ON ASSESSING THE PREDICTIVE VALIDITY OF
FAST AND FRUGAL HEURISTICS

by

Y. Ganzach
M. Givon

Working Paper No 8/2010 October 2010

Research No. 01680100

Corresponding author: Prof. Yoav Ganzach, Faculty of Management, The Leon Recanati Graduate School of Business Administration, Tel Aviv University, Tel Aviv, Israel.
Email: yoavgn@post.tau.ac.il, Telephone: 972-3-640-5892.

This paper was partially financed by the Henry Crown Institute of Business Research in Israel.

Abstract

We show that when irrelevant cues are included, the fit of a model based on a fast and frugal heuristic called the matching heuristic can exceed the fit of two compensatory models, the Franklin rule and regression, even when the true generating model is compensatory. We discuss the issue of cue relevance and emphasize the need to keep the level of cue selection constant in comparative model fit studies.

Comparative model fit is an approach to theory testing that assesses the validity of competing theories by comparing their statistical fit to the data. This approach has been used in many areas of the behavioral sciences in general (e.g., Cohen, Dunbar & McClelland, 1990) and in behavioral decision making in particular (e.g., Goldberg, 1970; Ganzach, 1995). Recently it has been used to test the validity of 'fast and frugal' heuristics, by comparing the fit of models based on the assumption that people rely on such heuristics in making decisions to the validity of models based on the assumption that people make compensatory decisions (e.g., Dhami & Ayton, 2001; Dhami, 2003; Gigerenzer & Goldstein, 1996).

Whereas the current paper offers a critique of some comparative model fit procedures used in the study of fast and frugal heuristics, it by no means suggests that people do not rely on such heuristics. Furthermore, in contrast to some authors (e.g., Roberts & Pashler, 2000), we do not argue against the practice of assessing theory validity based on model fit. However, we do suggest that more care should be exercised in using comparative fit procedures in confirming hypotheses about decision heuristics, and we explore some possible pitfalls of these procedures. To this end we critically examine a comparative model fit procedure used by Dhami (2003), and show that her procedure can often lead to the conclusion that people use a fast and frugal strategy even if the true strategy is compensatory. Note, however, that although our critique is built around Dhami's (2003) work, it is relevant to other work in the fast and frugal tradition – primarily work that relies on cross-validation – to assess model fit (e.g., Gigerenzer, Czerlinski & Martignon, 1998; Martignon & Blackmond Laskey, 1999; Martignon & Hoffrage, 2002).

The key concept in our critique is the concept of cue relevance. It refers to the researcher's decision as to which cues should be used in the model. Since there is a large number of cues that the researcher may consider in building her model, a crucial decision is which cues should be selected. In comparative model fit research there are either one or two stages in the selection. In the first of these two stages – a stage which is often not explicitly stated – the researcher chooses a limited number of cues out of the many possible cues. Thus Dhami (2003) chose 25 cues that may affect a judge's bailing decision, and Gigerenzer & Goldstein (1996) used nine cues that may be related to the sizes of German cities. But obviously it is possible to select more or less cues. This somewhat arbitrary initial selection of cues lies at the heart of the biased comparison between compensatory rules and fast and frugal heuristics discussed in this paper.

Some of the work in comparative model fit in the fast and frugal tradition stops at this first stage (see for example, Gigerenzer & Goldstein, 1996; Dhami & Ayton, 2001). That is, after the initial selection of cues, the fits of the competing models are estimated on the basis of this set of cues. The current paper does *not* directly concern this type of comparative model fit studies, though some of our results are also relevant to it.

The second type of comparative model fit studies, the type which is the focus of the current paper, involves a second stage, a cross-validation stage. The database is divided into two samples, a modeling sample and a validation sample. Based on the first stage, performed only on the modeling sample, the weights of the cues are determined. Using these weights, the researcher builds a validation model which is used to predict the criterion (e.g., participants' decisions) in the validation sample. The correlation (or other measures of match) between the predictions of the validation model and the criterion is taken as a measure of the predictive fit of the model, and consequently of the validity of the corresponding cognitive process underlying the model.

