**When Should We Use Simple Decision Models?**

**A Synthesis of Various Research Strands**

September 8, 2017

Konstantinos V. Katsikopoulos1, 2, Ian N. Durbach3, 4 and Theodor J. Stewart5, 6

1 Department of Decision, Analytics and Risk, Southampton Business School, University of Southampton, UK

2 Centre for Adaptive Behaviour and Cognition, Max Planck Institute for Human Development, Germany

3 Centre for Statistics in Ecology, the Environment and Conservation, Department of Statistical Sciences, University of Cape Town, South Africa

4 African Institute for Mathematical Sciences, South Africa

5 Department of Statistical Sciences, University of Cape Town, South Africa

6 Alliance Manchester Business School, University of Manchester, UK

**Abstract**

Many decisions can be analyzed and supported by quantitative models. These models tend to be complex psychologically in that they require the elicitation and combination of quantities such as probabilities, utilities, and weights. They may be simplified so that they become more transparent, and lead to increased trust, reflection, and insight. These potential benefits of simplicity should be weighed against its potential costs, notably possible decreases in performance. We review and synthesize research that has used mathematical analyses and computer simulations to investigate if and when simple models perform worse, equal, or better than more complex models. Various research strands have pursued this, but have not reached the same conclusions: Work on frequently repeated decisions as in inference and forecasting—which typically are operational and involve one or a few decision makers—has put forth conditions under which simple models are more accurate than more complex ones, and some researchers have proposed that simple models should be preferred. On the other hand, work on more or less one-off decisions as in preference and multi-criteria analysis—which typically are strategic and involve group decision making and multiple stakeholders—has concluded that simple models can at best approximate satisfactorily the more complex models. We show how these conclusions can be reconciled. Additionally, we discuss the theory available for explaining the relative performance of simple and more complex models. Finally, we present an aid to help determine if a simple model should be used, or not, for a particular type of decision problem.

**Keywords**

Decision making/process, decision support systems, forecasting, heuristics, multicriteria, statistics

**1. Introduction: Quantitative Models for Decision Analysis and Support**

Few decision analysts today would blindly use quantitative models. And just as few would outright reject them. Models make clear that decisions consist of entities such as options, attributes of options, values of attributes, utilities of values, that tradeoffs may need to be made, that some options are dominated by others, and so on. Models may be used to derive initial solutions, which can then be accepted as they are or improved further, or be used to inspire other models and solutions.

Whereas some decision problems are so “wicked” (Rosenhead and Mingers, 2011), or “complex” and “chaotic” (French, 2013), that quantitative models cannot help, there are important decision problems to which models do apply: For example, personal choices can be supported by multi-attribute utility functions (Keeney and Raiffa, 1976), credit scoring can be tackled by regression (Thomas, Edelman, and Crook, 2002), and product demand can be forecasted by time series models (Makridakis, Wheelwright, and Hyndman, 1998).

An important practical difficulty is that decision makers often resist the models offered by decision theorists. It is not hard to see why: The models tend to be too complex psychologically, in the sense that they require the elicitation and combination of quantities such as probabilities, utilities, and weights. Even if software does most of the work, decision makers have reasons to feel overwhelmed or fail to understand the premises, concepts, and computations of the models.

Simplification holds a strong appeal for decision analysis and support. Trust in the decision process and opportunities for reflection and insight are critical elements of good decision analysis and support, and these elements are typically easier to achieve when decision makers understand the tools they are using. This favors the use of simple rather than complex models. But, of course, caution is necessary: One can go too far and oversimplify. The potential benefits of simplicity, such as increased transparency, trust, reflection and insight, should be weighed against its potential costs, notably possible decreases in performance.

In considering the role of quantitative models—simple or not—for decision analysis and support, it is important to recognise the diversity of problem settings to which these models may be applied. The following categories are near the ends of a spectrum of these settings.

1. Frequently repeated decisions, typically operational, involving one or a few decision makers. Examples include predicting consumer choice behaviour and credit scoring.
2. More or less one-off decisions, typically strategic, involving group decision making and multiple stakeholders. Examples include deciding the amount of corporate investment in a new production facility and its location.

