Two Methods for Estimating Category Statistics – Which is Better?

Malcolm Wright

This paper reports partial results from a broader research programme into the use of purchase probability scales to replace panel data as a source of market statistics. In this part of the research two methods of estimating category penetration (number of buyers) and category purchase frequency were evaluated. One was a direct method that involved probabilistic questions about category consumption. The other was an indirect method that involved probabilistic questions about the major brands (and "other" brands), with the results subsequently aggregated to form category estimates. To evaluate the two methods, the probabilistic estimates were matched to panel observations for the same respondents in five product categories. The indirect method was slightly superior for estimating category penetration, but greatly superior for estimating category purchase frequency.

Keywords: Juster Scale, Verbal Probability Scale, Purchase Probability, Consumer Panel, NBD-Dirichlet Model.

Introduction

Is it better to measure product category penetration (the number of buyers) and average purchase frequency directly, from probabilistic questions asked about the whole category, or indirectly, by aggregating information obtained for each brand? This is worth knowing, because category penetration and category purchase frequency are statistics with a special role in market research and brand loyalty modelling.

The operators of consumer panels have long summarized purchase behaviour using measures such as category penetration and purchase frequency, market share, brand penetration and purchase frequency, share of category requirements, proportion of sole buyers, and repeat buying rates. Ehrenberg and his colleagues have shown how there are law-like relationships between these measures, such as Double Jeopardy, the Duplication of Purchase Law, and a variety of other regularities that are described and modeled by the well-known and highly successful NBD-Dirichlet model of purchase incidence and brand choice (Goodhardt, Ehrenberg & Chatfield 1984; Ehrenberg 1988; Ehrenberg, Goodhardt & Barwise 1990). Managers can analyse brand performance by examining trends over time, but also by comparing brand performance statistics with the theoretical norms provided by the NBD-Dirichlet model. Consumer panels provide the data for both these applications.

Unfortunately, there are many instances in which consumer panels are not available. Some countries have no consumer panels, many countries have panels for some categories but not others, and even when a country has a wide range of panels, there are still likely to be business-to-business, retail, and industrial market categories that are not covered.

Consequently, it would be useful to obtain key market statistics such as category penetration and purchase frequency, brand penetration and purchase frequency, and market share without having to rely on panel data. As well as being some of the most fundamental performance measures, these statistics are all that is required to estimate the NBD-Dirichlet model and generate theoretical norms for the other market statistics usually reported from consumer

panels. Thus, category penetration and purchase frequency, brand penetration and purchase frequency, and market share can be described as the key measures of purchase behaviour in a market, from which all other purchase behaviours can be derived.

Purchase recall and purchase intention - the most obvious alternative sources of data - have proved to be highly unreliable methods of obtaining such market statistics (Sudman 1964; Parfitt 1967; Sudman & Ferber 1974; Wind & Lerner 1979). Probabilistic approaches have generally performed much better (Juster 1966; Belk 1985; Jamieson & Bass 1989; Day, Gan, Gendall & Esslemont 1991; Marder 1997), but they have not yet yielded the range of market statistics required to estimate the NBD-Dirichlet model.

Stanton and Tucci (1982) did claim to show that recall measures could accurately reproduce panel diaries, as they found few significant differences in results at the aggregate level. However, their approach involved an extensive personal interview to gather 24-hour recall data on food category consumption, and comparison of this with diary information for the following two days. This hardly seems a challenging test of recall methods; when subjected to an extensive personal interview, there would be few people who could not remember what they had for lunch yesterday. Hu and Bruning (1988) conducted more interesting work. They used a survey to obtain pre-assessments of typical monthly usage rates for four methods of long distance communication, and compared this with 12 months of subsequent panel data from the same respondents. Their results showed a mean overstatement of 10% for the pre-assessments, although this was mostly present in two of the methods (long distance phone calls and cards), which consequently had almost twice this average level of error. The other two methods (letters and personal visits) had very low levels of error. This is a promising result, and in effect represents a naive version of one of the estimators used in this paper. However, Hu and Bruning's (1988) approach has only been tested in long distance communications, and unfortunately does not yield the required estimates of category penetration.

