ITS

WORKING PAPER
ITS-WP-00-05

The Use of Mixtures of Market and Experimental Choice Data in Establishing Guideline Weights for Evaluating Competitive Bids in a Transport Organisation

By

David A. Hensher,
Jordan J. Louviere, &
David E. Hansen.

March, 2000

ISSN 1440-3501

Established and supported under the Australian Research Council's Key Centre Program.

INSTITUTE OF TRANSPORT STUDIES
The Australian Key Centre in Transport Management

The University of Sydney and Monash University

NUMBER: Working Paper ITS-WP-00-05

TITLE: The Use of Mixtures of Market and Experimental Choice Data in Establishing Guideline Weights for Evaluating Competitive Bids in a Transport Organisation

ABSTRACT: The government sector is increasingly using competitive bidding for service deliveries such as the provision of bus and rail services as well as the purchasing of professional engineering services such as project planning, design, and project supervision. As part of a program to simplify and introduce consistency in the tender evaluation process, one government transport agency in New Zealand financed a study to investigate the potential of combined revealed and stated preference methods as a way of establishing weights to attach to the criteria used to evaluate offers of engineering services. These techniques have mainly been used in the study of travel choices yet they have a much broader appeal in studying the decision making process of organisations. In this paper we use a data-mixing model to capture the decision expertise of a transport organisation through the revelation of preference weights for each evaluation criterion. Using choice information based on both market-driven and experimentally-derived choice sets, we should be able to increase the robustness of the evaluation weights in comparison to the weights obtained from single data-sourced models. The resulting parameterised tool can be used in subsequent tender evaluations to provide an additional source of advise to supplement or replace that provided by current members of a bid evaluation team.

AUTHORS: David A. Hensher,
Jordan J. Louviere [*], &
David E. Hansen [**].

CONTACT: Institute of Transport Studies (Sydney & Monash)
The Australian Key Centre in Transport Management
C37, The University of Sydney NSW 2006
Australia

Telephone: +61 2 9351 0071
Facsimile: +61 2 9351 0088
E-mail: itsinfo@its.usyd.edu.au
Internet: http://www.its.usyd.edu.au

DATE: March, 2000

[*] *Department of Marketing, Faculty of Economics & Business, University of Sydney.*

[**] *Department of Marketing, Texas Southern University.*

Introduction

Previous research on modelling organisational decision making has lead to advances in expert and knowledge-based systems that rely on models, managerial intuition, and internal databases. Yet, despite significant advances in the accuracy and consistency of organisational decision models, it remains unclear how well they capture managerial expertise, which is the key component of "expert" systems. Indeed, capturing managerial expertise requires one to model human decision-making, which encompasses a wide array of methods that often yield conflicting empirical results (Hogarth 1987). For example, some researchers suggest that one should model actual decisions using historical data and/or so-called "representative experiments" (Hammond et al, 1986), while others rely on well-designed statistical experiments (eg, orthogonal arrays) because they enhance parameter estimation, model comparison and hypothesis tests (Louviere, Hensher and Swait in press 2000). More recently, a number of researchers, particularly in transport, have suggested that actual and experimental decisions and choices are complimentary and can be combined (Ben-Akiva and Morikawa 1990; Swait and Louviere 1993). This stream of research also suggests that it may be possible to develop better models by combining different sources of data to take advantage of the strengths of each, while minimising their weaknesses. The purpose of this paper is to investigate the possibility of improving our ability to model organisational decisions by combining data from hypothetical experiments with historical choice data to model organisational choices of bidders in a competitive tendering situation.

The latter application is of practical interest and importance because it is an activity routinely undertaken by many organisations worldwide, and to the extent that one can move in the direction of providing expert or knowledge-based systems that can perform much of the preliminary work, considerable savings in time and resources could be realised. This paper represents a very modest step in this direction in the form of a feasibility study that investigates the applicability of pooling sources of data to develop an organisational choice model. Future research will deal with generalising the approach to other, similar applications and determining the degree to which such models can assist organisations in early stages of the process by screening proposals and recommending a few for final consideration. Such an application is not unprecedented, as Dawes (1979) demonstrated that one could use simple statistical models to screen applicants to graduate programs, and that one could specify the models in such a way that they never rejected an applicant who actually would be selected.

There is evidence to suggest that pooling data from different sources can be an advantage in modelling decisions. For example, combining separate forecasts seems to improve the accuracy of forecasts, which in turn has led to methods for combining models and/or experts (Makridakis and Winkler 1983; Granger and Ramanathan 1984; Armstrong 1999; Blattberg and Hoch 1990). Such methods may do well because unique information from each data source reduces forecast errors much like stock portfolio diversification reduces risk. Thus, we propose that historical data-based and experimental data-based ways to model decisions also may exhibit offsetting properties that reduce forecast errors. We focus on organisational buying decisions in which one winner takes all, such as vendor selection problems that can be viewed as a discrete choice of one vendor. This view allows us to use well-developed theory and methods from random utility theory (RUT) to develop and test discrete choice models (Louviere and Woodworth 1983; Louviere, Hensher and Swait in press 2000). Although we focus our research on this seemingly narrow application, we think that our proposed approach can be generalised to many other discrete choices made by transport organisations, such as choosing from pools of job applicants, evaluating and choosing suppliers, and so forth.