The first of the above two categories is allied to statistical models for prediction or forecasting of the future. This is self-evident in problems of consumer choice behaviour, but it is also relevant in a mode of analysis and support, as in credit scoring. Although a credit scoring model may initially be viewed as an aid to decisions regarding whether or not to grant credit, it may also be seen as predicting what a well informed and experienced expert would decide. It is possible to compare recommendations of the model with a set of expert evaluations, which establish a kind of ground truth.

In the second of the two categories, however, there is no clear means of establishing a ground truth in order to test and validate a quantitative model. Experience with problems of the first category may provide some basis for adopting particular models in the second category, but caution is needed as there are no repeated trials to compensate for one decision being wrong.

Even though we recognise that many decision problems will not fall at either extreme, here we examine the potential of simple decision models in exactly these two categories. Additionally, we make a distinction between problems of repeated operational decisions, involving (*i*) inference; that is, determination of the category of an option—as when classifying a customer as active or not—or some other characteristic of an option, and (*ii*) forecasting; that is, predicting the worth or value of a decision option—such as a company’s stock—in the future, as this reflects a divide often encountered in the literature. Thus, we consider three research strands: *inference*, *forecasting*, and *strategic* decision making.

The authors of this article are decision modelers who have, for the most part, worked on a single one of the strands. We essentially used the same formal methods and often tested the same simple models, such as equal weighting of attributes or sequential processing of attributes (Hogarth and Karelaia, 2005). But we have not arrived at the same conclusions. And, the conclusions of the third strand were different as well. As we scanned the academic literature, we realized that this is a general issue and there is not much communication among the inference, forecasting, and strategic decision making strands. Analogously, our consulting experiences suggest that practitioners have diverse impressions.

We acknowledge that the formal research from which we draw here does not so much speak to how models fare on dimensions such as transparency and insight, and thus cannot provide a full answer to which models should be used in practice (Durbach and Stewart, 2012). In a sense, this work contributes more to normative rather than to prescriptive knowledge about decision models (Bell, Raiffa, and Tversky, 1988; Smith and von Winterfeldt, 2004). We believe, however, that such research is key for deciding whether or not to use simple models in practice. If simple models perform equally well or better than more complex models, then it seems that they should be employed (assuming that they fare better on dimensions such as transparency and insight).

Our article has three main goals. The first is to review in one place the main empirical findings of mathematical and simulation research on the relative performance of simple and complex decision models, and to synthesize those findings. Our synthesis reveals that the conclusions of the various strands can be reconciled. The second goal is to reflect on the theory available for specifying *a-priori* if simple or complex models should be preferred for a particular decision problem. Our third goal is to translate the available theory to an aid to help determine if simple models should be used, or not, for a particular decision problem.

The remainder of the article is structured as follows. Sections 2 and 3 review material on the performance of simple decision models for inference and forecasting, and strategic decision making respectively. In order to speak to a broad audience, the treatment is not mathematical, but we have strived to remain precise. We provide just a limited amount of technical detail. Notably, we refrain from providing a technical definition of simplicity—which is indeed a thorny issue (Cutting, 2000)—and confine ourselves to just labelling nested versions of models as simpler than the original models. Section 4 summarizes common elements and reconciles apparently contradictory findings from the research strands. Section 5 discusses the theory available for explaining the relative performance of simple and more complex models. Based on this theory, Section 6 presents an aid that can be used to help determine if a simple model should be used, or not, for a particular type of decision problem.

**2. Simple Models for Repeated Operational Decisions**

**2.1. Inference**

As indicated previously, repeated operational decisions are closely linked to statistical prediction. A typical empirical study of simple inference models is the following.

In a project sponsored by the Bank of England, Aikman et al. (2014) considered the problem of predicting which of the 118 global banks that had at least 100 billion USD in assets at the end of 2006, went bankrupt during the financial crisis and which did not. After the fact, it is known that 43 banks failed and 75 banks survived the crisis (for definitions of bank failure, see Laeven and Valencia, 2010). For each bank, the authors gathered data on a number of economic indicators, such as leverage ratio (i.e., the proportion of a bank’s capital that is not based on debt), the amount of wholesale funding (e.g., government or public funding), and so on. In decision analysis jargon, these indicators would be called attributes, and they would be called predictors in statistics, features in artificial intelligence, and cues in psychology.

