

Paper 262-27

Analyzing Discrete Choice Data on Monadic Cards

Jay Wu & Barry Zacharias, ACNielsen Vantis, San Ramon, CA

ABSTRACT

Although Discrete Choice Modeling typically involves multiple choice cards, monadic cards may add insights into [1] assessing market demands under current or most likely scenarios; [2] calibrating forecasts based on regular choice cards; [3] providing a quick store shelf simulation in addition to regular choice cards. We present a real life example to show how utilities can be estimated from a single monadic card and analyzed with the same rigor as the regular choice cards.

INTRODUCTION

Studies (e.g., Johnson & Orme, 1996) have shown that beyond a certain number of cards (up to 20), both response quality and dollar value of research deteriorate. Further, in analyzing data from discrete choices we commonly ignore the fact that repeated measures are obtained from multiple choice cards. However, even though we can advise respondents to focus on each individual card without referring back and forth to any other cards, it is unlikely that responses by the same respondent are totally independent. This puts a question mark on how reliable our choice data really are, despite the well documented superiority of tradeoff data in comparison to ordinary survey data (e.g., Louviere et al., 1990; Louviere & Woodworth, 1983)

Years of marketing research experiences in general, and new product forecasting in particular, suggest to us the importance of combining the best of two worlds – tradeoff exercises and ordinary survey instruments – in deriving in-depth market insights. Unlike the regular choice cards, we may choose to use monadic cards which show each respondent just

a single card – thus the name “monadic” – at the early stage of surveys, typically preceding other regular DCM cards.

Monadic cards occupy a unique place in the trade-off exercises for three reasons. *First*, a monadic card most closely resembles the *current* or most likely store shelves for the category of products in question, ranging from toothbrushes to high definition TVs. Take prices as an example, those listed on monadic cards reflect what most consumers would encounter in marketplaces either today or one, two, or five years down the road, while the regular choice cards may convey prices that are randomly determined. *Second*, a monadic card may be either more *complete* or more *focused* than regular choice cards in its list of models or brands. Models/brands expected to go away in the future will not be listed on regular choice cards for forecasting purposes, but may still be included on monadic cards for sensitivity checks. On the other hand, regular choice cards may sometimes show more models or options, anticipating future line extensions by competitors, while monadic cards may be designed to focus on today’s major competing brands accounting for more than 85% of the shares. *Finally*, because of their positions in the flow of survey, monadic cards typically command more respondents’ attention and thus higher response quality can be expected.

These features render monadic cards useful for improving forecasts on share of preferences. Monadic cards can be theoretically used as

- ❖ “shelf exposure” cards, where monadic cards deliberately simulate market scenarios today or tomorrow in order to get a direct read on specific demands under those conditions;

- ❖ “holdouts” cards, where monadic cards may be used to check on the predictive validity of the models and utilities derived from regular choice cards;
- ❖ “sample correction” cards, where shares of selection from monadic cards are compared with actual market shares to obtain a sense of how closely our sample is representative for the consumer population in the market;
- ❖ “add-on” cards to a *hybrid design*, where monadic cards depict a choice environment just like a regular choice card, so we can assess sources of volume, while the focus of the project is a conjoint exercise for an optimal product design; or
- ❖ “replacement” cards, where they are used to replace multiple regular choice cards when researchers are hard pressed with time and budget constraints. This last use is relatively rare.

Some monadic cards are deliberately designed to mimic current market. Thus one would expect frequencies of monadic picks align with market shares. Similarly, when monadic cards are configured to reflect the “most likely” scenarios not currently but in a perceivable future, one can obtain a direct read of how future market landscape will look like from these scenarios, subject to overstatement corrections and other biases associated with survey data.

Of course, monadic cards are not without limits of their own. Because only one scenario is presented, one risks over-generalization of a chance response or random picks out of low involvement, low motivation, ill-preparation, or even biased towards respondent’s last shopping experiences. Therefore, the best practice in using monadic cards to forecasts should include two features. First, concept demonstrations of new products should be presented prior to monadic cards to better prepare respondents to the tradeoff task (i.e., information acceleration, cf. Urban & Hauser, 1993.) Second, analytically it makes sense to estimate utilities from monadic data just like we do with regular multiple choice cards. By subjecting monadic choice data to the scrutiny of statistic models,

one obtains utilities – which are similar to the predicted values in a regression model – of each attribute and each model/brand. Because utilities are the workhorse of all DCM analyses, you are putting monadic data on a par with the rest of choice data, rendering more reliable benchmarks information for forecasting.

