilon_j \text{ for all } j)$$

$$= \Pr(\varepsilon_j < x_i'\beta - x_j'\beta + \varepsilon_i \text{ for all } j)$$

$$= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{x_i' b - x_j' b + e_i} \cdots \int_{-\infty}^{x_i' b - x_m' b + e_i} f(e_j) \cdots f(e_m)\, de_j \cdots de_m \right] f(e_i)\, de_i$$

Data Augmentation

The introduction of an augmented latent variable (z) simplifies the evaluation of the model parameters (b) given the data (y).

Obtain $p(z, b|y)$ by drawing iteratively from $p(z|y, b)$ and $p(b|z, y)$.

Obtain the marginal posterior distribution of b by ignoring the draws of z: $p(b|y) = \int p(z, b|y)\, dz$

Implementation

Model:

$y|z$ ← *Choose the alternative with the greatest z*

$z|x,\beta$ ← *z is distributed according to a normal distribution*

Conditional Distributions:

$z|y,x,\beta$ ← *Draw from a truncated normal distribution*

$\beta|z,x$ ← *This is simple OLS*

Consumer Choice

- Consumers choose offerings with attributes that provide benefits.
- In most product categories there are multiple offerings, with many attributes, resulting in a complex decision process.
- Consumers often use screening rules to simplify the choice process.

Digital Cameras

A selection of tested products (Within types, in performance order) **Brand & model**	Memory	Zoom lens	Manual controls	AA batteries	Charger	AC adapter	Eyeglasses
3- TO 5-MEGAPIXEL CAMERAS							
Sony DSC-F707	MS	5x	■		■	■	■
Canon PowerShot G2	CF	3x	■		■	■	■
Olympus Camedia C-3040 Zoom	SM	3x	■	■			■
Olympus Camedia D-40 Zoom	SM	2.8x	■	■			
Fujifilm FinePix F601 Z	SM	3x	■		■	■	
Sony Cyber-shot DSC-S75	MS	3x	■		■	■	■
HP PhotoSmart 812	SD	3x		■			
Kodak EasyShare DX4900	CF	2x		■			
Olympus Camedia E-10	CF, SM	4x	■	■			■
Canon PowerShot S40	CF	3x	■		■		
Casio QV-4000	CF	3x	■	■	■		■
Kyocera Finecam S4	SD	3x	■		■	■	
Panasonic Lumix DMC-LC5	SD	3x	■		■	■	■
Sony Cyber-shot DSC-P71	MS	3x		■	■	■	
Sony CD Mavica MVC-CD400	CD-R/RW	3x	■		■	■	
Minolta Dimage S404	CF	4x	■	■			
Toshiba PDR-3310	SD	3x	■		■	■	
Kyocera Finecam S3	SD	2x	■		■	■	

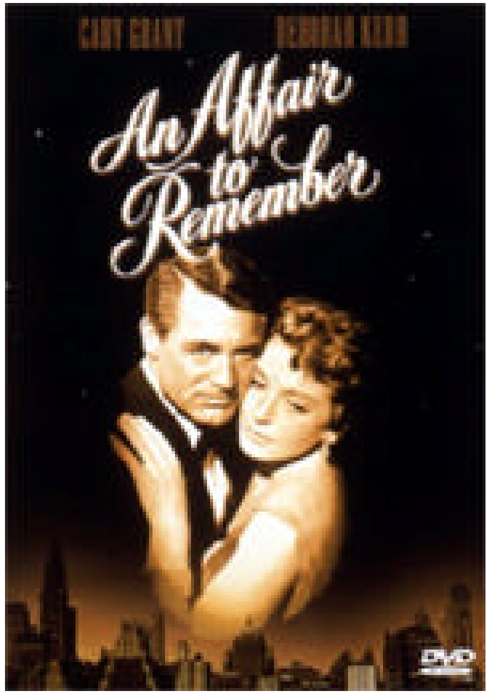
Alternative Screening Rules

- Conjunctive Rule
  - Alternative must be acceptable on each of a subset of attributes.
- Disjunctive Rule
  - Alternative must be acceptable on at least one attribute.
- Compensatory Rule
  - “Really good” attributes can make up for “really bad” ones.

Bayesian *Decision Theory* meets ***Behavioral*** *Decision Theory*

Will it be …..

The Thriller in Manila



An Affair to Remember

Consideration Set Choice Model

Let $I(x,\gamma)$ denote an arbitrary decision rule.