Combining Market and Experimental Choice Data

Our review of the research literature on combining forecasts to improve predictions suggests that the degree of improvement is related to the degree to which each separate forecast contributes uniquely to an overall forecast with reduced forecast errors, as well as the way in which forecasts are combined. Past research initially focused on estimating preferences from a single source of information, namely managers' past decisions (Dawes 1979), but attention shifted to ways of combining different types of information for prediction. Examples of combining information sources can be found in the decision calculus literature (eg, Little 1970), but more recent methods try to enhance forecasts by combining databases with managerial judgment (Blattberg and Hoch 1990). Another related stream of literature in marketing deals with so-called "hybrid conjoint" analysis methods that combine individual's estimates of model parameters (termed "self-explication") with conjoint experiments to estimate models (see Cattin, Gelfand, and Danes 1984).

Different paradigms in the judgment and decision making literature have tended to focus on historical or experimental data as a primary basis for studying and modelling decisions. For example, the Social Judgment Theory (SJT) paradigm focuses on historical decisions, but also makes use of experiments that reflect real decision environments. Such experiments are called "representative" because they deliberately incorporate decision attribute correlations that are observed in actual or "representative" decision environments (see Hammond et al. 1986); hence they are thought to be more ecologically valid and generalisable. SJT's roots lie in Egon Brunswick's (1952) conceptual framework that views human decisions as a mapping between a human and her decision environment. The decision environment is represented by a relevant set of fallible cues or attributes that humans use to interpret the environment and on which they base their decisions. Because the SJT approach to modelling decisions often uses actual decision or choice data as a basis for estimating statistical models that describe decision processes, it has much in common with "revealed preference" modelling approaches used in transport research. Unfortunately, data on real choices rarely are ideal for the estimation of statistical models of choice or decision processes because explanatory variables typically exhibit limited variability (sometimes constant) and high collinearity and/or new variables or options are not observed in the data history (Louviere 1988; Louviere, Hensher and Swait in press 2000). However, the external validity of historical data are obvious.

Other paradigms are conceptually similar to SJT, but instead rely on statistical design methods to construct hypothetical scenarios that are evaluated by human subjects who make decisions with reference to each. Subjects' decisions reveal their preferences in much the same way that observed choices can reveal preferences if certain statistical conditions are satisfied. Unlike "representative" experiments, these approaches to studying and modelling decision making rely on variations of factorial experiments in which attributes are uncorrelated. Information Integration Theory (IIT) pioneered by Norman Anderson provides an excellent example of this approach (see Anderson 1981, Louviere 1988). Conceptually similar approaches to IIT based on RUT have evolved in transport and marketing in which statistical design techniques are used to create sets of competing choice options from which subjects choose. Statistical design techniques are used to insure that certain classes of RUT-based choice models can be estimated from the subjects' choices (see, eg, Louviere and Woodworth 1983), which insures satisfaction of statistical properties of choice models at the expense of market or historical validity.

Research into pooling sources of preference or choice data in the RUT paradigm also suggests that one should be able to improve models by combining historical (market) and experimental choice data due to the unique and complementary information each provides. Moreover, RUT provides a common theoretical framework to combine, analyse, model and compare many types of preference data (McFadden 1981, Louviere, Hensher and Swait in press 2000). Thus, historical and experimental data can be viewed *theoretically* as complements, and research (Ben-Akiva and Morikawa 1989) suggests that there can be considerable benefit from pooling them and exploiting the strengths of each while minimising their weaknesses. Moreover, as we later discuss, research has demonstrated that choices made in real and hypothetical decision environments not only are related, but often yield similar information about preferences.

To our knowledge, pooling of market and designed choice experiment data has not been proposed or applied to organisational decision making. However, the above discussion suggests that such an approach should be potentially useful. For example, in choice contexts driven by market conditions, such as the industrial bidding situation considered below in which bidders seek to satisfy published requirements of transport projects, positive correlations among attributes of competing bids may well occur and/or bids may differ only on a few attributes. In such cases designed orthogonal arrays may yield better conditioned data for estimating the effects of the attributes on decisions (see McClelland and Judd 1993).

The remainder of the paper is organised as follows. In the next section we propose a formal conceptual model of organisational decision making based on RUT; then we introduce and explain the application of the conceptual framework to an organisational decision making problem involving choice of bidder in transport projects. This is followed by the results and their policy interpretation; and the paper concludes with suggestions for future research directions.

