Report Series

Correcting for CBC Model Bias: A Hybrid Scanner Data - Conjoint Model

Martin Natter
Markus Feurstein

Report No. 57
June 2001

Report Series

June 2001

SFB
'Adaptive Information Systems and Modelling in Economics and
Management Science'

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

Papers published in this report series
are preliminary versions of journal articles
and not for quotations.

This paper was accepted for publication in:
The International Review of Retail, Distribution and Consumer Research

This piece of research was supported by the Austrian Science Foundation
(FWF) under grant SFB#010 ('Adaptive Information Systems and Modelling
in Economics and Management Science').

Correcting for CBC Model Bias: A Hybrid Scanner Data - Conjoint Model

Abstract

Choice-Based Conjoint (CBC) models are often used for pricing decisions, especially when scanner data models cannot be applied. Up to date, it is unclear how Choice-Based Conjoint (CBC) models perform in terms of forecasting real-world shop data. In this contribution, we measure the performance of a Latent Class CBC model not by means of an experimental hold-out sample but via aggregate scanner data. We find that the CBC model does not accurately predict real-world market shares, thus leading to wrong pricing decisions. In order to improve its forecasting performance, we propose a correction scheme based on scanner data. Our empirical analysis shows that the hybrid method improves the performance measures considerably.

Keywords: Choice-Based Conjoint Analysis, External Validity, Latent Class Model, Pricing, Scanner Data Model

1 Introduction

Optimal pricing is among the most important tasks to be solved by marketing managers. The two major data sources for deriving quantitative models are scanner data and consumer-derived data. Choice-Based Conjoint (CBC) analysis (Louviere and Woodworth 1983) is one of the most frequently used methods for gathering and analyzing consumer data. Wittink and Cattin (1989) document 1062 commercial applications of conjoint analysis by 66 U.S. firms over the period 1981-1985. For Europe, Wittink, Vriens and Burhenne (1994) report 956 uses of conjoint analysis by 59 firms for the period 1986-1991 and identify pricing as the number one purpose of conjoint studies. Since scanner data are based on real purchase acts and conjoint data only reflect simulated choices, scanner data are typically preferred to CBC data. However, when the historical scanner data do not exhibit sufficient price variations, the estimation of response models may fail. According to AC Nielsen, only about 5% of consumer non-durables (accounting for about 20% of revenues) show significant price variations. Furthermore, price variations found in scanner data rarely reflect regular price variations but are due to price promotions. In contrast to scanner data models, conjoint models do not depend on historical data. However, CBC models struggle with several other deficiencies: In a CBC interview respondents make several hypothetical purchase decisions within a few minutes and there are no monetary consequences. Dynamic effects and other impacts which are not captured within a CBC study, such as increasing brand awareness, changes in the level of distribution, life-cycle effects, promotional activity, seasonal impacts, new market entrants, etc., may decrease the real-world validity of CBC models.

The aim of this work is twofold: On the one hand, we assess the predictive validity of a CBC study by comparing the CBC forecasts to the actual sales. On the other hand, we propose a hybrid method that can still be used when scanner data show insufficient price variation and CBC estimates are biased.

2 Design of the Validation Study

In our study, the validity of the Latent Class CBC model is tested for 8 different brands of mineral water. The CBC study was designed with brand (A1-A8) and price (B1-B9) attributes. The interviews with a total of 128 respondents were conducted via the Sawtooth interview software in the 25th week of 1997 at three different shopping malls. Each respondent was shown 14 choice sets with five concepts per set plus a none-choice or other option. The respondents were asked to pick one of the concepts. As an external validation of the Latent Class CBC model we use scanner data consisting of 95 weekly price and sales data (Jan. 1996 - Oct. 1997). To test our methodology of incorporating scanner data information, we split the 95 weeks of scanner data into 7 quarters and use the last 6 quarters for validation only.

