

# A Comparison of Ranking, Rating and Reservation Price Measurement in Conjoint Analysis

SHLOMO KALISH

*Leon Recanati Graduate School of Business Administration, Tel Aviv University, Tel Aviv, Israel 69978*

PAUL NELSON

*William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, New York 14627*

**Key words:** Reservation Price, Attribute Levels, Conjoint Analysis

[January 1991]

## *Abstract*

This paper empirically compares the traditional preference measures of ranking and rating in conjoint analysis with a direct monetary measure of product value – reservation prices. Experimental results are as expected. While reservation prices do very well in terms of fit, they are inferior in terms of predicting choice on a holdout sample. In addition, surprisingly little difference is found in the performance of ranks and ratings.

In recent years the reservation price (RP) concept has received considerable use in theoretical and algorithmic papers concerning, for the most part, the pricing and positioning of a product line and price promotion (see, e.g., Rosen, 1974; Varian, 1980; Moorthy, 1984; Oren, Smith and Wilson, 1984; Kalish, 1985; Dobson and Kalish, 1988; Gerstner and Hess, 1990; Moorthy and Png, 1991). These papers model the consumer's willingness to pay (reservation price) for a product as depending on the product's attribute levels (exclusive of price) and base choice on the difference between the purchase price and the consumer's RP for the product.<sup>1</sup> Key advantages of this methodology relative to the more widely established multiattribute (MA) model are: (i) the RP model is directly connected with traditional economic demand theory (e.g., the demand function plots the number of consumers with a reservation price greater than a said price vs. price); and (ii) the model provides a simplified algorithmic framework in which to evaluate price and positioning alternatives through choice simulation. Despite the analytical appeal, empirical applications have been few (see, e.g., Pessemier, 1975; Hauser and Urban, 1986; Cameron and James, 1987). Given the wide acceptance of the MA model in pricing and positioning applications and that reservation prices are more difficult to measure (i.e., response difficulty is higher) than preference ranks or ratings this is not unexpected. Is this failure to use the RP methodology in empirical applications warranted? Measurement problems seem to imply the an-

swer is yes. However, an empirical investigation has not been undertaken. Our objective is to compare the RP approach to the traditional MA model primarily from the predictive validity perspective. Using consumer survey results, we compare the predictive ability and goodness of fit of the RP model using directly measured reservation prices with those of the MA model using the common rank and rating preference measures.<sup>2</sup>

The remainder of the paper consists of four sections. First, a brief review of the reservation price concept is given. Next, the experimental design is presented. Experimental results then are discussed and, finally, conclusions presented.

## 1. Background

The reservation price approach hypothesizes that each customer has a maximum price they are willing to pay for a given product which equals the product's value to the consumer. This price is the consumer's reservation price for the product. The consumer compares her reservation price for each product with its purchase price and chooses the product that offers the largest differential.<sup>3</sup> That is, the consumer chooses the product that maximizes her utility,  $U(z,P) = R(z) - P$ , where  $P$  is the purchase price and  $R(z)$  is the consumer's reservation price for a product consisting of attribute bundle  $z$ .  $R(z)$  is thus a monetary (ratio) scaled measure of preference that *does not* take price into account. Alternatively, the traditional MA model treats price as an additional attribute, utility =  $f(z,P)$ . Thus, utility measures used in estimation (ranks and ratings) are "util" (interval or ordinal) scaled measures of preference which *do* take price into account. As with the RP model, the consumer chooses the product with the highest utility. It follows that the two methodologies are analytically equivalent if price enters the MA model linearly (Ratchford, 1979; Srinivasan, 1982). In this case, utility, when re-scaled by dividing by the price coefficient is measured in monetary units. However, two key differences between the models arise with respect to measurement and application: (i) preference as measured by stated reservation prices does not take price into account; and (ii) reservation prices are ratio scaled.