ANALYZING MONADIC CHOICE DATA: WHAT ARE AT STAKE

Analyzing monadic choice data requires some customization at work. Unlike the regular choice cards, which are derived from an experimental designs, monadic cards are typically not part of random designs. The reason: the values of attributes are generated not randomly but *deliberately* to simulate either current or most likely store shelves in the future. It involves negotiation with clients on whichever prices are best representing current markets, for example. There are well developed SAS® procedures to analyze regular choice data, with PROC FACTEX and PROC OPTEX for deriving the optimal designs and PROC TRANSREG and PROC PHREG for analyzing the data. We do not however have the luxury of automated procedures with monadic choice data. Instead we have to do several steps manually. This article discusses sample codes for accomplishing such a task.

The major steps involved are to first translate monadic cards into an experimental array of variables/attributes, read in the monadic responses, and then merge the design variables with responses or picks to a dataset ready for PROC TRANSREG and PROC PHREG to analyze.

The pivotal part of this process is to set up a *post hoc* experiment design file as if the monadic cards were generated from such a design – even though in reality the cards were drawn by hands on a piece of paper and then handed over to graphics department for polish before mass produced to the field execution.

A REAL LIFE EXAMPLE OF ANALYZING MONADIC CHOICE DATA

The Card and The Design. The Appendix shows a monadic card for jumbo TV sets. The card itself is "jumbo" in size and in number of possible choices. Each cell in the grid denotes a unique choice for respondents to pick, and there are 12 rows and 17 columns, making a total of $12 \times 17 = 204$ possible cells, of which 46 were marked "N/A" due to unavailability of certain brands at certain sizes (e.g., Brand A for all PLATFORM As; all brands for wide screen format with standard signal type,) thus reducing the number of non-missing cells to 158.

Eyeballing of the card points to five experimental factors: technological platforms (Platforms A, B, and C,) screen formats (Formats X and Y,) signal types (Signals 1 and 2,) six brands (with faceless names from Brand A to Brand F for client confidentiality,) and screen sizes (ranging from S1 to S14.)

The following codes are written with the above monadic card in mind:

```
proc plan ordered;
  factors platform=3 scrnform=2 signal=2 brands=6
    sizes=14;
  output out=cand1;
run;
```

PROC PLAN is good for generating full factorial designs, which amount to $3 \times 2 \times 2 \times 6 \times 14 = 1,008$ unique combinations, which are cut down to 158 by the following data step:

```
data cand2;
  set cand1;

  if (platform=1 and scrnform=1 and sizes in(2,4) and
    brands ne 1) or
    (platform=1 and scrnform=2 and sizes in(1,3,6) and
    brands ne 1) or
    (platform=2 and scrnform=1 and sizes in(9,11,13)) or
    (platform=2 and scrnform=2 and signal=2 and sizes
    in(5,7,10,14)) or
    (platform=3 and scrnform=2 and signal=1 and sizes
    in(2,5,8)) or
    (platform=3 and scrnform=2 and signal=2 and sizes
    in(2,5,8,10,12));
run;
```

the IF conditions selects the combinations consistent with the monadic card. CAND2 dataset now has 158 cases.

Frequencies of Picks From Monadic Card.