The research problem is how to mathematically combine the available attributes in order to predict bank failure reasonably well. A standard macro-economic solution is to use logistic regression (Laeven and Valencia, 2010). In order to keep the model tractable, the four most statistically informative attributes were used. The authors compared the performance of logistic regression with that of a family of simpler models, *fast and frugal decision trees*. An example of a fast and frugal decision tree, developed based on economic intuition by some of the authors, is depicted in Figure 1: Instead of always trading off all four attributes to categorize a bank, this simple tree goes through attributes one at a time, asks a binary question on each attribute’s value, where for each question there is a possible answer that allows a decision to be made so that the process terminates. More generally, such models are also *lexicographic.* Because the tree is intended as a decision support tool for financial regulators, it does not assign a bank to a “fail” or “survive” category, but rather suggests whether a bank is at a risk for failing and thus should be monitored (red flag), or not (green flag). Taking the thresholds used to dichotomize attributes (e.g., 4.1% for the leverage ratio) as given, this tree is a model nested in the family of all linear models. The thresholds were estimated statistically, and we acknowledge that one might see this as an indication of some complexity. On the other hand, note that logistic regression estimates thresholds as well.



Figure 1. A fast and frugal tree for deciding whether a bank is at a risk for failing and thus should be monitored (red flag), or not (green flag). This simple tree, developed based on intuition, was not outperformed by any of 20 statistically calibrated logistic regressions (Aikman et al., 2014).

An important point of this study is that models were evaluated on a test set, after their parameters—for example, weights of attributes in a regression or order of attributes in a tree—had been calibrated on a training set and their values fixed. This process is typically repeated hundreds or thousands of times, in order to average out random variation. Because there is no overlap between the test and training sets, this procedure measures predictive power, uncontaminated by the possible over-fitting of a particular known dataset.

The Figure 1 tree assigned a green flag to 82% of the survived banks and a red flag to 50% of the banks that failed. None of 20 statistically calibrated versions of logistic regression—each version had a different value on a parameter controlling the tradeoff between hits and false alarms—could outperform this intuitive tree. Furthermore, 20 statistically calibrated fast and frugal trees performed competitively with the 20 logistic regressions. In fact, the trees performed more accurately than the logistic regressions when the wholesale-funding attribute was omitted, a treatment which was meant to simulate the case of more sparse or lower quality data.

More generally, empirical studies of simple inference models tend to share three features: (*i*) At least one simple and one complex model are evaluated, (*ii*) the evaluation is focused on predicting, rather than fitting data, and (*iii*) the data used is real, not synthetic.

Theoretical studies sometimes do not conform to (*ii*) and (*iii*) because mathematical analysis may be easier when fitting a dataset which is assumed to have some convenient properties (Baucells, Carasco and Hogarth, 2008; Katsikopoulos, 2013). Some empirical studies also do not conform to (*i*) when it is not clear if and how the available data suffices for building a complex model. In this case, the performance of a simple model can be compared with the performance of unaided decision makers. For example, when categorizing traffic approaching a military checkpoint as hostile or nonhostile, a fast and frugal tree has been compared with the actual decision making of the soldiers manning checkpoints (Keller and Katsikopoulos, 2016). Had the tree been used in 1,053 incidents in NATO checkpoints in Afghanistan between 2004 and 2009, the number of civilian casualties would have decreased from 204 to 78.

The examples provided so far hint that simple models may compete well with more complex models (Todd, 2007; Katsikopoulos, 2011). Is there more systematic evidence for or against this claim? We discuss three studies which used a range of inference problems to compare the performance of simple and more complex models.

