

The Value of Purchase History

Rossi, McCulloch and Allenby (1996)
Models in Quantitative Marketing
Jason Bell · October 23, 2026

Who should get the coupon?

Household A bought canned tuna 20 times

Household B bought it once

What coupon face value should
each household receive?

Why this paper matters

Made hierarchical Bayes standard in marketing

Estimates tastes for every household, even with 1 trip

Puts a dollar value on customer purchase histories

Today's questions

Why do one-household estimates fail on short panels?

How does a hierarchical model borrow strength?

How much taste variation do demographics explain?

What is purchase history worth in targeted couponing?

Two ways to use purchase history

Week 8: Guadagni & Little

One shared β for everyone
History is a variable $L_{ij,t}$

Week 9: Rossi et al.

Each household gets its own β_h
History is data in Bayes' rule

From a hand-built loyalty index to household-level preference inference

The targeting setting

The canned tuna scanner panel

Brand (6-oz can)	Share	Avg. Price	Display	Feature
Chicken of the Sea (Water)	0.413	\$0.681	0.201	0.372
Starkist (Water)	0.294	\$0.758	0.126	0.256
House Brand (Water)	0.053	\$0.636	0.101	0.142
Chicken of the Sea (Oil)	0.134	\$0.694	0.164	0.307
Starkist (Oil)	0.104	\$0.751	0.195	0.232

400 Springfield, MO households · 5,292 purchases (1 to 61 per household, mean 13.23)

Source: Rossi, McCulloch & Allenby (1996), Table 2, p. 328 (A.C. Nielsen scanner panel)

Two easy options, both flawed

Each household alone

**Uses only that household's trips
Explodes when trips are few**

Everyone pooled together

**One set of tastes for all
Sends everyone the same
coupon**

A hierarchical model sits in between: partial pooling

Why separate estimates break down

Household	Purchases (T_h)	Separate Estimate of Price Sensitivity
Household A	20	−6.8 (usable, still noisy)
Household B	3	−18.4 (severe overfit)
Household C	1	No finite estimate (7 parameters, 1 trip)

Each household has 7 parameters (4 brand intercepts + log price, display, feature)

Illustrative diagnostic for a 7-parameter choice model when T_h ranges from 1 to 61

A hierarchical model of tastes

Stage 1: household choice (multinomial probit)

$$y_{ht} = X_{ht}\beta_h + \varepsilon_{ht}, \quad \varepsilon_{ht} \sim N(0, \Lambda)$$

y_{ht} : latent $J \times 1$ utilities on trip t ; household buys highest $y_{ht,j}$

X_{ht} : brand intercepts, log price, display, feature β_h : household h 's 7 tastes

$\Lambda = \text{diag}(1, \sigma_2^2, \dots, \sigma_J^2)$: independent normal errors with unequal variances

Chicken of the Sea (Water) is the reference brand (intercept = 0, $\sigma_1^2 = 1$)

Stage 2: how tastes vary across households

$$\beta_h = \Delta z_h + v_h, \quad v_h \sim N(0, V_\beta)$$

z_h : household demographics (intercept, income, family size, retired, unemployed, single mom)

Δ : how average tastes shift with observable demographics (7×6 matrix)

V_β : full 7×7 covariance matrix of unobservable taste differences

Off-diagonal elements of V_β let brand preferences covary with price sensitivity

Bayes' rule for one household

$$p(\beta_h \mid y_h, z_h, \Delta, V_\beta) \propto p(y_h \mid \beta_h, \Lambda) p(\beta_h \mid z_h, \Delta, V_\beta)$$

Left: our updated belief about household h 's tastes after seeing its history

First factor: likelihood of household h 's observed choices y_h

Second factor: population prior centered at demographic mean Δ z_h

Every household gets a full posterior distribution, not just a single point estimate

Intuition: partial pooling as shrinkage

$$\hat{\beta}_h \approx w_h \hat{\beta}_h^{\text{own}} + (1 - w_h) \Delta z_h, \quad w_h = \frac{T_h}{T_h + \kappa}$$

T_h : number of purchases observed for household h

κ : ratio of within-household noise to between-household taste variance V_β

Small T_h : lean on the population prior Large T_h : lean on the household's own data

Exact for normal linear models; the probit posterior shrinks the same way

Shrinkage in numbers

Purchases (T_h)	Weight on Own Data (w_h)	Posterior Mean Price Sensitivity
0 (New customer)	0.00	−7.3 (population mean)
1 (Checkout trigger)	0.20	−8.2
4 (Light buyer)	0.50	−9.7
20 (Heavy buyer)	0.83	−11.2 (near own data −12.0)

Population mean −7.3; household's own choices suggest −12.0

Illustrative normal-normal calculation with $\kappa = 4$, anchored at the paper's population mean (−7.3)

Borrowing strength

Households with little history
look like the population

Households with a long history
look like themselves

Shrinkage protects short panels from extreme overfitting

Computing the posterior

Why use a probit with Gibbs sampling?