2.1 The Latent Class CBC Model

DeSarbo, Ramaswamy and Cohen (1995) propose to use a latent class version of CBC to overcome the limitations of aggregate analyses or a priori segmentations. The authors generalize the Kamakura and Russell (1989) scanner data response methodology to a latent class CBC model considering within subject replications over choice sets. The respondents (segment specific) choice probability for segment s, P_s, is given by

$$P_s(j \in C_n) = \frac{exp\,(\beta_b(j,s) + p(j)\beta_p(s))}{\sum_{i \in C_n} exp\,(\beta_b(i,s) + p(i)\beta_p(s))} \tag{1}$$

where $\beta_b(j,s)$ is the intrinsic utility of brand $j = 1,\ldots,J$ to segment $s = 1,\ldots,S$ and $\beta_p(s)$ the price utility for segment s. The "none-option or other brand" has a price utility of zero. $p(j)$ reflects the price of brand j in choice set C_n. The share-of-preference (SoP) can then be written as

$$SoP(j) = \sum_{s=1}^{S} \alpha_s P_s(j) \tag{2}$$

where α_s represents the relative segment size of s. For the estimation of the CBC segment-specific parameters we use the proposed maximum likelihood procedure. For the determination of the number of segments, we split the data into an estimation (12 choice sets per respondent) and a validation set (2 choice sets per respondent). The model with the highest out-of-sample hit rate is chosen for external validation. The hit rate is defined as the percentage of correctly predicted choices. For the data at hand the three-segment solution with segment sizes (α) of 0.38, 0.23 and 0.39, respectively, shows the highest hit rate (47.6%). The parameter estimates of the three-class solution are indicated in Table 1.

2.2 Validity Measures

From the choice data, we build probabilistic choice simulators of the BTL-type and determine the shares-of-preference (SoP) of all available products based on the weekly scanner prices. The CBC estimates are then matched with the scanner data market shares (MS). We measure the validity of the CBC model by means of squared correlation between SoP and MS (r^2). High values of r^2 indicate good estimates of the price utility. Often management is not only interested in a high correlation between SoP and MS, but in good estimates of the real market shares. Therefore, we include the Variance-Accounted-For measure (VAF), given by $VAF = 1 - MSE(SoP, MS)/\sigma^2(MS)$. MSE(SoP,MS) denotes the mean squared error between CBC forecasts and the actual market shares of a brand over all periods and $\sigma^2(MS)$ the variance of the market shares. In order to measure the bias of the CBC model given by the difference between the average share-of-preference $\bar{SoP}$ and the average market share $\bar{MS}$, we define the following measure:

$$\Delta_{MS} = abs(\bar{MS} - \bar{SoP}) \tag{3}$$

High values of Δ_{MS} indicate a large difference between the CBC forecasts and real market share levels, i.e. a low validity of the brand utilities.

2.3 Results

Table 2 shows the brand-specific measures of the CBC forecasts. The columns illustrate the brand index, VAF, r^2, Δ_{MS} and the average market share for each brand. The average market-share-weighted correlation, r^2, between MS and SoP is 60.5% which indicates that price effects are nicely captured by the Latent Class CBC model. This is reflected by Figure 1 which displays the time series of the real MS (bold line) and the SoP (thin line) for each of the 8 brands. Although the SoP are highly correlated with MS, we identify two major problems:

- Interview bias: the levels of SoP and MS differ considerably for most of the brands, i.e., the brand utilities are biased.

- Stationarity: market shares may show a drift (brand 2) or a shock (brand 7). This may be caused by (sudden) changes of the degree of distribution (brand 7) or other external effects.

The other two measures, VAF and Δ_{MS}, reflect these two problems: VAF is less than zero for all brands except brand 4. Market share weighted VAF is as low as -21.6. The difference between SoP and MS, Δ_{MS}, is 3.7% on (MS weighted) average. For brands 5 and 6 the difference is even higher than MS itself. As the optimal pricing strategy is not only dependent on the price effect but also on the market share level, the marketing consequences of these problems are drastic. Therefore, the use of CBC share-of-preference estimates should, in general, not be taken as forecasts of the market shares without adjusting the external effects. An approach for coping with this difficulty is the calculation of a correction vector. In the following section, we demonstrate how scanner data sources can be used to account for such external effects without changing the price utilities of the CBC model.