Consequently, this study sought to use the best developed probability scale, the Juster scale, as the basis for estimators of some of the market statistics usually found in consumer panels. The Juster scale was originally developed to predict purchasing of consumer durables, where it was found to outperform purchase intention scales (Juster 1966). It has since been extended to services (Clawson 1971, Day et al. 1991), fast moving consumer goods (Day et al. 1991), brands (Brennan & Esslemont 1994), telephone surveys (Brennan, Esslemont & Hini 1995a), multiple purchase quantities (Hamilton-Gibbs, Esslemont & McGuinness 1992, Seymour, Brennan & Esslemont 1994, Brennan, Esslemont & Hini 1995b), self completion questionnaires (Gendall, Esslemont & Day 1991), and customer loyalty (Danenberg & Sharp 1996). In this paper, the focus is on the use of the Juster scale to estimate *category* penetration and purchase frequency, and in particular on the choice between direct and indirect estimation of these statistics. The use of the Juster scale to estimate brand-level statistics is taken up elsewhere (Wright, Sharp & Sharp 2002).

The rest of this paper sets out the direct and indirect methods of estimating category statistics using the Juster scale, describes the data available to assess these two methods, and presents the results of a comparison of the methods across five product categories. For three of these categories, indirect method results were reported in Wright et al. (2002) as part of a general validation of Juster estimators of key market statistics; however, no comparison was made with the direct method. By making this comparison, this paper justifies the choice of the indirect method in Wright et al. (2002).

Method

The Survey Questions

The estimators were based on the Juster scale (Juster 1966, shown in Figure 1) or the Verbal Probability scale (Brennan et al. 1995a), a spoken analogue of the Juster scale which can be applied using a telephone survey (see Appendix A for examples of the full question wording for each scale). The Juster scale has previously been used to estimate category penetration (Juster 1966; Clawson 1971; Gabor & Granger 1972; Day et al. 1991) but validation has typically relied on purchase recall, a somewhat unreliable measure, with considerable fluctuation in results. In this research, panel observations are the basis for comparison.

Figure 1: The Juster Scale

----	----	10	- CERTAIN, PRACTICALLY CERTAIN (99 in 100)
----	----	9	- ALMOST SURE (9 in 10)
----	----	8	- VERY PROBABLE (8 in 10)
----	----	7	- PROBABLE (7 in 10)
----	----	6	- GOOD POSSIBILITY (6 in 10)
----	----	5	- FAIRLY GOOD POSSIBILITY (5 in 10)
----	----	4	- FAIR POSSIBILITY (4 in 10)
----	----	3	- SOME POSSIBILITY (3 in 10)
----	----	2	- SLIGHT POSSIBILITY (2 in 10)
----	----	1	- VERY SLIGHT POSSIBILITY (1 in 10)
----	----	0	- NO CHANCE, ALMOST NO CHANCE (1 in 100)

The number of buyers, or penetration, is estimated simply as the mean Juster (or Verbal Probability) score, adjusted appropriately to convert it to either a proportion or a percentage.

An alternative approach is to substitute the numerical descriptions for the scale points, using .99 instead of 1 for a Juster score of 10, and .01 instead of 0 for a Juster Score of zero. The zero point adjustment is necessary when fitting a probability distribution to the data, as maximum likelihood estimation techniques will fail if required to estimate the logarithm of zero. However, the endpoint adjustment seems otherwise unattractive. Conceptually, it must introduce a systematic bias in extreme cases, when either everybody or nobody performs a behaviour. It is also unclear how it can be justified when using the Verbal Probability scale when there is no verbal prompt corresponding to the endpoint adjustment. In any event, the number of zero scores seldom exceeds the number of 10 scores by more than 40% of responses, so the maximum effect of the endpoint adjustment would be an increase in the penetration estimate of .004 or .4 percentage points.

In addition to estimating penetration, an estimate must also be obtained for the volume of purchase occasions. Brennan et al. (1995b) found that a question along the lines of "*and how many times are you most likely to* <*buy/visit/shop at*> <brand j> *in the next* <period>"

worked best for this purpose. Their approach is adopted in this research, including weighting the answer by the probability of purchase.

Category Penetration Estimators

There are two methods of probabilistically estimating category penetration (B); directly from questions about the overall category, or indirectly from questions about each brand (all major brands plus a question on "other" brands).

The direct method involves the standard Juster procedure of averaging the individual purchase probabilities, which are obtained through a question such as "*now, taking everything into account, and using the Juster scale, what are the chances that you, personally, will <buy/shop at>* <brand j> *within the next* <period>". Equation (1) expresses this as a percentage.

$$B_{direct} = ((\Sigma_i P_i) / n) * 100 \quad (1)$$

where P_i is the probability that individual *i* will purchase from the category in the relevant period.