The fit of a model in the cross-validation sample (i.e., the predictive fit) critically depends on the number of cues which were considered in the modeling stage and on the procedure for selecting cues for the validation model. Particularly important for the current issue is a comparison between the following two possible procedures: (1) many cues are considered in the modeling stage followed by a careful selection of cues for the cross-validation stage – only cues that were found to be valid in the modeling stage are used in the validation model; (2) many cues are considered in the modeling stage but no selection is performed. The two procedures will lead to different results: the predictive fit of the second procedure will be lower than the predictive fit of the first since arbitrary weights assigned to invalid cues in the modeling stage will 'ruin' the fit in the cross-validation stage.

Dhami (2003) used the first method in estimating the predictive fit of a model derived from a fast and frugal decision rule called the matching heuristic – a variant of Take the Best heuristic (Gigerenzer, Hoffrage & Kleinboalting, 1991; Gigerenzer & Goldstein, 1996), the principal heuristic in the fast and frugal literature. At the same time she used the second method in estimating the predictive fit of a compensatory decision rule called the Franklin rule. Thus, it is possible that her results – a better predictive fit of the matching heuristic over the Franklin rule – is due to her method of estimating fit rather than to a better correspondence of the matching heuristic to the cognitive process underlying the decisions. Below we illustrate this argument using simulated data. We generate data in which – notwithstanding random error – the relationship between the independent variables and the dependent variable is perfectly described by a compensatory relationship and we show that the predictive fit of the matching heuristic is higher than the predictive fit of the Franklin rule (as well as the regression model).

Simulation

Dhami collected 'bail or jail' decisions of two courts, 159 and 183 respectively. In addition, she collected 25 dichotomous characteristic (non-dichotomous characteristics were dichotomized), or cues, that may affect the courts' decisions, some of them demographics (e.g., sex, age) and some legal (e.g., previous offenses). The average inter-cue correlation was 0.2 and 0.3 for the two courts, respectively, and the proportion of punitive decisions was 41% and 54%, respectively.

In order to evaluate the predictive fit of the matching heuristic, the Franklin rule and regression, we used Monte Carlo simulations. For each simulation we generated 100 datasets similar in their statistical characteristics to the datasets used by Dhami (2003). For each data set we generated 172 values for each of 25 variables from a multi-variate standard normal distribution with all correlations terms equal to 0.25. Subsequently, all variables were transformed to dichotomous *cues*. A value of 1 was given to a cue if the value of the variable was greater than or equal to zero; a value of zero otherwise. The dependent variable was created by a weighted sum of the first three cues to which a standardized normal random 'error' was added. The dependent variable was then transformed into a dichotomous criterion and assigned a value of 1 if its value was above a certain value and zero otherwise.

In the simulations we varied the following parameters: the spread of attribute weights, the size of the error and the distribution of the criterion (i.e., the ratio of '1' to '0'). We had two reasons for varying these parameters. First, we wanted to replicate the important statistical characteristics of Dhami's data as closely as possible (i.e., not only the number of cues, sample size and inter-cue correlation, but also the distribution of the criterion and the estimated K, as defined below). Second, we were interested in the effect of these parameters on the relative fit of the competing models.

To calculate the fit of the models we used the first 86 'observations' as a modeling sample and the later 86 'observations' as a validation sample. A detailed description of the rules and the way they were applied to the data in the calculation of fit is given in the appendix. Here it will suffice to say that the Franklin rule is essentially a compensatory strategy in which cue weights are determined by the joint frequencies of criterion and cue values and that the matching heuristic is essentially a disjunctive rule that uses K cues, where K maximizes the fit in the modeling sample.

The results of the simulations are reported in Table 1. For the sake of brevity, only the results of representative simulations are reported. However, the results of the rest of the simulations were consistent with the reported results.

The main result of the simulations is that – although the true generating process is compensatory – over a wide range of parameter values the predictive fit of the matching heuristic exceeds the predictive fit of the Franklin rule as well as the predictive fit of the regression model. This result is consistent with our analysis of cue relevance. Note also that the within-sample fit of the matching heuristic also exceeds the fit of the Franklin rule. This result is consistent with Broder (2002).

The results of our simulation also indicate some issues that should be considered in assessing the validity of the matching heuristic in comparative model fit studies. First, ceteris paribus, an increase in the error affect the estimated K; and second, the relative fit of the matching heuristic vs. the compensatory rules depends on the parameters that were varied, and, in particular, the size of the error and the distribution of the criterion.