Quantifying Organisational Choice

We assume that we can capture organisational expertise by modelling contract award decisions as discrete choices, given that only one supplier wins. That is, we model the probability that a transport organisation will award a contract to the i^{th} supplier and view this as a random utility process in which an organisation seeks to award a contract to that bidder who offers the highest utility. As outsiders, we cannot know the organisation's true utility function, hence

$$U_i = V_i + \varepsilon_i \tag{1}$$

where

U_i = the latent, unobservable utility of the i^{th} bidder
V_i = the explainable component of utility
ε_i = the stochastic, unexplainable component of utility

Due to the stochastic component, we can only establish up to a probability that the organisation chooses the i^{th} bidder, hence

$$P(i|C) = P[(V_i + \varepsilon_i) > \text{Max } (V_j + \varepsilon_j)],\ \forall\ j \in C,\ i \neq j \tag{2}$$

where C = the set of competing bids.

For now we assume that ε_i is an independently and identically distributed (IID) Gumbel $(0, \sigma_\varepsilon^2)$ random variate for all i=1,…,I (we relax this assumption later). This assumption leads to the well-known multinomial logit (MNL) choice model, which can be expressed as follows:

$$P(i|C) = \exp(\lambda V_i) / \sum (n=1 \text{ to } j)\ \exp(\lambda V_j) \tag{3}$$

and

$$V_i = \beta_i X_i. \tag{4}$$

λ is a scale constant, unique to a particular data set or model ($\lambda^2 = \pi^2/6\sigma_\varepsilon^2$ in the case of the Gumbel), and β_i is a K-element parameter vector associated with K-1 project evaluation scores and price (X_i).

In our application β_i is generic (ie, the effects of project attributes and price are constant for all bidders) because bidders are unknown (at least to the analyst) and vary across projects. This assumption implies that an organisation evaluates all bidders in the same way, and chooses the bidder with the highest overall utility score. When ones combines market and experimental data there is a risk that the IID error assumptions will be violated because these distribution of the error components in each data source may not be identical. To allow for this possibility, we express the choice process as a nested choice problem (see Ben-Akiva and Lerman 1985, Hensher and Greene in press), which allows us to handle potential IID violations by grouping the market and experimental data into different branches of the nest during estimation.

Several researchers have shown that despite their differences, historical and choice experiment data can be combined successfully in the branches of a nested logit model (Ben-Akiva and Morikawa 1989; Hensher and Bradley 1993). This structural specification has the advantage of recognising that the primary difference between utility expressions for each data source lies in the distribution profile of the error components. Nested logit (NL) models provide a way to partition choice sets such that subsets of alternatives can have different unobserved (ie random) component variances. Thus, partitioning the pooled data into two sets, namely historical and experimental data, takes variance differences into account and provides a test of the degree to which one needs to rescale the parameter estimates of one data set by the ratio of the variances of the unobserved effects from the two data sets. That is, in RUT choice models the variance of the random component of utility is inversely proportional to the scale of the utility parameters. Hence, if we denote the scale of the parameters as λ, we can normalise the variance of one data set to unity and estimate the ratio of the error component variances in both data sources relative to that restriction, which provides a test of the econometric representation versus naïve pooling.

An Application to Organisational Purchasing through Competitive Bidding

Obtaining and evaluating bids

To test the pooled data model and compare it to stand-alone historical (market) and experimental choice data models for understanding organisation choices we model the selection decisions of a large transport organisation for out-sourced professional engineering services like project planning, design, and project supervision. Typically, the organisation formulates a request for proposals (RFP) that contains a description of the project to be completed, and receives two to six proposals from prospective bidders in response. Each proposal consists of a fixed contract price, and project attribute information that includes details of project method, team capabilities, track record, and expected outcomes (see Domberger at al 1995, Berechman 1993).

The organisation evaluated the bids in two stages: 1) Prior to seeing the prices of each proposal, decision makers evaluated proposal quality by rating each proposal on a 1 to 100 rating scale with respect to each attribute (ie, experience of staff relevant to this type of project, firm's track record on similar projects, relevant technical skills, project methodology, and project management skills). Decision-makers then calculate a quality index for each proposal by subjectively weighting and combining all five attributes. Weights are a function of decision maker past experience and unique requirements of a particular project. 2) Then the quality index is compared against each bid price, dominated alternatives are eliminated and the proposal with the highest value is awarded the project (assuming that value to be a function of quality and price). We try to capture this price-quality tradeoff in the estimated model parameters because it represents the expertise of the individual decision-makers, which implicitly is the organisation's policy or decision rule.