Conjoint analysis using directly measured reservation prices as input has both advantages and disadvantages relative to use of either preference rankings or ratings as input. Since it is likely that consumer input and response difficulty are larger for reservation prices than for ranks or ratings, poorer fit and predictive ability may be expected. Alternatively, product descriptions need not include price. Consequently, the number of product attributes is reduced by one. Since the number of product concepts evaluated by the respondent is typically a limiting factor, this reduction may be significant. In some applications a monetary valued utility function allows easier and more intuitive analysis (Parker and Srinivasan, 1976; Srinivasan, 1979; Ratchford, 1979). For example, in the linear model the value of each coefficient is the monetary value to the consumer of an additional unit of the attribute. Directly measured reservation prices are ratio scaled which poses no technical problems in their use in regression and other metric estimation

techniques. This scale also is comparable across individuals, which eliminates difficulties of interpersonal utility comparisons that exist with rating and ranking scales (see, e.g., Green and Krieger 1985; Gupta and Kohli, 1990).

## 2. Experimental design

A conjoint analytic framework using full profile representations of twelve hypothetical products was used. Each respondent stated their preferences over the set of products using only one of the three preference measures (ranking, rating and reservation price). The products were round-trip airline tickets from New York City to Hawaii. The products were described by three attributes with levels defined relative to currently available products. The attributes were service level (minimum, regular or premium), seating room (regular or spacious), and whether the flight was non-stop or not. This resulted in a total of twelve combinations. Product descriptions included price when ratings and rankings were collected but did not include price when reservation price data were collected. The price of each combination was determined by part worth costing of the features, where the cost of each feature reflected their approximate market prices at the time. To avoid correlation problems \$100 was either added or subtracted from the full cost of each product.

There are three ways to compare the different measurement scales. The most direct comparison is to evaluate the consistency of preferences across the three scales. To what extent are the preferences implied by the different measurement scales similar? Do the estimated attribute (and price) weights differ? To investigate these questions one needs to apply at least two different scales to each individual. This was not pursued because in a preliminary study we found respondent workload to be quite high and learning effects to be significant. A second comparison measure is the fit of the estimated model to the original preferences. The third measure, which is the most interesting from a practical point of view, is predictive validity. To investigate this issue, after some other tasks each respondent regardless of measurement scale was again presented with four of the hypothetical products now with different prices and asked to state their first choice, second choice, etc.

The questionnaires were administered to undergraduate and first year graduate students in six different business classes. In order to minimize response bias the classes were chosen so that none of the students had previous class exposure to conjoint analysis. This resulted in 255 useable questionnaires, with approximately one third of the sample using each measurement method. Ranks were measured in the usual way. The respondent was asked to rank the products from most preferred to least preferred. Ratings were measured on a scale of 0 to 100 points in five-point intervals. For reservation prices the respondent was asked to indicate what sum of money would make them indifferent between the product and the money.

Estimation was done at the individual level. Ranks and ratings were fit as a

linear function of the products' attribute levels and prices. Stated reservation prices were fit as a linear function of the products' attribute levels only.<sup>4</sup> The respondent's order of preference taking price into account was predicted for both the original twelve products and the four holdout products. For ratings and rankings,  $U(z,P)$  was estimated and then used to predict the order of preference on both the calibration and holdout samples. More specifically, stated preference ranks and ratings are used as proxies for product utilities in the estimation of  $\hat{U}(z,P) = \hat{\alpha} + \sum \hat{w}_j z_j - \hat{\beta}P$ . The  $\hat{w}_j$ 's and  $\hat{\beta}$  are the estimated importance weights of the attributes and price, respectively. For reservation prices,  $R(z)$  was estimated and then the estimated  $R(z)$  less price used to predict order of preference taking price into account on both samples. That is,  $\hat{R}(z) = \hat{c} + \sum \hat{\alpha}_j z_j$  is estimated using stated reservation prices as the measure for  $R(z)$ . The  $\hat{\alpha}_j$ 's are the estimated marginal values of the attributes to the consumer. The stated preference ordering including price effects is achieved by an ordering of  $R(z) - P$  where  $P$  is the price at the attribute bundle given in the rank and rating full profile descriptions. The estimated preference ordering including price is arrived at by an ordering of  $\hat{R}(z) - P$ .<sup>5</sup>

Estimation with all three preference measures was done using OLS regression analysis. While this is clearly justified for the monetary scaled measurements (since they are ratio scaled), it is partly justified for ratings and less so for ranks. However, since this is the most widely used estimation procedure for these cases (Cattin and Wittink, 1982) and there is little difference between OLS and monotonic regression results (see, e.g., Green and Srinivasan, 1978), we used this method.