Discussion

Our results demonstrate that tests regarding the predictive validity of the matching heuristic (and the Take the Best heuristic) are problematic. Consistent with Broder (2002) the results also suggest that tests that are based on within-sample fit (i.e., the fit in the modeling stage) are problematic as well. However, since the focus of the paper is on the former tests, our discussion here will concentrates on the first issue.

The reason that the matching heuristic fares better than the compensatory in cross-validation models is the inclusion of irrelevant cues that 'ruin' the predictive fit. Therefore, a fair test of the (predictive) fit of competing models requires some method for identifying *relevan*t cues. We do not think that such a general method can be found. However, in some situation the problem of cue identification may be less severe. In particular, it could be argued that in judgment and decision *experiments* – at least experiments that do not involve many cues – all cues that are presented to the subjects are relevant. Thus for example, the question of cue relevance does not arise in the voluminous literature on policy capturing (e.g., Brehmer & Joyce, 1988), most likely because consistent with the researchers' modeling practice, participants incorporated all the cues they were presented with in their responses (in the same vein Hogarth & Kareleia's, 2005, simulation of the predictive fit of the Take the Best heuristic also assumed the relevance of all cues).

On the other hand, in field studies such as Dahmi (2003) it is not possible to identify which cues are relevant. Thus, the choice of the number of cues – and therefore the predictive fit of compensatory models – is rather arbitrary. Furthermore, when Take the Best is applied to binary choice tasks such as the Gigerenzer & Goldstein (1996) task, it is in principle even impossible to distinguish between relevant and irrelevant cues, since when the two alternatives are equal on all other cues, any cue, no matter how 'unimportant' it is, can affect the choice.

What recommendations could be made for situations in which the researcher is interested in comparing the fit of two process models, one that involves a preliminary cue selection and another that does not? One possible solution is to use algorithms that maximize predictive fit for both models. Another possible solution is to initially select a subset of relevant cues and subsequently compare the two models based on these cues alone. Possible methods for such a selection may involve statistical screening (e.g., selecting only cues that are significantly related to the criterion) or self-report based screening (e.g., asking decision makers what are the relevant cues). Finally, it should be noted that from the point of view of the problems discussed in the current paper, comparing within-sample fit is more reliable than comparing predictive fit.

References

Brehmer, B., & Joyce, C. R. B. (1988). *Human judgment: The SJT views*. Elsevier Science Publisher, North Holland.

Bröder, A. (2002). Bailing and jailing the unknown way: A critical examination of a study reported by Dhami and Ayton (2001). *Berichte aus dem Psychologischen Institut der Universität Bonn,28* (1).

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332–361.

Dhami, M. (2003). Psychological models of professional decision making. *Psychological Science*, *14*, 175-180.

Dhami, M., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making, 14*, 141-168.

Dhami, M. K., & Harries, C. (2001). Fast and frugal versus regression models of human judgment. *Thinking and Reasoning, 7,* 5-27.

Ganzach, Y. (1995). Nonlinear models of clinical judgment: Meehl's data revisited. *Psychological Bulletin, 118,* 422-429.

Gigerenzer, G., Czerlinski, J., & Martignon, L. (1998). How good are fast and frugal heuristics? In J. Shanteau, B. Mellors, & D. Schum (Eds.), *Decision research from Bayesian approaches to normative systems: Reflections on the contributions of Ward Edwards*. Norwell, MA: Kluwer.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-699.

Gigerenzer, G., Hoffrage, U., & Kleinboalting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Goldberg, L. R. (1970). Man versus model of man: A rationale plus some evidence of improving clinical inference. *Psychological Bulletin*, 73, 422-432.

Hogarth, R., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less ‘‘more’’? *Journal of Mathematical Psychology, 49*, 115-124.

Martignon, L., & Blackmond Laskey, K. (1999). Bayesian benchmarks for fast and frugal heuristics. In G. Gigerenzer, P. Todd & the ABC Research Group, *Simple heuristics that make us smart* (169-188). Oxford, UK: Oxford University Press.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal and fit: Simple heuristics for paired comparison. *Theory and Decision*, *52*, 29-71.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358-367.