First, Czerlinski, Gigerenzer, and Goldstein (1999) estimated the performance of linear regression, *take-the-best* (which is a lexicographic model that makes a decision based on only one attribute and can be formally represented as a fast and frugal tree; Gigerenzer and Goldstein, 1996), and *tallying* (i.e., a linear model with all attribute weights equal to one) on 20 datasets from the fields of biology, environmental science, demography, health, psychology, and transportation. These models are similar to the outranking methods used in preference research. The main finding of the study was that, on the average, the predictive accuracy of take-the-best equaled that of linear regression, 76%, whereas tallying scored 69% (training sets consisted of 50% of all decision options and their attribute values). This result has been replicated and also extended to include other standard models from statistics and artificial intelligence (e.g., naïve Bayes) and very small training sets, which may be more realistic for decision making under changing conditions. For example, Katsikopoulos, Schooler and Hertwig (2010) tested training sets with sizes ranging from 3% to 15% of the datasets on average. Their main findings are that tallying is the best-performing method for the smallest training set, and that take-the-best is the best method for the other training sets, outperforming naïve Bayes by more than 5% on average (naïve Bayes scored 73% in the 50% training set). Şimşek and Buckmann (2015) extended these findings by using more datasets and more models.

Second, Martignon, Katsikopoulos and Woike (2008) estimated the performance of classification and regression trees (Breiman et al., 1984), logistic regression and two methods of building fast and frugal trees, on 30 datasets from the *UC Irvine Machine Learning Repository*. Eleven datasets referred to medical decisions. The size of the training set was 15%, 50% or 85% of each dataset. The main finding was again a strong effect of training set size. The best standard model outperformed the best simple model for the largest training set by 4% (82% vs. 78%) on the average, but this difference shrunk to 1% (76% vs. 75%) for the smallest training set.

In a third study looking at a broad range of inference problems, Şimşek (2013) extended the design of Czerlinski et al (1999), by using their 20 datasets plus 31 more, spanning the fields of biology, business, computer science, ecology, economics, education, engineering, environmental science, medicine, political science, psychology, sociology, sports, and transportation. She also used a state-of-the-art version of linear regression with elastic net regularization, plus a number of simple models, including take-the-best. The size of the training set was equal to the size of the whole dataset minus one. The main finding was that regularized linear regression scored 79% and take-the-best 78% on average.

In sum, one can see some empirical regularities about when do simple inference models perform better than more complex ones, and vice versa. Sparse or low-quality data tend to favor simple models. These problems can be called “difficult”. At the other extreme of “easy” problems, simple models also compete well with more complex models. For example, both fast and frugal trees and support vector machines achieve 100% hit rate and 2% false alarm rate in detecting unexploded ordnance (i.e., munitions used in war or military practice; Fernandez, Katsikopoulos, and Shubitizde, 2010). On the other hand, more complex models tend to be better than simpler ones on problems that are neither difficult nor easy. These empirics will be discussed further in Section 5.1.

**2.2. Forecasting**

Forecasting lies somewhere between inference and preference. It resembles inference more closely than preference because forecasts, can typically be proven right or wrong. There are exceptions, however, as in policy making where outcomes are sometimes subject to interpretation (Tetlock, 2006); and in challenging areas, such as product demand forecasting, a right forecast is often interpreted as right within an acceptable margin of error. On the other hand, forecasting is often about predicting preferences as for example in business forecasting.

We focus on two types of evidence-based forecasting models, called *causal* and *extrapolation* models. Consider the problem of forecasting product demand.

Causal models in forecasting may be linear or non-linear functions of input variables. The demand for a product could be predicted by a linear function of variables such as price, country of origin, and so on. As is the case in inference, a common way of simplifying linear regressions in forecasting is by setting the weights of all variables equal to one (Graefe, 2015), or by using only one variable (Nikolopoulos et al., 2007).

Extrapolation models use observed demands from the past in order to estimate future demand. In time-series models, it is common that the weight of an observation decreases the further back in the past this observation was made. A very straightforward way of implementing this idea is simple exponential smoothing, where the demand forecasted for tomorrow is a weighted linear combination of the demand that was forecasted for today plus the demand observed today, as depicted in Figure 2.