With the design file in place, we now switch to the responses to monadic card (i.e., monadic picks) that were stored in a previously created SAS file, DCM_RESPONSE.

```
data monadic;
  set dcm_response

  *---code screen size variable---;
  if a_x_s1 ne . then sizes=1;
  if a_y_s2 ne . or c_x_s2 ne . then sizes=2;
  if a_x_s3 ne . then sizes=3;
  if a_y_s4 ne . then sizes=4;
  if b_x_s5 ne . or c_x_s5 ne . then sizes=5;
  if b_x_s6 ne . then sizes=6;
  if b_x_s7 ne . then sizes=7;
  if c_x_s8 ne . then sizes=8;
  if b_y_s9 ne . then sizes=9;
  if b_x_s10 ne . or c_x_s10 ne . then sizes=10;
  if b_y_s11 ne . then sizes=11;
  if c_x_s12 ne . then sizes=12;
  if b_y_s13 ne . then sizes=13;
  if b_x_s14 ne . then sizes=14;

  *---array for screen formats---;
  array scrnf(*) a_y_s2 a_y_s4 b_y_s9 b_y_s11 b_y_s13
    a_x_s1 a_x_s3 a_x_s6 b_x_s5 b_x_s7 b_x_s10
    b_x_s14 c_x_s2 c_x_s12 c_x_s8 c_x_s10 c_x_s12;

  *---code screen format variable---;
  do i=1 to 5;
    if scrnf(i) ne . then scrnform=1; /*y format*/
  end;
  do i=6 to 17;
    if scrnf(i) ne . then scrnform=2; /*x format*/
  end;

  *---array for brands, tech platforms, and signal types---;
  array rinits(*) a_y_s2 a_y_s4 a_x_s1 a_x_s3 a_x_s5
    b_y_s9 b_y_s11 b_y_s13 b_x_s5 b_x_s7 b_x_s10
    b_x_s14 c_x_s2 c_x_s5 c_x_s8 c_x_s10 c_x_s12;

  *---code the tech platform variable---;
  do i=1 to 5;
    if rinits(i) ne . then platform=1; /*tech platform*/
  end;
  do i=6 to 12;
    if rinits(i) ne . then platform=2; /* tech platform*/
  end;
  do i=13 to 17;
    if rinits(i) ne . then platform=3; /* tech platform*/
```

```

end;

*---code signal type variable---;
do i=1 to dim(rinits);
  if rinits(i) in(7,8,9,0,11,12) then signal=2;
  else if rinits(i) in(1,2,3,4,5,6) then signal=1;
end;

*---code brands variable---;
do i=1 to dim(rinits);
  if rinits(i) in(1,7) then brands=1;
  else if rinits(i) in(2,8) then brands=2;
  else if rinits(i) in(3,9) then brands=3;
  else if rinits(i) in(4,0) then brands=4;
  else if rinits(i) in(5,11) then brands=5;
  else if rinits(i) in(6,12) then brands=6;
end;

drop I;
run;

```

Our goal is to align monadic response (i.e., picks) data with the design file. That is, dataset MONADIC created above contains all variables (PLATFORM SIZES SCRNFORM SIGNAL BRANDS) same as the design file except prices.

The following PROC FREQ moves one step closer to the goal:

```

proc freq data=monadic noprint;
  tables platform*scrnform*signal*brands*sizes
  /out=cc(drop=percent);
run;

```

The output dataset CC contains some combinations falling into the “N/A” cells on the monadic card. So we run another data step to delete those combinations in the new dataset NCC.

```

data ncc;
  set cc;
  if platform=2 and scrnform=2 and signal=1 then delete;
  if platform=. and scrnform=. and signal=. and brands=.
  and sizes=.
  then delete;
  size=input(put(sizes,sizef.),6.0);
  drop sizes;
run;

```

The first 5 observations of NCC are listed below.

```

brands    platform    scrnform  signal  size  COUNT

```

```

Brand B PLATFORM A  y format  Signal 1  S1  8
Brand B PLATFORM A  y format  Signal 1  S4  5
Brand C PLATFORM A  y format  Signal 1  S2  4
Brand C PLATFORM A  y format  Signal 1  S4  1
Brand D PLATFORM A  y format  Signal 1  S2  12
...

```

Merging Design with Monadic Picks. Next we merge the frequencies of picks with the design:

```

data ready;
  merge monadic ncc;
  by platform scrnform signal brands size;

  freq=count;
  c=1;
  output;

  freq=752-freq;
  if prices=0 and freq=791 then freq=0;
  c=2;
  output;
  drop count;
run;

```

PROC PHREG which we use to analyze monadic choices requires a censoring factor C with a value of 1 for pickings and 2 for non pickings (PHREG was designed for survival analysis, in which censored variables play a crucial role.) Thus for the first non-missing cell, the eight respondents picked this specific cell have a value of "1" on variable C, while 752-8=744 cases have a value of "2" for not picking (the total sample size for this study was 752.) This step essentially doubles the number of observations in dataset READY

Creating Main and Cross Effects for the Model. Dataset READY is in a shape ready for PROC TRANSREG, which creates several new variables to facilitate potential interaction or cross effects between PLATFORM, SCRNFORM and prices.

```

proc transreg data=ready design=2000 noz nor;
  model class(brands platform scrnform signal
  /zero=none)
  identity(size size_sq prices)
  class(platform scrnform
  /zero=none)*identity(prices)/lprefix=0;

```

```

output out=coded(drop=_type_ _name_
  intercept);
id c freq;
label prices='Prices' platform='Platform'
  scrnform='Screen format' signal='Signal Type'
  size='Screen size' size_sq='Size Squared';
run;