Table 1: The simulation results

	Equal ratio[a]				High ratio[a]			
	Low error[b]		High error[b]		Low error		High error	
	Low[c] attr. spread	High[c] attr. spread	Low attr. spread	High attr. spread	Low attr. spread	High attr. spread	Low attr. spread	High attr. spread
In sample fit								
Regression	.99	.97	.78	.78	.97	.98	.85	.85
Franklin rule	.78	.76	.67	.66	.68	.69	.64	.64
Matching heuris.	.84	.86	.68	.68	.94	.96	.79	.79
Predictive fit								
Regression	.93	.88	.57	.58	.84	.88	.69	.69
Franklin rule	.76	.74	.59	.60	.67	.67	.58	.57
Matching heuris.	.78	.83	.60	.61	.92	.94	.73	.73
Estimated K	1.4	1.2	1.3	1.3	2.9	2.3	4.5	4.2

Note: averages of 100 simulations
[a] equal ratio: 50% of '1's and 50% of '0's; high ratio: 75% of '1's and 25% of '0's
[b] low: error standard deviation of 0.1; high: error standard deviation of 1.0
[c] low: all three weights equal 0.5; high: weights are 0.3, 0.5, 0.7.

Appendix

<u>Franklin Rule</u>:

Let:

$P_{11j} \equiv P(Z_i = 1 | X_{ij} = 1)$ = the proportion in the estimation sample of observations in which the dichotomous criterion Z equals 1 out of all observations in which the dichotomous cue X_j equals 1.

$P_{10j} \equiv P(Z_i = 1 | X_{ij} = 0)$ = the proportion in the estimation sample of observations in which the criterion equals 1 out of all observations in which cue X_j equals 0.

$$U(X_{ij} = 1) = \begin{cases} 1 \text{ if } & \mathrm{P}_{11\mathrm{j}} \geq P_{10j} \\ 0 \text{ otherwise} \end{cases}$$

$$U(X_{ij} = 0) = 1 - U(X_{ij} = 1)$$

$$W_j = \underset{X}{Max}\{P_{11j}, P_{10j}\}$$

$$T_i = \sum_{j=1}^{J} W_j U(X_{ij})$$

$$\overline{T} = \frac{\sum_{i}^{n} T_i}{n}$$

J is the number of cues in the data set; n is the number of observations in the estimation sample.

Prediction rule: $\hat{Z}_i = \begin{cases} 1 \text{ if } & T_i \geq \overline{T} \\ 0 & \text{otherwise} \end{cases}$

Matching Heuristic

$$U(X_{ij}=1)=\begin{cases}1 \text{ if } & \sum_{i=1}^{n} X_{ij}Z_i > \sum_{i=1}^{n}(1-X_{ij})Z_i \\ 0 \text{ if } & \sum_{i=1}^{n} X_{ij}Z_i < \sum_{i=1}^{n}(1-X_{ij})Z_i\end{cases}$$

$$\textit{If} \quad \sum_{i=1}^{n} X_{ij}Z_i = \sum_{i=1}^{n}(1-X_{ij})Z_i \quad \text{Then}$$

$$U(X_{ij}=1)=\begin{cases}1 \text{ if } & \sum_{i=1}^{n} X_{ij}(1-Z_i) < \sum_{i=1}^{n}(1-X_{ij})(1-Z_i) \\ 0 \text{ if } & \sum_{i=1}^{n} X_{ij}(1-Z_i) > \sum_{i=1}^{n}(1-X_{ij})(1-Z_i)\end{cases}$$

$$\textit{If} \quad \sum_{i=1}^{n} X_{ij}(1-Z_i) = \sum_{i=1}^{n}(1-X_{ij})(1-Z_i) \quad \text{Then}$$

Assign randomly $U(X_{ij}=1)=1{,}0$

$U(X_{ij}=0)=1-\mathrm{U}(\mathrm{X}_{ij}=1)$

Let P_{11j} be the proportion of observations with Z_i=1 out of all cases in the estimation sample for which $U(X_{ij})=1$. Rank the J independent cues according to P_{11j} in descending order so that the first cue has the highest P_{11j}. Search the first K cues until the first occurrence of $U(X_{ij})=1$ and assign prediction $\hat{Z}_i=1$. If all first K cues yield $U(X_{ij})=0$, assign prediction $\hat{Z}_i=0$. Find 1<K<J that yields the highest overall fit. Apply matching rule with that K to the holdout sample.

Regression

Use the least squares method to regress Z on all J independent variables (X) in fitting the data. For fit and prediction apply the following rule: If estimated Z is greater than or equal to 0.5 give the observation a predicted value of one, otherwise predict it as zero.