Demand forecasted for tomorrow

 Demand observed today

Demand forecasted for today

 weighted by *w* weighted by 1 – *w*

Figure 2. Simple exponential smoothing: A time-series model for forecasting future demand for a product.

Beyond the first forecast that needs to be made, there is only one free parameter in this model, *w* (ranging between 0 and 1). More complex models may have more parameters and employ a non-linear functional form in order to take into account influences such as seasonality, intermittence, and other trends (Hyndman et al., 2002; Syntetos, Boylan, and Croston, 2005).

Empirical studies of simple forecasting models share the three features of empirical studies on simple inference models: (*i*) Both simple and complex models are evaluated, (*ii*) the evaluation is focused on prediction, and (*iii*) real datasets are used.

Makridakis and colleagues have run a number of large-scale empirical studies evaluating the performance of a range of forecasting methods over hundreds or thousands of datasets, where the creators of methods and other experts were invited to submit forecasts which were then evaluated by the competition team (Makridakis, Hibon, and Moser, 1979; Makridakis et al., 1998; Makridakis and Hibon, 2000; Makridakis, Hogarth, and Gaba, 2009). The main conclusions of the original study and of the three subsequent “M-competitions” were that (a) complex models did not, on average, perform better than simple models; (b) the relative performances of models depended on the accuracy measure assumed and the length of the forecast horizon; (c) combining forecasts from different models improved forecast accuracy.

While (b) and (c) are perhaps obvious and were accepted by the community, (a) provoked a strong and mostly negative response (see the commentaries to Makridakis et al., 1979, in the *Journal of the Royal Statistical Society*). With the exception of a small group of forecasting researchers which embrace, or at least give a chance to simple models, “complexity remains popular among researchers, forecasters and clients” (Green and Armstrong, 2015, p. 1678). Complex models are alive and well (Goodwin, 2001). On the other hand, practitioners are sometimes open to simple forecasting models (Katsikopoulos and Syntetos, 2016; Kolassa, 2016).

Despite the resistance to simple forecasting models, conclusive evidence favoring complex models has not been forthcoming, at least for the kind of large-scale comparative forecasts typified by the M-competitions. If anything, the balance of evidence seems to continue to support simple models (Ghandar, Michalewicz, and Zurbruegg, 2016; Meade and Islam, 2015). Only in a meta-analysis of tourism demand forecasting literature, Peng, Song, and Crouch (2014) found that complex models, such as dynamic econometric and artificial intelligence methods, achieved 10% and 13% mean absolute percentage error respectively, slightly outperforming time series (e.g., ARIMA, 13%) and naïve forecasting models (16%).

Furthermore, Green and Armstrong (2015) introduced a special issue on simple versus complex forecasting in the *Journal of Business Research* with a review of evidence on the relative predictive accuracy of simple and complex forecasting models. They appear to have made a very fair effort to include studies from researchers known to have found evidence or to advocate for either one of the two model families. In the end, they identified 32 journal articles and book chapters, of which 25 articles and chapters reported a total of 97 quantitative comparisons between simple and complex models. They found that, averaged across the different forecasting error measures reported, complex models increased error by 27% compared to simple models. The error increase was 28% for extrapolation models (based on 17 articles and chapters) and 25% for causal models (based on 8 articles and chapters).

Perhaps a balanced view on this controversial topic is that for complex models to outperform simple ones, two conditions must be met: (1) The real-world process must be complex, and (2) the forecaster must able to model this complexity correctly. Our own literature search revealed a few applications, mostly relating to forecasting physical phenomena, in which these conditions would appear to be satisfied and complex models do in fact outperform simple ones. These include short-term (daily or hourly) forecasts of wind power generation (Foley et al., 2008), in which complex atmospheric dynamics must be modeled, and new product sales (Fader, Hardie, and Huang, 2004), where buying behavior for the new product implies changing inter-purchase times, that are best captured by a dynamic change-point model. In these cases, sufficiently detailed data[[1]](#footnote-1) and knowledge of the underlying dynamics means that complex models developed specifically to capture these deliver a measurable improvement over simpler benchmarks.