3 A Hybrid Scanner Data - Conjoint Model

Other effects than price, brand or package may lead to significant changes of market shares, too. Building up channels of distribution may cause shifts in market shares (Golanty 1995). The life-cycle theory suggests that products follow a certain pattern of ups-and-downs during their time on the market. However, static conjoint models do not account for such effects. Due to the cross-sectional time-series nature, POS scanner data are an interesting source for improving choice modeling (Winer et al. 1994). If real market shares are known from the AC Nielsen Retail Index (NRI) or other data sources, a correction vector can be calculated as the fraction between average market shares from the NRI and the shares-of-preference forecasted by the CBC model at average observed prices. Due to the potentially evolving nature of market shares, we propose to use an adaptive correction scheme which calculates the correction factors based on the last quarter's average market shares. The advantage of this method is that the use of a factor has no effect on price utility and only corrects brand utility which we identified as the major source of error in our primary analysis. Given that many companies receive quarterly reports about the previous quarter's market shares, this is a very simple and practically feasible solution to the problem. The correction procedure works as follows:

1. Take average prices and market shares of the last quarter from NRI.

2. Compute the shares-of-preference as a function of the NRI prices.

3. Calculate the fraction of market shares and shares-of-preference.

4. Multiply this correction vector with the shares-of-preference of the forecasting period (weeks of the next quarter).

5. Normalize the forecasts to one (market share condition).

The quarter-specific correction factors are presented in Table 3. Factors greater (smaller) than 1 increase (decrease) the shares-of-preference. A factor of 1 does not change brand utility. There seems to be a tendency of the CBC model to overestimate (underestimate) the market shares of smaller (larger) brands. Very small factors in Table 3 may indicate that there are real business problems (availability, placement, etc.) and thus suggest to management to investigate the issue in more detail. The time dependence of the factors adds to the necessity of quarterly updates. The price utilities, which are not changed by this procedure, seem to be valid for longer periods of time (see Figure 2) than the brand utilities. Consequently, when brand utilities are updated by the proposed methodology it appears possible to use the model as a management tool over a longer time horizon. Table 4 shows that the use of a correction vector significantly improves the results for all measures considered. Market-share-weighted r^2 increases from 60.5% to 67.3%. Market-share-weighted VAF reaches 59.7% and is positive for each individual brand. The correction vector corrects shares-of-preference to the correct market-share level of the past quarter. Hence, Δ_{MS} decreases to 1.1%.

4 Conclusion

In our study, we examine the performance of an interview-based Latent Class CBC model in terms of scanner data. Our error measures indicate that the CBC model provides good estimates of the price utilities. However, the brand utilities are seriously biased so that the shares-of-preference differ considerably from market shares. As the brand utility influences pricing decisions, CBC models should not be directly used for pricing decisions. In some cases, scanner data models could be employed. However, when the historical scanner data do not exhibit enough price variations, the estimation of response models may fail. We propose a hybrid method that can still be used when scanner data show insufficient price variation and CBC estimates are biased. Our approach uses a correction vector which dynamically updates the brand utilities. This vector is based on the relation between the last quarter's average share-of-preference and average market shares. The results of our empirical study show that all performance measures considered can be improved significantly. Since quarterly market share information is available to most companies, our hybrid method is easily applicable in practice.

5 References

DeSarbo, W. S., Ramaswamy, V. and Cohen, S. (1995) 'Market Segmentation with Choice-Based Conjoint Analysis', Marketing Letters 6(2): 137-147.

Golanty, J. (1995) 'Using Discrete Choice Modeling to Estimate Market Share', Journal of Marketing Research 7: 25-28.

Kamakura, W. A. and Russell, G.J. (1989) 'A Probabilistic Choice Model for Market Segmentation and Elasticity Structure', Journal of Marketing Research 26: 379-390.

Louviere, J. J. and Woodsworth, G. G. (1983) 'Design and Analysis of Simulated Choice or Allocation Experiments: An Approach Based on Aggregate Data', Journal of Marketing Research 20: 350-367.

Winer, R. S., Bucklin, R. E., Deighton, J., Erdem, T., Fader, P. S., Inman, J. J., Katahira, H., Lemon, K. and Mitchell, A. (1994) 'When Worlds Collide: The Implication of Panel Data-Based Choice Models for Consumer Behavior', Marketing Letters 5(4): 383-394.

Wittink, D. R. and Cattin, P. (1989) 'Commercial Use of Conjoint Analysis: An Update', Journal of Marketing 53: 91-96.

Wittink, D. R., Vriens, M. and Burhenne, W. (1994) 'Commercial use of conjoint analysis in Europe: Results and critical reflections', International Journal of Research in Marketing 11: 41-52.