Historical (Market) choice information

The organisation's field offices supplied data on actual project award decisions made over a two year period for a variety of transport projects involving engineering design, planning or investigation. We received data from 95 recent decisions that contained ratings of non-price attributes, bid prices for each proposal and final project winners in each set of proposals. Projects varied in levels of complexity, which might impact price-quality tradeoffs: they spanned a wide range of prices ($20,000 to $600,000) and attribute quality ratings (45 to 100 on a scale of 1 to 100). We used the median price of each set of bids to divide the 95 sets of bids into two groups: one had a median price of about $70K and the other had a median price of about $300K. We made the assumption that complexity and price were related, which provided us with low (n=54) and high complexity groups of projects (n=41). The dependent variable or decision outcome was the choice of bidder. Two to six bids were received for every project. Thus, these bids constitute the choice sets faced by the organisational decision makers. We treated the winning bidder as the chosen option and the losing bidders were rejected options.

Experimental choice information

The design of a suitable choice experiment in this application was complicated by virtue of the fact that all attributes are quantitative[1] and the preference directionality of each is known a priori (e.g., as price increases, utility decreases, all else equal). Thus, we could not simply design the experiment using orthogonal fractions of complete factorial experiments (common practice in stated preference modelling applications) because these orthogonal design produce choice sets in which there are dominated choice alternatives. Choice sets that contain dominant/dominated choice options may compromise the credibility of the experimental exercise and/or provide less useful statistical information because subjects always choose the dominant option. We dealt with the dominance issue by creating a non-dominant design in the following way:

1. We first generated attribute combinations to describe different bidder profiles based on an orthogonal, main effects fraction of the 4^6 factorial. In particular, we assigned four numerical levels to each quality attribute and bid price to span the range of variation observed in the market data (65, 72, 79, 86 for the quality attributes, and median bid price plus zero, 10%, 20%, or minus 10%) and used a main effects design to create 64 profiles (i.e., a 4^{6-3} design). The levels of quality ratings did not span the 0 to 100 range of the rating scales because we observed very few ratings above 85 or below 70; hence we restricted the range of levels to that observed in the overwhelming majority of the actual decisions for which we had data.

2. Wiley (1978) demonstrated that the sum of the design codes for each attribute combination (profile) can be used to identify sets of non-dominant choice options. That is, bidder profiles can be sorted and arranged into sets based on their design code sums. For example, a profile with design code = 1,3,3,2,1,4 (representing the five quality dimensions and bid price) sums to 14; hence it can be combined with any other profiles whose sum also equals 14 without any of these profiles dominating or being dominated. This allowed us to put the 64 profiles into groups of profiles that had equal code sums. We then constructed as many non-dominant pairs of profiles as possible from the members of each group of profiles whose code sums were equal. This yielded 37 pairs and three singles. We then added the option of not awarding a contract to each pair and single, which produced 40 total choice sets.

3. Finally, because the range of prices was so large, we developed two tasks to allow for differences in project complexity reflected in the price range. The tasks differed according to whether the price range was above or below the median bid observed in the actual project award data supplied to us. For example, the median bid price in the low price/complexity range was

[1] It is well known that if all attributes in a design are quantitative, as distinct from qualitative, there will always be combinations of attributes that are dominated by other combinations or that dominate other combinations. Including a qualitative attribute (eg colour) whose utility direction on the scale is 'ambiguous' solves this issue.

$70,000, and the median bid price in the high price/ complexity range was $300,000. The two tasks differed only in the bid prices that were assigned to the four price levels; the design was the same for all other attributes in both tasks.