Overall though, the evidence suggests that in many application areas at least one of these conditions is missing, and simple models are, at the present time, capable of at least as good performance as complex ones. Consistent, Green and Armstrong, Green, and Gräfe (2015) advocate simple forecasting models. Furthermore, Armstrong, (2015) put together a number of simple guidelines for building simple forecasting models, which are deduced from what they call the “golden rule of forecasting”, *be conservative*.

These bold proposals by the very considerable authority of Armstrong were met in the same special issue with strong counterarguments by other authorities. For example, Fildes and Petropoulos (2015) asked whether there is really a single golden rule. They argued that there is not a universal model which can tackle every forecasting problem.

Even though we can see why, on the face of the preferential treatment that complex models currently enjoy, some forecasters may want to emphasize the value of simple models, we would also like to highlight the need for a theory of conditions under which simple, or complex, forecasting should be used. The forecasting literature per se has not produced a complete theory (for beginnings, see Katsikopoulos and Syntetos, 2016; and Kolassa, 2016). Because of the formal similarities between inference and forecasting modeling, however, there are some theoretical results, which will be discussed together in Section 5.1.

The empirical results on the performance of simple models in repeated operational decisions provide a *prima facie* case for exploring their use in strategic one-off decisions. But a significant issue is that for strategic problems it is not appropriate to evaluate simple models by averaging performance over repeated decisions, as the current decision is one-off. More carefully considered results are needed, and we review those in the next section.

**3. Simple Models for Strategic One-Off Decisions**

The models that we consider here aim to structure the evaluation of possible options so that the decision analysis process satisfies the preferences of the decision maker(s) (Belton and Stewart, 2002). These models cover a wide range of the modeling used for multi-criteria decision analysis and support, and we take the liberty of using the single label “preference” to refer to this work.

Some of the best-known preference models—multiattribute value theory and its stochastic version, multiattribute utility theory (MAUT)—have been shown to be logically equivalent to a collection of axioms, which, at least at a first glance, appear to be reasonable. Models like MAUT can be considered of intermediate complexity. On the one hand, observed choices often violate the axioms of these models, and some authors contend that these violations are detrimental for prescription as well (Bleichrodt, Pinto, and Wakker, 2001a). Thus, less restrictive models, which describe observed behavior better, have also been proposed. These models are invariably more complex, the most notable example being cumulative prospect theory, originally developed by Tversky and Kahneman (1992) and also extended for prescriptive use (Abdellaoui, 2000; Bleichrodt, Pinto, and Wakker, 2001b; Bleichrodt, Doctor, Filko, and Wakker, 2011; Bleichrodt, Schmidt, and Zank, 2009). On the other hand, the implementation of models such as MAUT requires the careful assessment of a number of parameters, mostly involving trade-offs. This is time consuming as well as cognitively taxing. Therefore, simpler models than MAUT such as those using ordinal weights (e.g. Barron and Barrett, 1996) or best-worst scaling (Rezaei, 2016) have also been proposed, which bypass the need for some of the assessment questions.

It has been suggested that many simplifications result in relatively little change to model inputs, which in turn have basically little impact on decision outcomes. Often referred to as the *flat maximum effect* (Edwards and von Winterfeldt, 1986; Lovie and Lovie, 1986; Katsikopoulos, 2011), this view contends that aggregate outcomes such as expected utility and value, as well as the choices derived from them, are seldom affected by the precision of the model form or of the input numbers (assuming that aspects of good decision practice, such as removing dominated options, are satisfied). Thus, in this case, gains in procedural quality obtained with simpler models need not necessarily come at the expense of significant loss in the quality of the chosen option.

For example, an important component of most preference models are cardinal attribute weights, which express acceptable trade-off levels between values on different attributes (Belton and Stewart, 2002) or more vague notions of attribute importance (Roy, 1985; Saaty, 1980). Several authors have explored simple models in which weights are assessed only imprecisely, either ordinally, using arbitrary linear constraints, or as intervals. Models then use this imprecise information in a number of ways, by creating cardinal summaries of the imprecise weights (Ahn and Park, 2008; Barron and Barrett, 1996), simulating over plausible weight ranges (Lahdelma and Salminen, 2001; Durbach and Calder, 2015), or building models that make only imprecise recommendations (Kadziński and Tervonen, 2013).