```

Modeling Monadic Picks. The total number of variables in the output dataset CODED from PROC TRANSREG is now 27 instead of 10 as in READY, due to variable expansion from the CLASS statement of PROC TRANSREG. Most of the 27 variables have a binary coding with values of 1 and 0, suiting the need for PROC PHREG.

```

proc phreg data=coded outest=betas nosummary;
  model c*c(2)= &_trgind /ties=breslow;
  freq freq;
  title "DCM for Monadic Data";
run;

```

PROC TRANSREG saves much time on coding cross effects. Once you have run it, your code for PROC PHREG is much shorter. Thus instead of citing all the effects on the right hand of MODEL statement, you only need to put "&_trgind" there, which contains all variables, all main effects and interactions. Further, if some cross effects turn out not significant, you can use the following trick: add a line "%put &_trgind;" immediately after PROC TRANSREG, which put all the model effects to the LOG window. You can simply cut and paste those effects to be part of PROC PHREG code, replacing MODEL C*C(2)=&_TRGIND with a long list of effects, and then cut off specifically those non-significant ones.

After submitting the above code, we obtain the following output:

Multinomial Logit Parameter Estimates

Parameters	DF	Estimate	SE	Chi-Square	Pr
Brand A	1	.75	0.15	25.17	<.0001
Brand B	1	.12	0.14	0.79	0.373
Brand C	1	-0.49	0.16	8.69	0.003
Brand D	1	-0.01	0.14	0.01	0.92
Brand E	1	1.03	0.13	67.29	<.000

Brand F	0	0	.	.	.
Plat A	1	-0.71	0.44	2.63	0.11
Plat B	1	-3.15	0.45	49.60	<.000
Plat C	0	0	.	.	.
Y Format	1	0.71	0.35	4.14	0.042
X Format	0	0	.	.	.
Signal 1	1	-1.23	0.13	85.95	<.0001
Signal 2	0	0	.	.	.
Screen size	1	0.40	0.057	48.07	<.000
Size Sq	1	-0.00	0.000	40.70	<.000
Prices	1	-0.00	0.00	43.18	<.000
Plat A*					
Prices	1	0.000	0.000	.2661	0.61
Plat B*					
Prices	1	0.00	0.000	14.03	0.00
Plat C					
*Prices	0	0	.	.	.
y format*					
Prices	1	-0.00	0.00	25.98	<.000
X format*					
Prices	0	0	.	.	.

We see that the cross effect between PLATFORM A and prices was not significant, but PLATFORM B and prices is. At this point we can choose to delete that specific effect and fit the rest of model.

Assessing Goodness of Fit. Having just one card makes it easier to compare the expected picks with the observed ones to assess goodness of model fit. The following steps are taken to achieve that goal.

First we obtain the product of model parameters and the design values. In other words, we compute the log of utilities (i.e., expected values) through PROC SCORE, and then obtain anti-log of utilities through another data step.

```

proc score data=coded(drop=c) score=betas type=parms
  out=p(keep=platform scrnform signal size size_sq
  prices brands freq c rename=(c=xb));
  var &_trgind;
run;

```

```

data p2;
  set p;
  /*cut off duplicated utilities due to C*/
  if mod(_n_,2) ne 0;
  exb= exp(xb);
run;

```

Using utilities we next compute shares of preference for each cell on the monadic card. First we obtain the total utilities to be used as the denominator for computing expected picks.

```
proc means data=p2 noprint;
  output out=s sum(exb) = t_exb;
run;

data p3(drop=_type__freq_ t_exb size_sq xb e_sop exb);
  if _n_=1 then set s;
  set p2;
  e_sop=exb/t_exb; /*expected share of preference*/
  exp_freq=e_sop*752; /*expected picks*/
run;
```

With expected and observed picks, we run a simple linear regression model with observed picks as dependent and expected picks as independent variables.

```
proc reg data=p3;
  model freq=exp_freq;
run;
```

The result shows that DCM model has successfully explained almost 80% of the observed picks, as revealed by the R-Squared value.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3258.86	3258.86	454.69	<.0001
Error	141	1010.57	7.17		
Correct. Total	142	4269.43			
Root MSE		2.68	R-Square	0.7633	
Dependent Mean		5.26	Adj R-Sq	0.7616	
Coeff Var		50.90865			

Parameter Estimates					
Variable	DF	Parameter	SE	t Value	Pr > t
Intercept	1	-0.478	0.35	-1.37	0.17
exp_freq	1	1.136	0.05	21.32	<.00

CONCLUSION

This paper discusses usage of monadic cards in forecasting consumer demands and preferences. It demonstrates one way of analyzing monadic

data to estimate utilities, using a real life forecasting project in consumer electronics. We show that even large monadic cards with about 160 valid cells can be analyzed successfully.