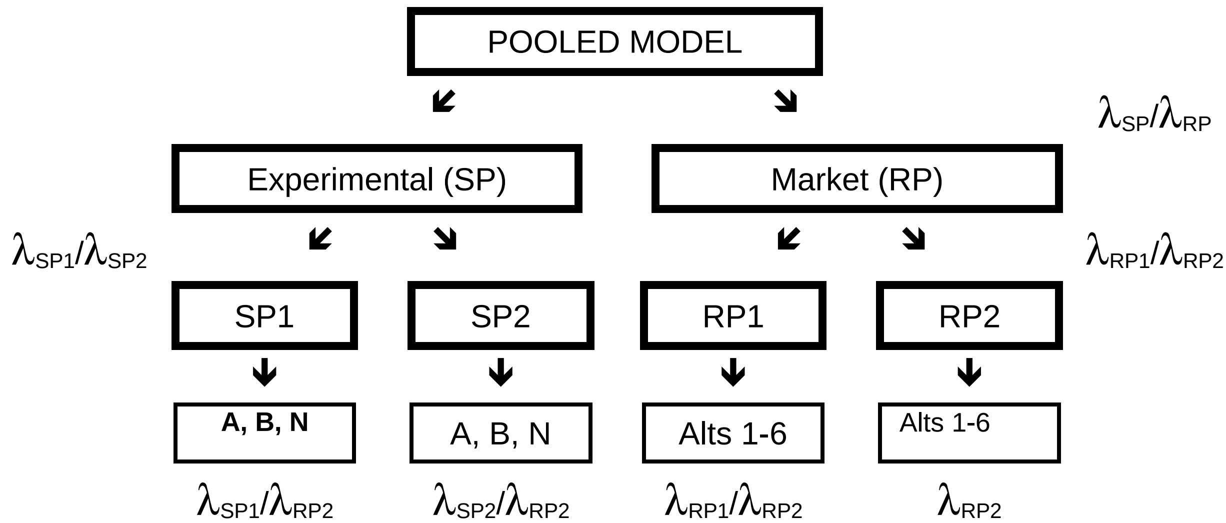
Twenty employees participated in the experiment and completed both tasks (high/low price-complexity). In each of these 80 choice sets (40 low and 40 high complexity) they awarded the contract to one bidder or they awarded no contract. They were not allowed to use mechanical devices to help with the decisions and were told to assume that the attribute+price profiles described typical proposal scores for professional services projects. Just as in a real project award situation, we required the 20 employees to *reach a consensus regarding the winner in each choice set*, and the overall winner defined the chosen alternative in each choice set. The number of choice sets may seem large to transport researchers familiar with SP practice in transport applications, however, in many SP paradigms it is not uncommon for researchers to ask subjects to evaluate 100, 200 or even larger numbers of profiles or choice sets (see, Norman and Louviere 1974; Louviere 1974; Louviere, Oppewal, Timmermans and Thomas 1993). Indeed, as noted by Louviere, Oppewal, Timmermans and Thomas, there is surprisingly little evidence to suggest that such numbers of profiles or choice sets effect mean model parameters. Rather, as noted by Louviere and Street (1999) and Louviere and Hensher (2000), available evidence suggests that larger number of choice sets most likely impact error component variability, not mean model parameters. Thus, larger numbers of choice sets may increase error component or response variability, which is a reliability and not a validity issue.

Model estimation

The two sets of choice data were pooled and models were estimated using the approach pioneered by Ben-Akiva and Morikawa (1989), and subsequently extended by others (Hensher and Bradley 1993; Swait and Louviere 1993). That is, we pooled the actual and experimental data, and tested if the model parameters in each data source were proportional by applying the nested logit estimation approach of Hensher and Bradley (1993). The market and experimental data were treated as branches of a nested logit model. The inclusive value parameter in the nested logit model is an estimate of the error variance ratio of market to experimental data. Specifically, we let the "award no contract" option in the choice experiment have a unique error variance by treating it as a separate nest at the lowest level of the tree. Then we estimated separate error variance ratios for each experimental data source (award vs no award) and the market data. We found that the best model constrained the error variance ratios of the market alternatives and the no-choice alternative to unity. It also is important to note that the decisions being modelled are consensus or group choices; hence heterogeneity is irrelevant. Thus, we estimated one set of parameters for all observations.

The process of combining data from sets of choice settings is known as “rescaling,” which Swait and Louviere (1993) define as the process by which error variance ratios for two or more sources of preference data can be placed on the same “scale” by multiplying the target data source parameters by the estimated error variance ratio. Figure 1 shows that this involves estimation of $\sigma^2_{SP}/\sigma^2_{RP}$, or, $\sigma^2_{SP1}/\sigma^2_{RP1} + \sigma^2_{SP2}/\sigma^2_{RP2}$, where SP1 and RP1 represent the low complexity projects, and SP2 and RP2 represent the high complexity projects.

Figure 1 Nested Logit Model with Scaling Factors for Each Branch



Notes: SP1, RP1 refers to low complexity projects, and SP2, RP2 refers to high complexity projects, A, B are high and low complexity bids, N = non-award, Alts1-6 = six bids in each observed choice set.

Summary of estimation results

The results in Table 1 are for the pooled RP-SP model. We arrived at this model as follows:

1. We effects coded the four levels of each attribute and price in the SP data set, and estimated a main-effects only (ie, strictly additive) model specification that incorporated only the effects-coded quality attribute and price main effects. Effects codes are exactly like dummy codes, except that the omitted value or level is coded uniformly –1 instead of uniformly 0. This centers the estimation about the intercept, allows the intercept to be interpreted and permits one to create uncorrelated cross-products. When one uses 0, 1 dummy codes cross-products (interactions) are not orthogonal and model intercepts typically are uninterpretable.

2. We graphed the resulting estimates of the attribute and price effects, and found that three attributes (relevant experience, track record and management skills) not only seemed to be non-significant, but their results were not interpretable. Taken together, this finding suggests that these attributes are not significant, there are unobserved interactions or both. The remaining three attributes (price, technical skills and methodological expertise) appeared to exhibit diminishing marginal utility as their levels increased.