Table 1: Parameter estimates for the Latent Class CBC model

parameter	s1	s2	s3
β_b(1)	1.143	0.132	0.013
β_b(2)	1.901	0.019	0.087
β_b(3)	-1.564	-0.090	-0.380
β_b(4)	-1.212	-0.409	0.009
β_b(5)	0.352	-1.043	-0.853
β_b(6)	-0.042	-1.084	-0.477
β_b(7)	-4.313	0.243	-1.165
β_b(8)	-1.203	-0.564	-0.948
β_b(none)	-1.354	-2.243	-5.000
β_p	-0.240	-0.105	-0.959

Table 2: Standard model: Variance Accounted For (VAF) and r^2 for all brands, distance between average shares-of-preference and real market share levels (Δ_{MS}) and average market shares ($\bar{M}S$)

brand	VAF	r^2	Δ_{MS}	$\bar{M}S$
1	-3.26	0.825	0.040	0.208
2	-0.79	0.349	0.033	0.275
3	-5.56	0.807	0.067	0.175
4	0.70	0.752	0.015	0.148
5	-928.62	0.282	0.059	0.016
6	-158.26	0.576	0.034	0.026
7	-0.09	0.254	0.018	0.079
8	-3.98	0.621	0.024	0.071

Table 3: Correction factors for each brand and quarter Q_i

brand	Q1	Q2	Q3	Q4	Q5	Q6
1	1.43	1.31	1.25	1.21	1.24	1.26
2	1.17	1.16	1.17	1.17	1.18	1.13
3	1.72	1.65	1.76	1.58	1.60	1.55
4	0.86	0.89	0.90	1.00	0.90	0.92
5	0.26	0.24	0.24	0.21	0.21	0.21
6	0.41	0.44	0.44	0.40	0.41	0.44
7	0.60	0.61	0.64	0.84	0.87	0.97
8	0.80	0.78	0.75	0.70	0.71	0.76

Table 4: External effect model: Variance Accounted For (VAF) and r^2 for all brands, distance between average shares-of-preference and real market share levels (Δ_{MS}) and average market shares ($\bar{MS}$)

brand	VAF	r^2	Δ_{MS}	$\bar{MS}$
1	0.737	0.821	0.008	0.208
2	0.366	0.489	0.016	0.275
3	0.743	0.743	0.010	0.175
4	0.696	0.730	0.014	0.148
5	0.410	0.486	0.001	0.016
6	0.461	0.490	0.001	0.026
7	0.604	0.806	0.009	0.079
8	0.590	0.620	0.005	0.071