$I(x,\gamma) = 1$ if the rule is satisfied, $=0$ if not.

<u>Example</u>:

x=price, $\gamma = \$500$,

$I(x,\gamma) = 1$ if $x < \$500$

$I(x,\gamma) = 0$ if $x \geq \$500$

$$\boxed{\Pr(i) = \Pr(x_i'\beta + \varepsilon_i > x_j'\beta + \varepsilon_j \text{ for all } j \text{ such that } I(x_j,\gamma) = 1)}$$

Data Augmentation and Consideration Sets

<u>Model:</u>

$y|z, I(x,\gamma)$ ← *Choose the alternative with the greatest z, among those considered*

$z|x,\beta$ ← *z is distributed according to a normal distribution*

<u>Conditional Distributions:</u>

$z|y,x,\beta,I(x,\gamma)$ ← *Draw from a (truncated) normal distribution*

$\gamma|\ z, y, x$ ← *Draw from the set of allowable cut-offs*

$\beta|z,x$ ← *This is simple OLS*

Empirical Application

- Discrete choice conjoint study
  - Advanced Photo System (APS)
  - 302 participants, pre-qualified
  - Each participant evaluated 14 choice sets
  - Each set comprises 7 alternatives
    - 3 APS cameras
    - 3 35mm cameras
    - None option

Attributes and Levels

Attribute	Levels
Basic body style and standard features	Low
	Medium
	High
*Mid-Roll Change**	None
	Manual
	Automatic
*Annotation**	None
	Pre-Set
	Customized List
	Custom List 1
	Custom List 2
	Custom List 3
*Camera Operation Feedback**	No
	Yes

Attribute	Levels
Zoom	None
	2x
	4x
Viewfinder	Regular
	Large
Camera Settings Feedback	None
	LCD
	Viewfinder
	Both
Price	$41 - $499

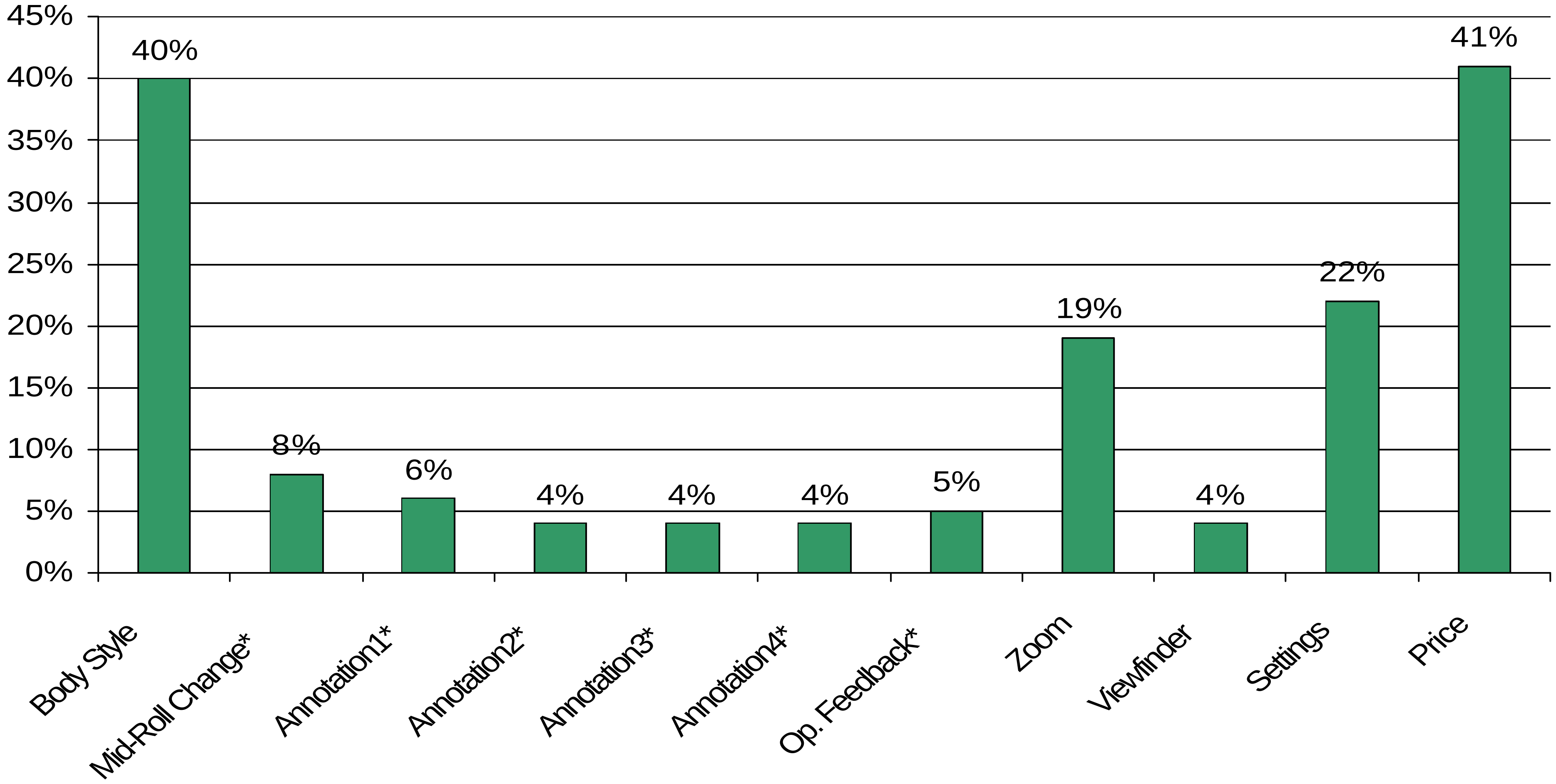
*Attributes unique to APS cameras at the time of the study