### 3. Results

#### 3.1. Goodness of fit

Fit was measured using the average across individuals of the Spearman rank correlation between the ranks of the actual measurements and fitted values.<sup>6</sup> Table 1 reports the average fit statistics. Fit is high in all cases. Fit is best when reservation prices are used. A likely rationale for this result is that reservation prices

Table 1. Average goodness of fit statistics across individuals

Measurement scale	Spearman rank correlation	Sample size
Rank	.91	68
Rate	.88	97
Reservation Price	.96	89

Note: The numbers reported are the averages of the rank correlation between the implied stated preference ranks and the fitted ranks.

allow a much larger range of values for the dependent variable than do ratings or rankings. Consequently, better discrimination among the calibration products results. Some bias in favor of ranks and ratings arises because five rather than four parameters are estimated. However, this is mitigated since the RP model incorporates price before preferences (which thus includes price effects) are ranked relative to the other products. Alternatively, ratings score lowest. The fact that fit is lower for ratings than rankings could be a result of response style bias (Kalwani and Silk, 1982).

### 3.2. Predictive validity

A more interesting comparison criteria is the ability to predict choice or preference. While actual choices would be ideal, we only have simulated choices, as described above.<sup>7</sup> As with the fit, a statistic that is comparable across the different measurement scales is the rank correlation between stated and predicted preference ranks. Contrary to goodness of fit measures, there are no biases of the kinds reported earlier. We report both the average and median Spearman rank correlations. We also report the proportion of correctly predicted first choices. Results are presented in Table 2.

In all cases, predictive ability is significantly above the random choice yardsticks of 25% for the percentage of correctly predicted first choices and zero for the rank correlation. However, the rank and rate measurements consistently outperform reservation prices. Whereas the proportion of correctly predicted first choices is 62% when ranks or ratings are used, it is 46% for reservation prices. In addition, the average and median rank correlations are consistently larger for ranks and ratings. Note the consistent superiority of the median over the mean rank correlations. This is caused by a limited number of respondents for whom preferences are predicted quite poorly, possibly due to lack of interest.<sup>8</sup>

At least for this data set, the MA model using either a ranking or ratings predicts better than the RP model using directly measured reservation prices. In order to

Table 2. Prediction of choices on holdout products-average and median rank correlations and proportions of correct first choices

Comparison method	Measurement scale	Average rank correlation	Median rank correlation	Percent #1 correct
Aggregate	Rank	.57	.80	62
	Rate	.49	.80	62
	Reservation Price	.43	.60	46

Note: The correlations reported are the averages and medians of the rank correlations between the choice ranking of the four holdout products and their predicted ranks. The percentage of correct first choices is the percentage of respondents for whom their stated and estimated first choice are the same.

test whether these apparent differences are statistically significant, we formulated the following regression and logit models:

$$\text{Correlation}_{ij} = \mu + \eta_j + \varepsilon_{ij}$$

and

$$\text{Probability}_{ij} (\text{first choice correctly predicted}) = \frac{1}{\{1 + \exp[-(\mu + \eta_j)]\}}$$

In the regression model,  $i$  is the individual index,  $\mu$  is the base case (reservation price) rank correlation,  $\eta_j$  is the average correlation difference between the other two measurement scales and the base case, and  $\varepsilon_{ij}$  is the correlation unexplained by the measurement scales. In the logit model,  $i$  is the individual index,  $\mu$  is the base case (reservation price) effect on the probability, and  $\eta_j$  is the average additional effect on the probability which is related to the other two measurement scales.