Many empirical findings in this area are consistent with the flat maximum effect. Barron and Barrett (1996), for example, tested various rank-based summaries of ordinal weights, and found that a linear model using rank-order centroid weights identified options that (*a*) were the same as the option selected using known cardinal weights in between 80% and 90% of simulated cases, (*b*) achieved well over 95% of the range between minimum and maximum values, and (*c*) improved in terms of approximation accuracy when attributes were positively correlated. Others who have performed similar simulation-based tests on models with simplified weights include Ahn and Park (2008), Ahn (2011), Barron and Barrett (1996), Durbach and Calder (2016), and Kirkwood and Corner (1993).

Each of these findings is representative of a broader theme: (*a*) Mean or median approximations are often very good; (*b*) the value or utility of sub-optimal options is often very close to that of the optimal option; and (*c*) environmental features can impact the approximation accuracy of simple models.

A second component of preference models amenable to simplification is the within- and between-attribute evaluation of the performance of an option. Typically, marginal (single-attribute) value or utility functions are assessed using trade-off questions similar to those employed for weights as discussed above. The choice of a mathematical form for aggregating marginal functions is closely associated with conditions restricting the kinds of dependencies that can exist between attributes. Assessment questions, again of the trade-off variety, can be used to test which of these conditions hold for any particular problem.

Stewart (1995, 1996) used a number of simulation experiments to test the effects of simple attribute-independence assumptions, and of the number of points used per attribute. He found that three intermediate points were adequate to achieve close approximations, and that in many cases only one point was sufficient, but that poor results were often obtained if a linear aggregation function was assumed incorrectly. Whereas an additive utility function led to very little deterioration in the utility of the selected option when multiplicative aggregation was more appropriate (Stewart, 1995); incorrectly assuming preferential independence—which guarantees an additive value function—led to a substantial worsening of the utility of the chosen option (Stewart, 1996).

Abdellaoui, Bleichrodt, and Paraschiv (2007) evaluated a simplification to the assessment of prospect-theory utility functions, which assumes a power function to reduce the number of assessment questions required. In experiments with human participants, they showed little or no difference between utility functions obtained using their method and those obtained with more complex assessment.

A final area where simplifications can be fruitfully applied is in the assessment of attribute values. Often these values will be uncertain. In these cases the canonical approach is to represent uncertainty via probability distributions, as done in MAUT for example. A number of protocols for assessing subjective probability distributions exist (O’Hagan et al., 2006), but they are time consuming and cognitively difficult. Simpler approaches replace the probability distributions with a variety of summary measures: expected values, variances, or a small number of quantiles. In simulation studies the use of quantiles has been observed to closely approximate the performance of full-information MAUT (Durbach and Stewart, 2012), and even expected values performed surprisingly well (Durbach and Stewart, 2009). Importantly, good approximations were found to depend crucially on the summary measures being assessed with good accuracy—indeed this was more important that the summary measure used. Variances appear to perform poorly on average, although Kirkwood (1992) provides conditions under which good approximation accuracy can be achieved with variances.

The discussion above referred to value or utility function models because most empirical work on simplification of preference models has been carried out in this area. Many of these findings, however, would surely apply to other schools of decision aiding such as outranking approaches. Outranking methods can themselves be viewed as simple models. They construct preference relations indicating whether there is sufficient evidence that one alternative is at least as good as another. Some of their key features are: focus on pairwise comparisons, non-compensatory preferences, and allowing for incomparability. In these ways, outranking methods are similar to simple models for inference such as take-the-best and tallying (see Section 2.1).

Even within outranking methods, variants range from simple, such as Electre I, to complex, such as Electre III (Vincke, 1999). Imprecise outranking methods have also been proposed (Corrente, Figueira, and Greco, 2014; Durbach, 2014; Hokkanen, Lahdelma, Miettinen, and Salminen, 1998). Very little work, however, has been done evaluating the effects of these simplifications, either by comparing outranking methods to each other, or by comparing outranking and other methods.