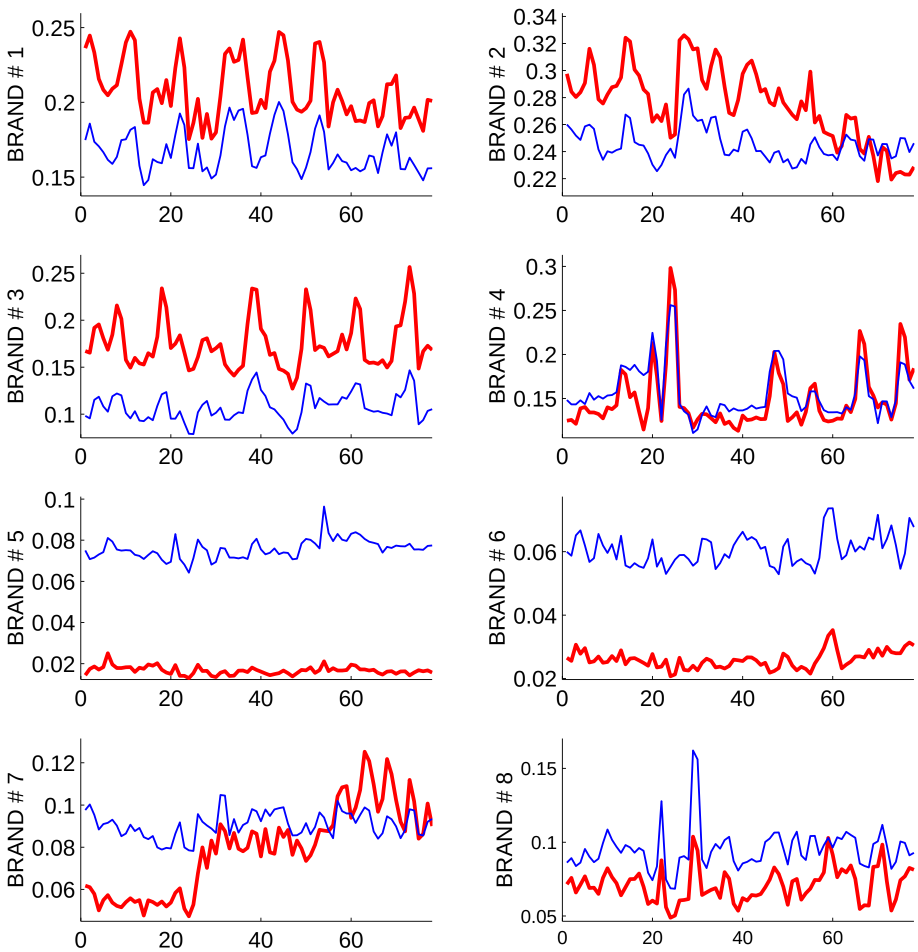

Figure 1: Real market shares (bold line), shares-of-preference (thin line) of the Latent Class CBC model

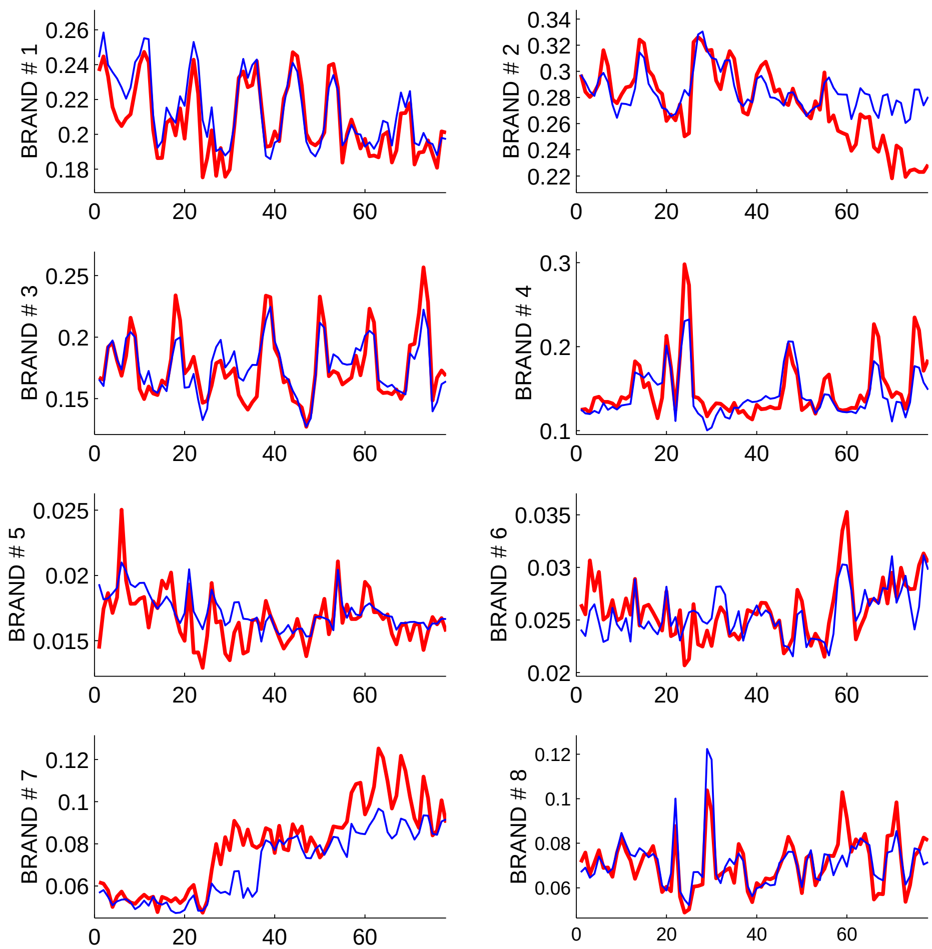

Figure 2: Real market shares (bold line), quarterly updated shares-of-preference (thin line) of the Latent Class CBC model