The indirect method involves identifying the chances that respondents will *not* buy. First, all brand purchase probabilities are subtracted from one to give the probability of *not* buying each brand. Next, for each individual, the product (Π) of the probabilities of not buying each brand is calculated. This product is equivalent to the probability, for that individual, of not buying *any* brand at all (ie. not buying the category). The probability that this individual *will* buy from the category is then just one minus the probability of not buying any brand at all. These individual purchase probabilities are then averaged to give a population estimate. Equation (2) gives the relevant estimator reported from Wright et al. (2002), again expressed as a percentage.

$$B_{indirect} = (\Sigma_i (1 - \Pi_j (1 - p_{ij})) / n) * 100 \quad (2)$$

where p_{ij} is the probability that individual *i* will purchase from brand *j* in the relevant period.

Category Purchase Frequency Estimators

Category purchase frequency (W) can also be estimated either directly from questions about the most likely number of purchases from the category, or indirectly from questions about the most likely number of purchases from each brand (all major brands plus a question on "other" brands). The direct method requires a direct estimate of total category volume to be divided by the direct estimate of the number of buyers (Equation 3). The indirect method, reported in Wright et al. (2002) requires the estimate of the individual brand volumes to be summed and then divided by the indirect estimation of the number of buyers (Equation 4).

$$W_{direct} = (\Sigma_i (P_i * V_i)) / (n * B_{direct}) \quad (3)$$

where V_i is the most likely number of purchases by individual *i* from the category, and $n * B_{direct} = \Sigma P_i$ and it does not matter which is used as the denominator in Equation 3.

$$W_{indirect} = (\Sigma_i \Sigma_j (p_{ij} * v_{ij})) / (n * B_{indirect}) \quad (4)$$

where v_{ij} is the most likely number of purchases by individual *i* from brand *j*.

Data

Three data sets were available covering five categories. In each case, the Juster estimates were obtained for a four-week period, and matched to four-weekly diary panel observations for the same respondents. As the same respondents were used, there is, strictly speaking, no sampling error. However, panel observations are known to fluctuate considerably from period to period, so adjacent four-weekly panel observations were pooled to minimise these fluctuations. For example twelve weeks of data would be divided into three four-weekly periods, and an average of these taken. This procedure assumes market stationarity, treating the Juster estimates as a reflection of the normal or steady state of the market.

The details of the panels were as follows.

- A panel of individuals' Bill Payment methods from an Australian city (eg. by mail, by phone, in person, etc). The Juster estimators were applied during the panel recruitment telephone survey using the Verbal Probability scale. They were subsequently matched to ten weeks of panel observations, yielding a useable sample size of n = 365. Only the first eight weeks of the panel, representing two four-week periods, were used in the analysis.

- A panel of car drivers' Retail Fuel purchase occasions in New Zealand. The Juster estimators were again applied during panel recruitment telephone survey using the Verbal Probability scale. There were again ten weeks of panel observations with just the first eight weeks, or two four-week periods, used in the analysis. The useable sample size was n = 587.

- A panel of individuals' Supermarkets, Retail Fuel, and Department Store purchase occasions in an Australian city (the data set used by Wright et al. 2002). In this case the Juster scale was mailed to respondents some months after the completion of the panel, and data then collected via telephone. After matching, the useable sample size was n = 271, with panel data pooled over three four-week periods. The Juster estimates applied to October, while the matched panel data covered a period five to eight months earlier. It turned out that seasonal differences were present in the Department Store category, undermining the assumption of market stationarity and confounding the results. To deal with this, the Juster estimates for this category were instead matched to smaller panel data set from October in the previous year, yielding a useable seasonally matched sample size of n = 138 which covered a single four-week period.

Results

To assess the direct and indirect methods, Juster estimates were calculated and compared with the corresponding panel observations. Three measures of fit were calculated: Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE) and simple correlation (r).

Table 1. Category Penetration (B)

	Panel Observation	Indirect Estimate	Direct Estimate
Supermarkets	98	96	96
Bill Payment	98	97	95
Retail Fuel - NZ	96	91	92
Retail Fuel - Aus.	87	81	83
Department Stores	83	80	60
MAD		3%	7%
MAPE		.04	.08
r		.98	.93

Table 1 shows that both methods gave reasonable estimates of category penetration, although the direct method showed considerable error in the Department Store category. While both methods showed high correlation with the panel observations (demonstrating that the *order* across the different categories was adequately captured), the indirect method had much lower MAD and MAPE statistics. Indeed, the fit statistics for the indirect method are objectively excellent, as well as being relatively better.