We investigated the statistical significance of the differences in the rank correlation and the proportion of correctly predicted first choices reported in Table 2 (i.e., whether the  $\eta_j$ 's differ from each other or zero). Ranking, which has the largest average rank correlation, is found to be statistically significantly different at the 9% level from reservation prices.<sup>9</sup> The hypotheses that the average correlations are equal when ranks and ratings or ratings and reservation prices are used cannot be rejected at any reasonable significance level. In addition, the hypothesis of no difference between the three measurement scales can not be rejected ( $P(>F) = 24\%$ ). Logit analysis relating the ability to predict first choices to the three measurement scales provides more significant results. The hypothesis of no difference between the three scales can be rejected at the 5% level. Furthermore, ratings outperform reservation prices in a statistically significant manner (3% level) while rankings show a statistically significant (5% level) improvement over reservation prices. No statistically significant difference between rankings and ratings is found.<sup>10</sup>

#### 4. Conclusions

While goodness of fit favors the use of reservation prices, predictive validity results support the use of the traditional ranking and rating measures. In terms of correctly predicting first choices on a holdout sample, this superior performance is statistically significant. Predictive validity differences as measured by Spearman's rho are not, in general, statistically significant. In addition, no significant differences in fit or predictive validity are found between the use of ranks or ratings with the MA model. These results, in contradiction to the current trend towards the use of ratings, favor simple ordinal comparison of products.

Note that the results reported above pertain to a single study of a single product class. They may or may not be generalizable. To begin evaluation of this issue, using a similar experimental design a series of small studies (with about 30 respondents each from the actual target population) concerning buyers in a variety of product classes was undertaken. Product classes included were service elevators for single family homes, cleaning robots, hotel rooms, television, day care, soft drinks, and automobiles. The results of these studies support the findings reported above. Ranking averaged 56% correct first choice prediction, while RP averaged only 40%. This difference was larger for the low ticket items and very small to negative for the high ticket, high involvement items (cleaning robot, automobiles and elevator). Based on preliminary evidence it appears that when respondents are very well informed all three measures perform roughly equally. Given the similarity of the RP and MA models it is not surprising that consumer use of each is supported. However, RP's generally inferior performance supports the notion that RP measurement is not as robust to respondent involvement as are ranks and ratings. (That is, measurement errors by the respondent or researcher are more likely and their effect on estimated product evaluations are more pronounced.) Consequently, given the current state of RP measurement and the lower response difficulty associated with ranks and ratings, continued use of the MA model with either ranks or ratings is supported. Future improvements in RP measurement, however, may make it an acceptable alternative to ranks and ratings in conjoint analysis.

## Notes

1. We use the term product to mean a specific brand/model within a particular product class.
2. Hauser and Urban (1986) undertake a comparison of directly measured reservation prices and stated purchase probabilities. However, they look at budget priorities across product classes rather than preferences within a single product class. They find estimation using purchase probabilities to outperform reservation prices but do not undertake a statistical comparison.
3. Economists refer to this differential as consumer surplus. Hauser and Urban (1986) call it net value.
4. We also fitted a main effects part worths model but it provided no statistically significant improvement in fit or predictive validity. Since only one attribute has more than two levels this is not surprising. Consequently, we report the linear model results.
5. The RP model and measurement is analogous to cost-benefit analysis (Benefit - Cost = Net Gain). Rather than measure and then estimate each product's net gain directly, only the benefit associated with each product is measured and then estimated or predicted. Estimated or predicted net gain for a product is derived through comparison of the product's benefit with its cost (i.e., price).
6. We cannot use more traditional measures of fit (e.g.,  $R^2$ ) since the dependent variables are measured on different scales.
7. Note that all the following tests are really joint tests of the choice model and measurement instruments.
8. Although the questionnaires were administered during a snowy winter in Rochester, N.Y., lack of interest by some respondents, especially given their student status, is likely. Improved data

quality through use of respondents known to be knowledgeable or purchasers of the product class should result in better fit and prediction relating to all three preference measures. Reservation prices may benefit most due to their greater response difficulty. We found RP answers by some respondents to be fairly uniform across products and some people stated that they had difficulty providing RP measurements. Both observations degrade fit and predictive validity. A given reference price may be used to address the second observation but we felt this might bias the RP measurements. Certainly if the respondent is knowledgeable of the product class no need to provide a reference price arises.