Model Fit

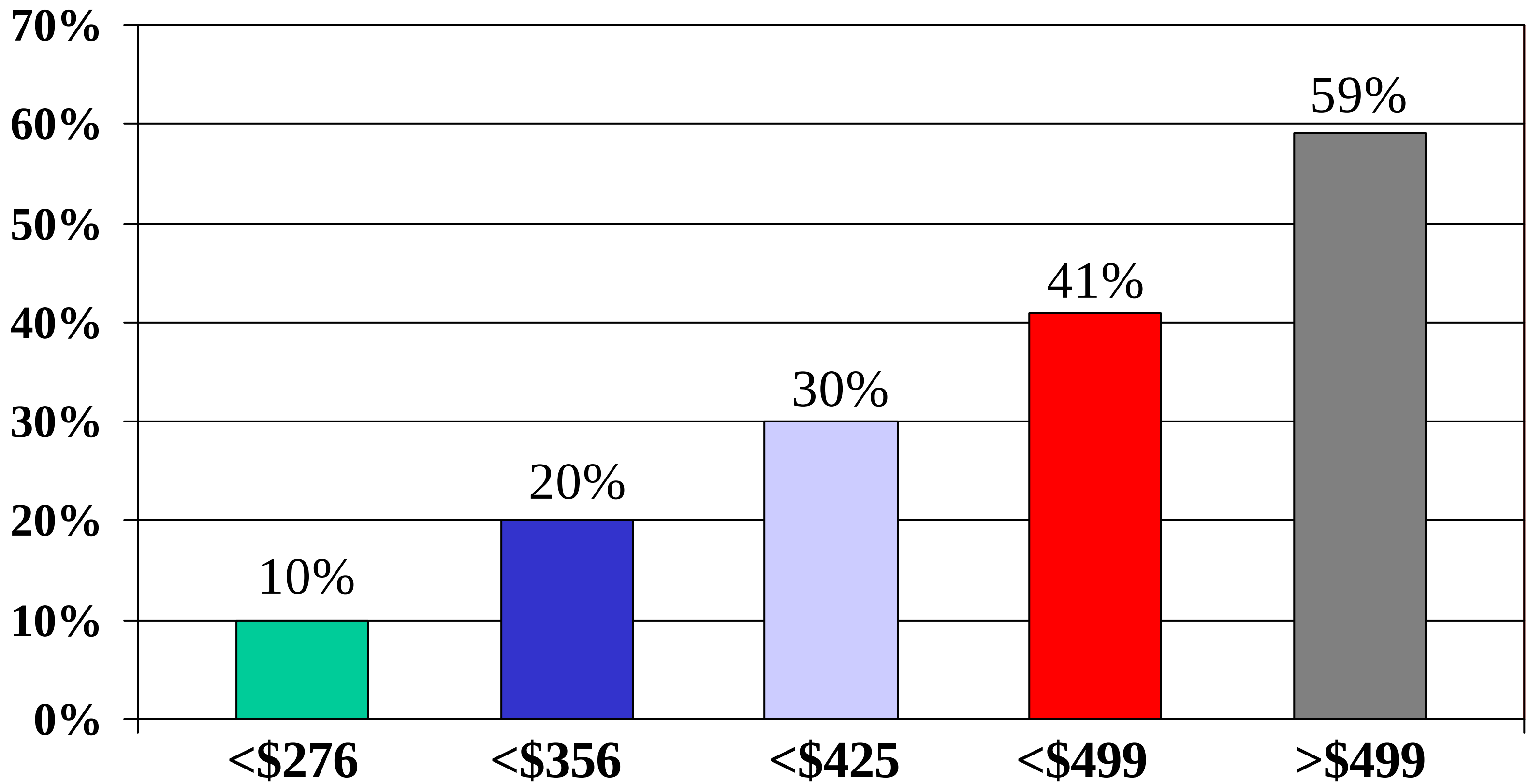
	In Sample		Predictive	
	LMD	Hit Probability	Hit Probability	Hit Frequency
Probit	-3468.8	0.525	0.391	266
Compensatory	-3449.6	0.527	0.393	267
Conjunctive	***-2990.8***	***0.583***	***0.418***	***276***
Disjunctive	-3586.2	0.528	0.39	267
Structural Hetero.	-3012.9	0.583	0.417	277