Table 2. Category Purchase Frequency (W)

	Panel Observation	Indirect Estimate	Direct Estimate
Supermarkets	9.9	9.2	7.0
Bill Payment	8.4	5.4	3.6
Retail Fuel - NZ	5.1	4.5	4.4
Retail Fuel - Aus.	4.5	4.3	4.3
Department Stores	4.4	4.7	2.7
MAD		0.96	2.06
MAPE		13%	29%
R		.87	.68

Table 2 shows that the indirect method also outperformed the direct method when estimating category purchase frequency, although in this case the difference is more marked. In particular, the direct method performed poorly for Supermarkets, Bill Payment, and Department Stores. While the indirect method performed relatively poorly for Bill Payment, it still gave much better results than the direct method for this category. The fit statistics for the indirect method are again both relatively better and objectively excellent, although they are not quite as good as the 10% mean overstatement achieved by Hu and Bruning (1988). This is due to the poor fit for Bill Payment; for the other four categories, the MAPE is just 7%.

Conclusion

When using purchase probability scales to estimate category penetration and purchase frequency, better estimates can be obtained by aggregating brand statistics than by asking about the category directly.

Direct estimates of category purchase frequency are particularly prone to misleading results, in this case considerable under-estimates for three out of five categories. The direct estimates of category penetration were somewhat better, but still not as good as the indirect estimates. By contrast the indirect estimates were not only relatively better but were also objectively excellent, as reported in Wright et al. (2002), with overall levels of error no greater than the normal period-to-period fluctuations seen in panel data. The one exception is the somewhat low estimate of category purchase frequency for Bill Payment. However, this estimate still represents a considerable improvement on the direct estimate for that category.

As noted earlier, other work has been undertaken to validate Juster estimators of a range of key market statistics (Wright et al. 2002). Further research is needed to develop probabilistic estimators for other market statistics, and for longer and shorter time periods, to allow brand managers to determine whether their brands are deviating from the theoretical norms generated by the NBD-Dirichlet model in these areas. Also, all the categories investigated in this article were relatively frequently used, and it would be useful to extend this research to less frequently purchased categories as well.

Meanwhile, it seems clear that indirect probabilistic methods are the better approach to estimating category statistics in the absence of panel data.

References

Belk RW (1985). Issues in the Intention-Behaviour Discrepancy, in *Research in Consumer Behaviour*, J N Sheth (Ed.) Vol. 1. Greenwich, CT: JAI Press.

Brennan M & Esslemont D (1994). The Accuracy of the Juster Scale for Predicting Purchase Rates of Branded, Fast-Moving Consumer Goods. *Marketing Bulletin*, 5, 47-53.

Brennan M; Esslemont D & Hini D (1995a). Obtaining Purchase Predictions Via Telephone Interviews. *Journal of the Market Research Society*, 37 (3), 241-250.

Brennan M; Esslemont D & Hini D (1995b). A Test of Three Methods for Estimating Level of Purchase, in *Academy of Marketing Science - World Marketing Congress*, K Grant and I Walker (Eds.) Vol. VII-I. Melbourne.

Clawson JC (1971). How Useful Are 90-Day Purchase Probabilities? *Journal of Marketing*, 35 (October), 43-37.

Danenberg N & Sharp B (1996). Measuring Loyalty in Subscription Markets using Probabilistic Estimates of Switching Behaviour, in *Australia New Zealand Marketing Educators Conference* Vol. 1. Auckland.

Day D; Gan B: Gendall P & Esslemont D (1991). Predicting Purchase Behaviour. *Marketing Bulletin*, 2, 18-30.

Ehrenberg ASC (1988). *Repeat-Buying: Facts, Theory and Applications* (2nd ed.). London: Griffin.

Ehrenberg ASC; Goodhardt GJ & Barwise P (1990). Double Jeopardy Revisited. *Journal of Marketing*, 54 (July), 82-91.

Gabor A & Granger C (1972). Ownership and Acquisition of Consumer Durables: Report on the Nottingham Consumer Durables Project. *European Journal of Marketing*, 6 (4), 234-248.

Gendall P; Esslemont D & Day D (1991). A Comparison of Two Versions of the Juster Scale Using Self-Completion Questionnaires. *Journal of the Market Research Society*, 33 (3), 257-263.

Goodhardt GJ; Ehrenberg ASC & Chatfield C (1984). The Dirichlet: A Comprehensive Model of Buying Behaviour. *Journal of the Royal Statistical Society*, 5, 621-655.