3. We eliminated the three insignificant attributes[2] and estimated a second model in which we transformed the three remaining variables as follows: a) we took the

[2] In both the SP and RP data models, the expected directionality of the marginal utility effects were as expected for all attributes, except for relevant experience. We cannot determine whether this sign reversal reflects a real aspect of these decisions, namely that the utility of a bid decreases at an increasing rate with increases scores on this dimension. Ordinarily, such a reversal in RP data would be due to multi-

log of each attribute level and then centered the log values about their respective means (ie, we subtracted the geometric mean from each log-transformed level); b) then we estimated a model that included all main effects and all interaction effects of the log-transformed and mean-centered attributes (price, technical skills and methodological expertise). We report that model in our results below because it was significantly superior statistically to a main effects only model.

Mean-centering attributes about their means eliminates or significantly reduces collinearity that results from simply multiplying each attribute column to create cross-products or interactions. Indeed, if attribute levels are evenly spaced, mean-centering exactly orthogonalises interactions with respect to each other and main effects, design permitting (see Louviere, Hensher and Swait in press 2000). Separate results for RP (market) and SP (experiment) models with the reduced set of attributes are presented in Table 1. They reveal that both RP and SP models have correct signs and significant interactions, but the sign of the MP interaction differs between the data sources. The coefficient for the "None" option in the SP data model is significant, suggesting that decision makers rejected all available options when necessary, possibly because neither option was worth the bid price. The latter leads to increased price sensitivity, and in turn a larger absolute value of the price coefficient given that price is generic across all observations.

Table 1: Separate RP and SP Models

Table 1a: RP model

Variable	Coeff	t-stat
PRICESM	-3.924	-38.604
TECHSKM	15.009	32.291
METHODM	7.459	24.398
TM	-27.175	-6.520
TP	4.283	7.106
MP	-8.303	-21.277
TMP	-65.339	-12.045

Log-Likelihood at solution = -5685.297
Adjusted rho-squared = 0.578

Table 1b: SP Model

Variable	Coeff	t-stat
SP12	-3.112	-19.358
PRICESM	-7.319	-35.712
TECHSKM	4.980	21.224
METHODM	3.755	16.311
TM	-21.557	-10.086
TP	0.033	0.105
MP	0.658	2.096
TMP	-10.197	-3.495
IND30KN	-10.179	-32.801

Log-Likelihood at solution = -5170.9345
Adjusted rho-squared = 0.375

The pooled model is presented in Table 2. All coefficients have the expected signs, and the main effects are significant well beyond conventional confidence interval levels (eg,

collinearity among the independent variables, but this cannot explain the same reversal in the SP data. Thus, a strong possibility is that this effect is confounded with an unobserved interaction that cannot be estimated.

.05). The interaction between technical and methodological skills is highly significant; the methodological skills by price interaction and the three-way interaction approach significance, which is encouraging given the small sample size. For example, McClelland and Judd (1993) demonstrate that one has very little power to detect non-linear and interaction effects with small samples, a problem that is exacerbated with discrete response data. The pooled model fits the data very well by conventional choice modelling standards, as indicated by a very high McFadden's rho-square (.86); and is much higher than the pooled MNL model rho-square in Table 2a. The models being compared are non-nested, so we rely on goodness-of-fit and prediction success measures; both strongly favour the pooled nested logit model in Table 2b. The inclusive value estimates in Table 2b used to combine the RP and SP data sources are significant and greater than one, indicating that the RP data variance was larger than the SP data variance[3] Interestingly, the dummy indicator variable for high versus low complexity IND30KN) projects was significant, suggesting differences in the overall likelihood of choosing a project (relative to not choosing) depending on value. That is, subjects were less likely to choose high complexity-high cost projects relative to choosing not to award the project to any of the bidders.

Table 2 Pooled Model

Table 2a: MNL Model Estimated as Starting Solution

Variable	**Coeff**	**t-stat**
SP	-0.595	-6.359
PRICESM	-3.912	-47.668
TECHSKM	8.188	41.132
METHODM	5.437	31.963
TM	-27.581	-15.772
TP	-0.066	-0.242
MP	-3.062	-13.425
TMP	-22.941	-9.625
IND30KN	-5.210	-34.185

Log-Likelihood at solution = -11528.87
Adjusted rho-squared = 0.675

Table 2b: Pooled, Nested Logit model

Variable	**Coeff**	**t_stat**
SP	-3.233	-14.499
PRICESM	-7.107	-35.401
TECHSKM	4.694	19.235
METHODM	3.765	16.488
TM	-9.762	-4.363
TP	-0.187	-0.617
MP	0.462	1.518
TMP	-4.596	-1.675
IND30KN	-12.61	-12.276
	IV parameters	
SP	1.260	13.526
NONE	1.000	(Fixed)
RP	1.000	(Fixed)