- ‘None’ option is assumed to be in every consideration set
- Heterogeneity is introduced for all parameters
- Estimation is by MCMC: 10,000 iterations

Proportion Screening on Each Attribute



Distribution of Price Threshold



Across Unit Analysis

The distribution of heterogeneity is typically specified as <u>**iid**</u> draws from a mixing distribution:

$$\beta_i \sim \text{Normal}(\mu, V_\beta)$$

"i" denotes the respondent/household/unit of analysis

Interdependent Preferences

- Influence of the taste of others
  - Minivans
  - Abercrombie and Fitch
- Extended product concept
  - Utility is derived from factors beyond the physical formulation of the offering.
- Mechanisms
  - Social concerns, network externalities, signaling effects, omitted variables (?), ...

The Standard Binary Choice Model

$$\Pr(y_i = 1) = \Pr(U_{i2} > U_{i1}) = \Pr(z_i > 0)$$

$$\left.\begin{array}{l} z_i = x_i^{'} b + e_i \\ \\ e_i \sim Normal(0,1) \end{array}\right\} \quad z \sim Normal(Xb, I)$$

Implication of iid Errors

- Conditional and unconditional expectations of preference are equal

$$E[z_2 \mid z_1] = X_2^{'}\boldsymbol{b} + \cancelto{0}{\Sigma_{21}}\Sigma_{11}^{-1}(z_1 - X_1\boldsymbol{b}) = X_2^{'}\boldsymbol{b}$$

Proposed Specification

Exogenous Covariates

$$z = Xb + e + q$$

$$q = rWq + u$$

Reflects interdependence among individuals

$$e \sim Normal(0, I)$$

$$u \sim Normal(0, s^2 I)$$

$$z \sim Normal(Xb, I + s^2 (I - rW)^{-1} (I - rW')^{-1})$$

Autoregressive Time Series Models

The AR(1) Process $y_t = \rho y_{t-1} + \varepsilon_t$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = r \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

or $y = \rho W y + \varepsilon$, where ρ measures the degree of association

Spatial Autoregressive Models

Observations that are connected circularly:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = r \begin{bmatrix} 0 & .5 & 0 & 0 & .5 \\ .5 & 0 & .5 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & .5 & 0 & .5 \\ .5 & 0 & 0 & .5 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

The rows of W should sum to 1.0

$\rho > 0$ implies conformity, $\rho < 0$ implies individualistic behavior

Generalized Spatial Autoregression

$$q = rWq + u$$ ← Augmented error term

$$W = \sum_{k=1}^{K} f_k W_k$$ ← W_k reflects potential dependence for k^{th} covariate (e.g., social class, physical proximity, etc.).

$$\sum_{k=1}^{K} f_k = 1$$ ← ϕ_k is the influence of the k^{th} covariate.