Hamilton-Gibbs D; Esslemont D & McGuinness D (1992). Predicting the Demand for Frequently Purchased Items. *Marketing Bulletin*, 3, 18-23.

Hu M & Bruning E (1988). Using Prior Experience to Explain Survey Versus Diary Recorded Usage Data. *Journal of the Market Research Society*, 30 (1), 59-72.

Jamieson LF & Bass FM (1989). Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods. *Journal of Marketing Research*, 26 (August), 336-345.

Juster TF (1966). Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design. *National Bureau of Economic Research*. New York: Columbia University Press.

Marder E (1997). *The Laws of Choice: Predicting Customer Behaviour*. New York: Simon and Schuster.

Parfitt JH (1967). A Comparison of Purchase Recall with Diary Panel Records. *Journal of Advertising Research*, 7, 16-31.

Seymour P; Brennan M & Esslemont D (1994). Predicting Purchase Quantities: Further Investigation of the Juster Scale. *Marketing Bulletin*, 5, 21-36.

Stanton JL & Tucci LA (1982). The Measurement of Consumption: A Comparison of Surveys and Diaries. *Journal of Marketing Research*, 19 (May), 274-277.

Sudman S (1964). On the Accuracy of Recording of Consumer Panels: II. *Journal of Marketing Research*, (August), 69-83.

Sudman S & Ferber R (1974). A Comparison of Alternative Procedures for Collecting Consumer Expenditure Data for Frequently Purchased Products. *Journal of Marketing Research*, 11 (May), 128-135.

Warshaw PR (1980). Predicting Purchase and Other Behaviors from General and Contextually Specific Intentions. *Journal of Marketing Research*, 17 (February), 26-33.

Wind Y& Lerner D (1979). On the Measurement of Purchase Data: Surveys Versus Purchase Diaries. *Journal of Marketing Research,*16 (February), 39-47.

Wright M; Sharp A & Sharp B (2002). Market Statistics for the Dirichlet Model: Using the Juster Scale to Replace Panel Data. *International Journal of Research in Marketing*, 19, (1), 81-90.

Acknowledgements
Thanks to Anne Sharp and John Dawes for assistance with data collection

Malcolm Wright is a Senior Lecturer at Massey University

Appendix A - Examples of Question Wording

The Juster Scale (mailed out)

Please look at the Juster Scale on the back of the letter. As you know we are going to be asking you about the chances of you doing various things in the next four weeks. That is, we will be asking you about what you will be doing in the 28 days between now and <*date*>.

For each question, if you were certain that you would do it, that is if the chances were 100% or 99% you would answer 10. If you thought there were, say, 3 chances in 10 you would answer 3, and so on.

Now, using the Juster scale, and taking everything into account, what are the chances that you personally will <buy/shop at > <*brand j*> within the next <*period*>.

The Verbal Probability Scale

Now I would like to ask you about your chances of <buying/shopping at> <*category*>. How probable is it that you personally will <buy/shop at > <*brand j*> within the next <*period*>?

I would like you to answer on a scale of 0 to 10. If you are certain, or practically certain that you will <buy/shop at > <*brand j*> within the next <*period*> you should choose the answer 10. If you think that there is no chance, or almost no chance, the best answer would be zero. If you are uncertain about the chances choose another number as close to zero or 10 as you think it should be. You can think of the numbers as chances out of 10. For example, 3 would mean 3 chances out of 10 that you will <buy/shop at > <*brand j*> within the next <*period*>, while a 7 would mean 7 chances out of ten, and so on.

So, taking everything into account, and using a scale of zero to 10, how probable is it that you will <buy/shop at > <*brand j*> within the next <*period*>?

<u>Verbal Probability Scale - Second Consecutive Reading (eg. next brand)</u>

And how probable it is that you will <buy/shop at > <*brand j+1*> within the next <*period*>? Again, I would like you to answer on a scale of 0 to 10. If you are uncertain, you should choose an answer as close to zero or 10 as you think it should be. For example, 3 would mean 3 chances out of 10 while a 7 would mean 7 chances out of ten, and so on. So, taking everything into account, how probable is it that you will <buy/shop at > <*brand j+1*> within the next <*period*>?

<u>Verbal Probability Scale - Third and Subsequent Consecutive Readings</u>

And how probable it is that you will <buy/shop at > <*brand j+2*> within the next <*period*>? Again, please answer on a scale of 0 to 10.