Log-Likelihood at solution = -4831.100
Adjusted rho-squared = 0.864

Note: SP = stated preference specific constant (1,0), PRICESM = natural log of price mean centred, TECHSKM = natural log of technical skills mean centred, METHODM = natural log of methodological

[3] The scale parameter in nested logit (inverse of the inclusive value parameter) has been shown by a mixed logit model to account for the major source of preference heterogeneity. This is due we suspect to the fact that the SP data is derived from a group response and hence no preference heterogeneity, but the RP data is a sample of individuals. Thus any preference heterogeneity appears to be well handled by the nested logit form.

skills mean centred, Two-way interactions of natural logs of technical skills (T), methodological skills (M) and price (P), TMP = three-way interaction of T, M and P; IND30KN = dummy variable distinguishing projects of low complexity $70K (0) and of high complexity $300K (1).

Forecasting effectiveness

The predictive ability of each model was evaluated by simulating the choices observed in each data source to determine how well the model predicted the organisation's choices. A useful indicator of predictive accuracy is a simple summation of the number of times that an alternative receives the highest choice scores. The joint model predicts the actual winning project bidder correctly 67.9% of the time in the RP sample, and the application of the SP parameter estimates to the RP data predicts correctly 62.8% of the time, supporting the gains from the joint SP-RP model. Coincidentally the stand-alone RP model also predicts 67.9% of the RP choices correctly. In contrast, the stand-alone SP model correctly predicts 93.75% of the SP experiment choices; hence the SP data has very good reproduction credentials (albeit for the experimental choice responses). Because the ability to predict the actual market decisions is the most relevant criterion, we conclude that the joint model and the RP stand alone model both reproduce the winning bidders correctly nearly 70% of the time.

Policy Implications for Organisational Decision Making

The use of SP and RP data to identify the relative importance that a bid evaluation team places on the criteria used to rank bids and select the preferred offer is a novel but powerful application of a method now well established in the study of travel choices. Similar to the way in which one applies a set of parameterised utility expressions to predict travel mode shares, one also can use the parameterised group utility expression to prioritise and even short-list bidders in a multi-stage evaluation process.

The resulting evaluation formula derived from a group utility maximising paradigm can play two roles: 1) It complements the assessments of a given team of bid evaluators, adding what might be described as a broader set of representative weights for all attributes and extra evaluative input for a specific set of bids. The latter functionality can be viewed as a quality-control mechanism that ensures that a bid evaluation team maintains a degree of consistency with evaluations provided by previous evaluators in similar contexts. 2) The evaluation formula also can provide an alternative, stand-alone method for evaluating and screening all bids that can be used to screen out or eliminate bids that differ significantly from previous representative assessment.

Although the set of statistically significant attributes was smaller than the available set that could have been used to evaluate transport projects in this application, this is less relevant than demonstrating how our approach can add value by establishing relative weights for the evaluation criteria. On the other hand, the fact that evaluators did not use/pay attention to some of the evaluative criteria suggests that these criteria may not be as relevant to project proposal evaluation and the ranking and selection of suppliers as the organisation has assumed. That is, our results suggest that if decisions are based only on three attributes, as our empirical results imply, the transport agency probably should reappraise the set of evaluative criteria it uses. Thus, at a minimum, the approach we proposed and implemented provides a way to ascertain the suitability of specific evaluation criteria in a systematic and scientifically valid manner.

Conclusions

This paper introduced the idea of mixing data sources to reveal the preferences of a team of individuals who evaluate competitive bids for the delivery of professional services to a large organisation. The use of experimental choice data to enrich revealed preference data provides a way to establish a set of weights to represent the relative importance of each evaluation criterion that enhances the statistical precision of weights derived from stand-alone market data. These advantages arise from the fact that such choice experiments allow one to observe larger numbers of service offerings and choice responses than usually are available in real markets. Thus, in an application like this, experimental data helps to deliver more statistically precise weights that enhance our understanding of real market decisions. The empirical application in this paper supports this conclusion in terms of overall statistical model fits and parameter standard errors, but the predictive accuracy of our approach was equal to a RP stand alone model. The latter result is somewhat surprising because very few similar outcomes have been reported in previous research; hence its generality should be assessed in future research.

Finally, it is worth noting that the approach we proposed and applied in this paper is not intended to fully replace human decision makers. Rather, to the extent that our results can be generalised to demonstrate comparable levels of predictive ability, then the models can be used as decision aids and sources of additional, quality-control relevant information. Much as the increasingly ubiquitous “doctor in a box” expert and knowledge-based systems assist general practitioners to more accurately and quickly diagnose symptoms presented by patients, so also can choice-model based expert and knowledge systems assist individuals and organisation faced with demanding evaluation tasks that require considerable expenditure of organisational time and resources.