MCMC Estimation with Data Augmentation

$$p(z_i|\sim) = Truncated \quad Normal(x_i' b + q_i, 1)$$

Augmented parameter simplifies calculations

$$p(b|\sim) = N(v, \Omega)$$

$$v = \Omega^{-1}(X'(z - q) + D^{-1} b_0), \quad \Omega = (D^{-1} + X'X)^{-1}$$

$$p(q|\sim) = N(v, \Omega)$$

$$v = \Omega^{-1}(z - Xb), \qquad \Omega = (D^{-1} + s^{-2} B'B)^{-1}, \; B = I - rW$$

Empirical Study

- Mid-sized car purchases – Japanese versus non-Japanese.
- Two sources of dependence – geographic and social neighbors.
  - Geographic neighbors revealed by zip codes.
  - Social neighbors revealed by demographics.
    - Age, income, ethnic origin and education
- 857 consumers, 122 zip codes
  - 666 used for model calibration, 191 holdout

Sample Statistics

Variable	Mean	Standard Deviation
Car Choice (1=Japanese, 0= Non-Japanese)	0.856	0.351
Difference in Price ($000)	-2.422	2.998
Difference in Options ($000)	0.038	0.342
Age of Buyer (years)	48.762	13.856
Annual Income ($1000)	66.906	25.928
Ethnic Origin (1=Asian, 0=Non-Asian)	0.117	0.321
Education (1=College, 0=Below College)	0.349	0.477
Latitude (relative to 30)	3.968	0.484
Longitude (relative to –110)	-8.071	0.503

Model Fit

	In-Sample Fit (log marg.den)	Predictive Fit (MAD)
Probit	-237.7	0.177
Random-Effects	-203.8	0.158
Geographic Neighbors	-146.4	0.136
Demographic Neighbors	-151.2	0.139
Mixture	-133.8	0.127

Posterior Estimates of q: Not Linearly Related to Covariates

Average	q > 0	q £ 0	“p-value”
Age	48.3	48.7	0.96
Income	69.8	71.9	0.42
Ethnic Origin	0.12	0.94	0.28
Education	0.37	0.38	0.67
Latitude	4.15	4.16	0.50
Longitude	-8.28	-8.19	0.03

Across Unit Analysis

- Interdependent preferences
  - Pervasive
  - Not well represented by linear model structures
  - Not “iid”
- Extensions:
  - Opinion leaders (extreme realizations of θ)
  - Aspiration groups (asymmetric W)
  - Temporal evolution (e.g., buzz)
    - Longitudinal and cross-sectional data
  - Multinomial response (multiple θ)

Marketing Actions

Loss functions are often easy to identify.

Examples:

- Direct Marketing
  - Coupon with face value (F)
    - $\pi = \Pr(i \mid \beta, x, \text{price} - F)(\text{Margin} - F)$
- Identification of individuals for further study
  - Characterization of extremes

Alternative Information Sets

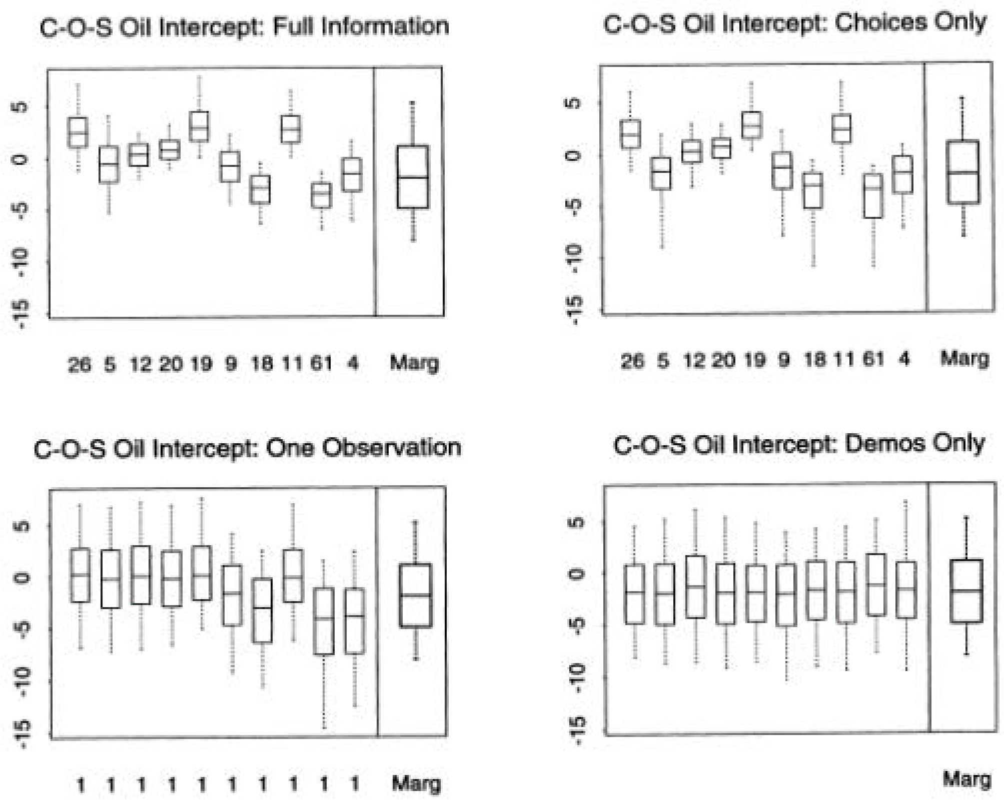
- Base
  - Information on the distribution of preferences, no specific info on individuals
- Demographic
  - Base info + demographic info on individuals
- Choices Only
  - Demo set + info on purchase history without causal environment (e.g., price)
- One Observation
  - Demo set + brand choice with causal info for 1 obs.
- Full
  - Complete information on purchase history and causal environment