REFERENCES:

- Johnson, R. M. and Orme, B. K. "How Many Questions Should You Ask In Choice-Based Conjoint Studies?" ART Forum, Beaver Creek, 1996
- Louviere, J. J., Anderson, D., White, J. B., and Eagle, T. C. "Predicting Preferences for New Product Configurations: A High Tech Example", in Carnevale, M, Lucertini, M. and Nicosia, S (Ed.) Modeling the Innovation (North Holland, 1990)
- Louviere, J. J. and Woodworth, G., "Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregated data," Journal of Marketing Research 20(1983) 350-367
- Urban, G. L. and Hauser, J. "Design and Marketing of New Products," Prentice Hall, 1993

CONTACT INFORMATION

Jay Wu, Ph.D.
ACNielsen Vantis
3130 Crow Canyon Place, Suite 400
San Ramon, CA 94583
Phone: 925-327-2032
Fax: 925-820-1905
jay.wu@vantis.acnielsen.com

Barry Zacharias
ACNielsen Vantis
3130 Crow Canyon Place, Suite 400
San Ramon, CA 94583
Phone: 925-327-2022
Fax: 925-820-1905
barry.zacharias@vantis.acnielsen.com
www.vantisintl.com

Sample Monadic Choice Card for A Consumer Electronic Project																		
		Platform Type A					Platform Type B										Platform Type C	
		Format Y		Format X			Format Y			Format X							Format X	
Signal One		S1	S2	S1	S4	S6	S9	S11	S13	S5	S7	S10	S14	S2	S5	S8	S10	S12
Brand A	1	NA	NA	NA	NA	NA	Price 5	Price 7	Price 8	NA	NA	NA	NA	Price 47	Price 49	Price 54	NA	NA
Brand B	2	Price 6	Price 16	Price 18	Price 25	Price 32	Price 2	Price 9	Price 16	NA	NA	NA	NA	Price 47	Price 51	Price 54	NA	NA
Brand C	3	Price 8	Price 17	Price 20	Price 28	Price 35	Price 3	Price 10	Price 17	NA	NA	NA	NA	Price 47	Price 51	Price 54	NA	NA
Brand D	4	Price 4	Price 13	Price 17	Price 23	Price 30	Price 1	Price 7	Price 14	NA	NA	NA	NA	Price 49	Price 49	Price 53	NA	NA
Brand E	5	Price 11	Price 19	Price 22	Price 31	Price 39	Price 5	Price 15	Price 19	NA	NA	NA	NA	Price 47	Price 52	Price 54	NA	NA
Brand F	6	Price 8	Price 17	Price 20	Price 28	Price 35	Price 2	Price 9	Price 16	NA	NA	NA	NA	Price 49	Price 49	Price 53	NA	NA
Signal Two																		
Brand A	7	NA	NA	NA	NA	NA	Price 30	Price 37	Price 44	Price 46	Price 47	Price 48	Price 47	Price 49	Price 51	Price 54	Price 56	Price 60
Brand B	8	Price 16	Price 24	Price 27	Price 36	Price 42	Price 32	Price 37	Price 46	Price 46	Price 40	Price 46	Price 49	Price 51	Price 54	Price 54	Price 57	Price 60
Brand C	9	Price 17	Price 26	Price 29	Price 39	Price 44	Price 35	Price 40	Price 47	Price 48	Price 41	Price 47	Price 50	Price 51	Price 54	Price 55	Price 57	Price 60
Brand D	10	Price 12	Price 21	Price 25	Price 34	Price 40	Price 30	Price 34	Price 48	Price 45	Price 38	Price 43	Price 48	Price 48	Price 53	Price 55	Price 57	Price 60
Brand E	11	Price 19	Price 31	Price 33	Price 41	Price 47	Price 39	Price 42	Price 48	Price 49	Price 46	Price 48	Price 52	Price 51	Price 54	Price 55	Price 57	Price 61
Brand F	12	Price 17	Price 26	Price 29	Price 39	Price 44	Price 32	Price 37	Price 46	Price 46	Price 40	Price 46	Price 49	Price 48	Price 53	Price 55	Price 58	Price 61