Evaluating 400 seven-dimensional integrals
by numerical quadrature is intractable

Data augmentation fills in latent utilities
so every step is a standard draw

Albert & Chib (1993); McCulloch & Rossi (1994)

The Gibbs sampling loop

1. Draw latent utilities y_{ht} from truncated normals
2. Draw each household's β_h via Bayesian regression
3. Draw population (Δ, V_β) and Λ given all β_h
4. Repeat thousands of times; keep draws after burn-in

Conditional on the latent utilities y_{ht} , updating β_h is exact normal regression

How tastes are modeled across frameworks

Feature	Week 8: Logit + Loyalty	Finite Mixture (3-Class)	Week 9: Continuous HB
Taste structure	One shared β for all	3 discrete support points	Continuous $\beta_h \sim N(\Delta z_h, V_\beta)$
Past purchases	Covariate $L_{ij,t}$ in X	Updates class weights	Updates full 7D vector β_h
Posterior range	Single point	Inside 3-point simplex hull	Unbounded across R^7
Holdout log-lik	Baseline	−2,036 (Appendix B)	−1,170 (much better fit)

Appendix B: continuous HB beats a 3-class latent mixture (−1,170 vs. −2,036)

**What do demographics and history
reveal?**

Demographics explain very little of tastes

Parameter	Posterior Mean	Unobserved SD	Demographic ρ^2
Price (log price)	-7.30	3.31	0.073 (7.3%)
Display	0.26	0.64	0.132 (13.2%)
Feature ad	0.41	0.65	0.120 (12.0%)
House Brand intercept	-4.30	2.49	0.330 (33.0%)
Starkist (Water) intercept	-0.11	1.01	0.200 (20.0%)

Six household demographics explain only 7.3% of the variance in price sensitivity

Source: Rossi, McCulloch & Allenby (1996), Table 4, p. 329 ($\rho^2 = 1 - \text{Var}(v_h)/\text{Var}(\beta_h)$)

Five information sets the firm might hold

Information Set	Data Observed for Household h	Marketing Mechanism
Base (Blanket)	Population distribution only	Untargeted mass drop
Demographic	6 household demographics (z_h)	Mailing list / census targeting
One Observation	z_h + 1 trip & store environment	Catalina checkout coupon
Choices-Only	z_h + brand history (no prices)	Frequent shopper card
Full Information	z_h + full history + prices & promos	Complete scanner panel

How does Choices-Only infer price sensitivity without seeing prices? Through V_β !

How purchase history sharpens beliefs

Demographics barely move the posterior from the base prior

Even one purchase strongly updates brand and price beliefs

Choices-Only infers price sensitivity via correlated tastes V_{β}

Full history tightly pins down household price response

Figures 1–4 (pp. 331–335): behavioral data dominate demographic covariates

From posterior beliefs to coupon targeting

Optimal targeted couponing decision

$$E[\pi_h(F) \mid I_h] = \int \Pr(1 \mid \beta_h, \Lambda, p_1 - F, X) (M - F) p(\beta_h \mid I_h) d\beta_h$$

Manufacturer of Chicken of the Sea (Water) earns wholesale margin $M = \$0.35$

Chooses coupon face value $F \in \{0\text{¢}, 5\text{¢}, 10\text{¢}, \dots, 35\text{¢}\}$ to maximize expected net revenue

Computed by averaging profit over MCMC posterior draws of β_h given information I_h

Averaging over posterior draws accounts for uncertainty; plugging in $E[\beta_h]$ does not

The dollar value of purchase history

Information Set	Net Rev / Occasion	Gain vs. No Coupon	Multiple of Blanket
No Coupon ($F = 0\text{¢}$)	\$0.1380	\$0.0000	0.00×
Blanket (uniform optimal F)	\$0.1459	\$0.0079	1.00×
Demographics Only	\$0.1467	\$0.0087	1.12×
One Observation	\$0.1500	\$0.0120	1.56×
Choices-Only (no prices)	\$0.1529	\$0.0149	1.93×
Full Purchase History	\$0.1570	\$0.0190	2.55×

One purchase beats 6 demographics (1.56× vs. 1.12×); full history yields 2.55× blanket gain

Source: Rossi, McCulloch & Allenby (1996), Table 5, p. 337 (evaluated on holdout sample)

Why behavioral history beats demographics

Demographics add only 12% over blanket couponing (1.12×)

A single checkout trip adds 56% over blanket (1.56×)

Choices-Only nearly doubles the blanket gain (1.93×)

Full history tailors coupons from 0¢ to 25¢ (2.55×)

Figure 5 (p. 336): blanket drops cluster at 5–10¢; full history spreads from 0¢ to 25¢

Bridge to the modern paper

What is inside the product vector X ?

Rossi, McCulloch & Allenby
(1996)

5 tuna brands with tabular
attributes
(brand dummies, price, display,
ad)

Sisodia, Burnap & Kumar (2025)

High-dimensional product
visuals
(images mapped to
disentangled z)

Same hierarchical Bayes preference model; new learned representation of the product

Questions for Sisodia, Burnap & Kumar (2025)

Why do entangled image latents fail in HB conjoint?

How does the model learn disentangled visual factors z ?

How does HB estimate individual visual partworths β_i ?

How can the firm generate a new optimal product image?

Before we finish

Why does separate MLE fail when $T_h = 1$?

How does Choices-Only learn price sensitivity without prices?

Why average profit over draws of β_h instead of $E[\beta_h]$?

Why does 1 purchase beat 6 demographics for targeting?

Readings

Rossi, McCulloch & Allenby (1996), Marketing Science 15(4), 321–340
The value of purchase history data in target marketing

McCulloch & Rossi (1994), Journal of Econometrics 64, 207–240
An exact likelihood analysis of the multinomial probit model

Allenby & Rossi (1998), Journal of Econometrics 89, 57–78
Marketing models of consumer heterogeneity

Burnap, Hauser & Timoshenko (2023), Marketing Science 42(6), 1029–1050
Product aesthetic design: a machine learning augmentation

Sisodia, Burnap & Kumar (2025), Journal of Marketing Research 62(3), 406–429
Generative interpretable visual design: using disentanglement for visual conjoint analysis