9. All significance levels relating to t or z statistics refer to two sided hypothesis tests.
10. In addition to the direct measurements, we also collected paired product comparison data using each of the three preference measures. The relative predictive abilities of the measures is consistent with those of the direct measurements. In addition, for all three measures the percentage of correctly predicted first choices and the rank correlation are statistically significantly less than they are for direct measurement. This finding confirms Green and Srinivasan's (1978) conjecture that a rank ordering should do better than simple paired comparisons. Our results imply this is true for ratings and reservation prices as well.

## References

- Cameron, T. A., and M. D. James. (1987). "Estimating Willingness to Pay from Survey Data: An Alternative Pre-Test-Market Evaluation Procedure," *Journal of Marketing Research* 24, 389-395.
- Cattin, P., and D. R. Wittink. (1982). "Commercial Use of Conjoint Analysis: A Survey," *Journal of Marketing* 46, 44-53.
- Dobson, G., and S. Kalish. (1988). "Positioning and Pricing a Product Line: Formulation and Heuristics," *Marketing Science* 7, 107-125.
- Gerstner, E., and J. D. Hess. (1990). "Can Bait and Switch Benefit Consumers," *Marketing Science* 9, 114-124.
- Green, P. E., and A. M. Krieger. (1985). "Models and Heuristics for Product Line Selection," *Marketing Science* 4, 1-19.
- Green, P. E., and V. Srinivasan. (1978). "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research* 5, 103-123.
- Gupta, S., and R. Kohli. (1990). "Designing Products and Services for Consumer Welfare: Theoretical and Empirical Issues," *Marketing Science* 9, 230-246.
- Hauser, J. R., and G. L. Urban. (1986). "The Value Priority Hypothesis for Consumer Budget Plans," *Journal of Consumer Research* 12, 446-462.
- Kalish, S. (1985). "New Product Adoption Model with Price, Advertising and Uncertainty," *Management Science* 31, 1569-1585.
- Kalwani, M., and A. J. Silk. (1982). "On the Reliability and Predictive Validity of Purchase Intention Measures," *Marketing Science* 1, 243-286.
- Moorthy, K. S. (1984). "Market Segmentation, Self Selection and Production Line Design," *Marketing Science* 4, 288-307.
- Moorthy, K. S., and I. P. L. Png. (1991). "Market Segmentation, Cannibalization, and the Timing of Product Introductions," MERC Working Paper 90-11, William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, New York.
- Oren, S., S. Smith, and R. Wilson. (1984). "Pricing a Product Line," *Journal of Business* 57, S73-S100.
- Parker, B. R., and V. Srinivasan. (1976). "A Consumer Preference Approach to the Planning of Rural Health Care Facilities," *Operations Research* 24, 991-1025.

- Pessemier, E. A. (1975). "Market Structure Analysis of New Product and Market Opportunities," *Journal of Contemporary Business* 4, 35-67.
- Ratchford, B. T. (1979). "Operationalizing Economic Models of Demand for Product Characteristics," *Journal of Consumer Research* 6, 76-85.
- Rosen, S. (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 34-55.
- Srinivasan, V. (1979). "Network Models for Estimating Brand-Specific Effects in Multi-Attribute Marketing Models," *Management Science* 25, 11-21.
- Srinivasan, V. (1982). "Comments on the Role of Price in Individual Utility Judgements," Choice Models for Buyer Behavior, *Research in Marketing* Supplement 1, JAI Press, Inc., 81-90.
- Varian, H. (1980). "A Model of Sales," *American Economic Review* 70, 651-659.