References

Anderson N.H. (1981) *Foundations of Information Integration Theory,* Academic Press, New York.

Armstrong, J. (1999), Bootstrapping Experts to Improve Forecasts, *working paper*, Wharton School of Business, Philadelphia.

Ben-Akiva, M. E., and Morikawa, T. (1989). Estimation of Switching Models from Revealed Preferences and Stated Intentions, *Transportation Research*, 24A(6), 485-495.

Berechman, J. (1993) *Public Transit Economics and Deregulation Policy*, Studies in Regional Science and Urban Economics Vol. 23, North-Holland, Amsterdam.

Blattberg, R. C. and Hoch, S. J. (1990). Database Models and Managerial Intuition: 50% Model + 50% Manager, *Management Science*, 36(8), 887-899.

Brunswick, E. (1952). The Conceptual Framework of Psychology, in International Encyclopedia of Unified Science, 1 (10), University of Chicago Press, Chicago

Cattin, P., Gelfand, A. and Danes, J. (1984), A Simple Bayesian Procedure for Estimation in a Conjoint Model, *Journal of Marketing Research*, 20, 29-38.

Dawes, R. M. (1979), The Robust Beauty of Improper Linear Models in Decision Making, *American Psychologist*, 34 (July), 571-582.

Domberger, S., Hall, C., Ah Lik Li, E. (1995) The Determinants of Price and Quality in Competitively Tendered Contracts, *The Economic Journal*, 105, 1454-70.

Granger, C. W. J. and Ramanathan, R. (1984). Improved Methods of Combining Forecasts, *Journal of Forecasting*, 3, 197-204.

Hammond, K. R., Hamm, R. M., and Grassia, J. (1986). Generalizing Over Conditions By Combining The Multi-Trait Multi-Method Matrix And The Representative Design Of Experiments, *Psychological Bulletin*, 100, 259-269.

Hensher, D.A. and Bradley, M. (1993) Using Stated Response Data to Enrich Revealed Preference Discrete Choice Models, *Marketing Letters*, 4(2), 139-152.

Hensher, D.A. and Greene, W.H. (in press) Specification and Estimation of Nested Logit Models: Alternative Normalisations, Transportation Research B.

Hogarth, R. (1987). *Judgment and Choice*. Seventh Edition, New York: Wiley.

Little, J. D. C. (1970). Models and Managers: The Concept of a Decision Calculus, *Management Science*, 16(8), B466-B485.

Louviere J. (1974) Predicting the Evaluation of Real Stimulus Objects From an Abstract Evaluation of Their Attributes: The Case of Trout Streams, *Journal of Applied*

Psychology, 59(5),572-577.

Louviere J. (1988) *Analyzing Decision Making: Metric Conjoint Analysis*, Newbury Park, CA:Sage Publications, Inc.

Louviere, J.J. and Hensher, D.A. (2000) Combining sources of preference data, A Resource Paper for IATBR 2000, *9th International Association for Travel Behaviour Research Conference*, Gold Coast, Queensland, Australia, 2-7 July, 2000.

Louviere, J.J. and Street, D. (2000) Stated preference methods, in Hensher, D.A. and Button, K. (eds) *Transport Modelling*, Handbooks in Transport, Volume 1, Pergamon Press, Oxford, Chapter 8.

Louviere, J.J. and Woodworth G. (1983) Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data, *Journal of Marketing Research*, 20, 350-36.

Louviere, J.J., Hensher, D.A. and Swait, J. (in press 2000) *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*, Cambridge University Press, Cambridge.

Louviere, J., Oppewal, H., Timmermans, H. and Thomas, T. (1993), Handling large numbers of attributes in conjoint applications: who says existing techniques can't be applied? But if you want an alternative, how about hierarchical choice experiments? Paper presented to the *American Marketing Association's Advanced Research Techniques Forum*, Beaver Creek, CO (June).

Makridakis, S. and Winkler, R. (1983). Averages of Forecasts: Some Empirical Results, *Management Science*, 29, 987-996.

McClelland, G. and Judd, C. (1993), Statistical difficulties of detecting interactions and moderator effects, *Psychological Bulletin*. 114 (2), 376-390.

McFadden, D. (1981) Econometric models of probabilistic choice in Manski, C.F. and McFadden, D. (eds.) *Structural Analysis of Discrete Data*, MIT Press, Cambridge, 198-272.

Norman K. L., and Louviere J. (1974) Integration of attributes in public bus transportation: Two modelling approaches, *Journal of Applied Psychology*, 59 (6), 753-758.

Swait, J. and Louviere, J. J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models, *Journal of Marketing Research*, 30(3), 305-314.

Wiley, J.B. (1978) Selecting Pareto Optimal Subsets from Multi-Attribute Alternatives, in K. Hunt, (ed.), *Advances in Consumer Research*, V,171-74.