Data

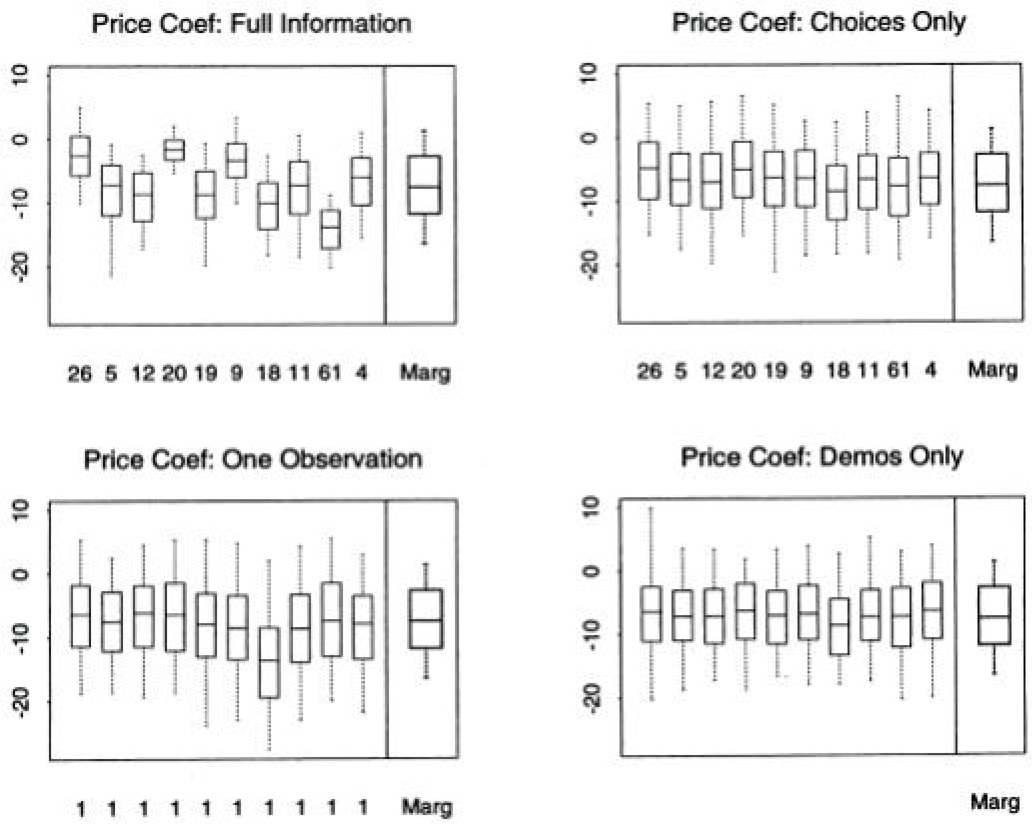
Scanner panel dataset of canned tuna purchases:

- Brands:
  - Chicken of the Sea (water and oil)
  - Starkist (water and oil)
  - House Brand
- Data:
  - Choices, prices, display and feature information.
- Individual-level posteriors of β based on different information sets.

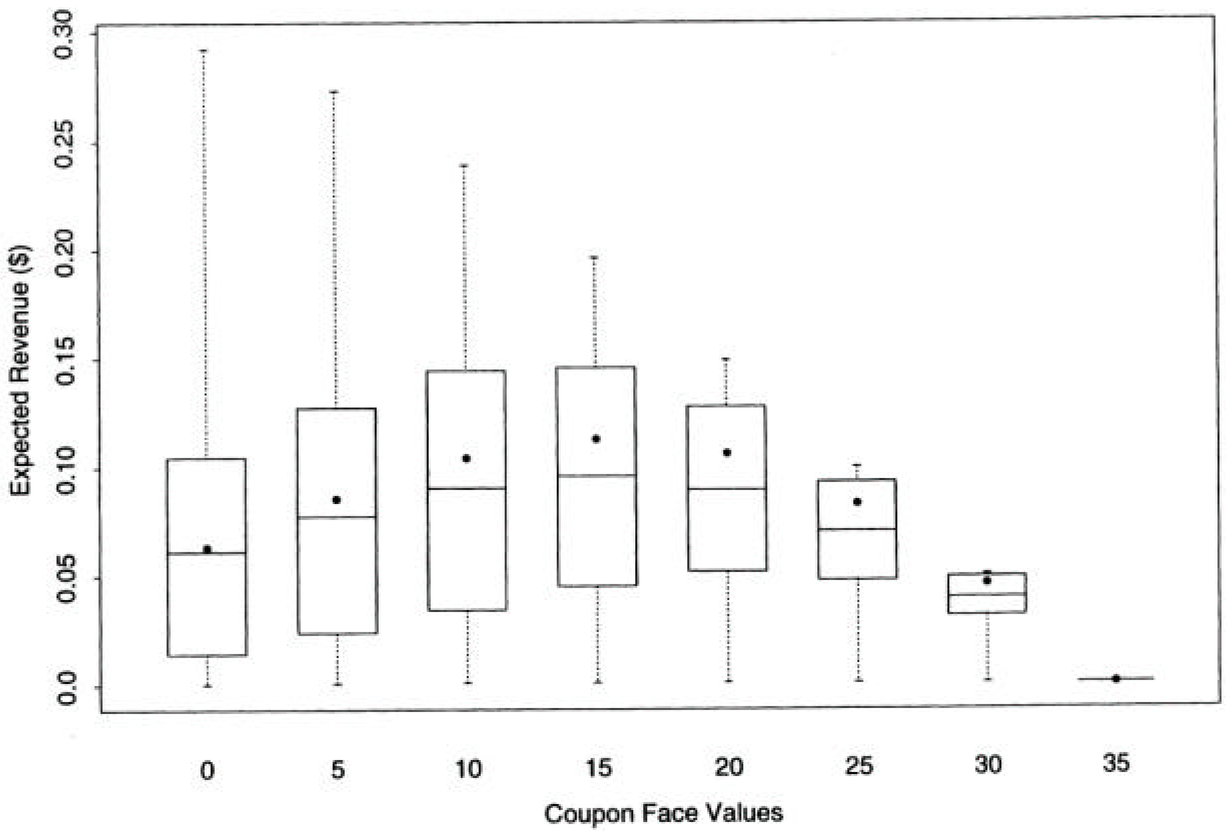
Predictive Distributions: Model Intercepts



Predictive Distributions: Price Coefficients



Optimal Face Values for One Customer



The Value of Information Sets

- Computed via the loss function by comparing:
  - Profits when face value is zero ($F=0$)
  - Maximum profits when face value not zero ($F \neq 0$)

Relative Value

Information Set	Net Revenue (per purchase)	Gain Relative to Blanket Drop
Full	0.1570	2.55
Choices-Only	0.1529	1.93
One Obs.	0.1500	1.56
Demos-Only	0.1467	1.12
Blanket	0.1459	1.0
No Coupon	0.1380	

Bayesian Statistics and Marketing

- Within-unit analysis
  - Choice and quantity, simultaneous purchase, state-space models.
- Across-unit analysis
  - The unit of analysis (person-activity), role of the objective environment, relationship to motivating conditions, market segmentation.
- Action
  - Choice simulators, product line design, exploration of extremes, analysis of economic systems.