In sum, simplification has a long history in preference. Many studies have compared how closely the performance of simple models approximates that of full-information models. Almost exclusively, these are simulation studies. Various environmental conditions—variability of and correlations between attribute values, numbers of alternatives and attributes, shape of marginal value/utility functions—have been found to have moderate effects on the ability of simple models to approximate more complex ones (Barron and Barrett, 1996; Stewart, 1996; Durbach and Stewart, 2012; Durbach and Calder, 2015). Generally though, most studies have found that simple models that simplify the assessment of weights, value/utility functions, aggregation forms, and attribute values, often perform surprisingly well.

**4. Reconciling the Conclusions of the Various Research Strands**

In the preceding sections, we reviewed various strands of formal research that have investigated the performance of simple decision models. A motivation for this work is that there is not much communication among the strands and that they have not arrived at the same conclusions.

The conclusions can be summarized as follows. Research on the inference strand has put forth conditions under which simple models are more accurate than more complex ones, and some forecasting researchers have proposed that simple models should be preferred; and the preference strand has shown that simple models can approximate satisfactorily the more complex models.

Of course, these summaries are our own interpretations of others’ work and it may be that some researchers disagree with the phrasing we used. Even in this case, however, it seems clear that there is a tension between the conclusions of preference research and the conclusions of inference/forecasting research: A prevailing point in preference is that simple models may perform surprisingly well, but that they still perform worse than more complex models, or at most equally well with them. This conclusion fits with the common assumption that preference models such as MAUT should perform well because they are equivalent to a set of logical axioms such as transitivity, even though there is no compelling evidence that more conformity to axioms is linked to better performance (Clausing and Katsikopoulos, 2008; Arkes, Gigerenzer, and Hertwig, 2016). On the other hand, a prevailing point in inference/forecasting is that, at least under some conditions, simple models perform better than more complex ones. Model axiomatization is rare in inference/forecasting and is not used as an a-priori reason for expecting better performance.

Can these conclusions be reconciled? Yes, and here is how: By definition, there is an important difference between preference and inference/forecasting. In preference, the correct decision (i.e., the ground truth) is *not* objectively known or, one may argue, not even knowable; On the other hand, in inference/forecasting, the correct decision *is* objectively known, or at least knowable. Because of this difference, preference research has needed to assume a correct decision in order to evaluate models, whereas inference/forecasting research has not done so. More specifically, work on preference assumes a model of the data, which is typically complex, and evaluates the loss that simplifications of this model incur relative to it, whereas work on inference/forecasting compares simple and complex models on predicting the data directly.

In general, it is logically possible for a simple preference model to perform better than a complex preference model. But it has been *impossible* in actual preference research because of the joint effect of two conditions present in this research: First, the simple models were simplifications of the complex models, that is, they were always *nested* within the complex models; Second, the nested and full models were not evaluated in predicting unseen data, but in *fitting* known data.

In sum, there is no logical contradiction between the conclusions of the preference and inference/forecasting strands. In fact, when some researchers (Hogarth and Karelaia, 2005; Katsikopoulos, 2013) viewed inference problems as having a correct decision given by a complex model, they did find that simplifications of this model performed worse, or at most equally well, just as it is claimed in work on preference.

For all three strands, whether simple models perform worse or better than more complex models is not the properly formulated research question. The properly formulated question is to specify the conditions under which simple models perform better than more complex models, and vice versa. We discuss the theory available for providing answers in the next section.

**5. Explaining the Relative Performance of Simple and Complex Models**

**5.1. Repeated Operational Decisions**

A framework for explaining the relative performance of simple and more complex models is provided by the statistical theory of prediction, and in particular the *bias-variance* decomposition of prediction error (Friedman, 1997). There are different decompositions for different decision problems. In research on simple decision models, the bias-variance decomposition commonly used is the one which applies to forecasting (Geman, Bienenstock, and Doursat, 1992). In a somewhat less precise but very insightful way, the lessons learned have also been applied to inference (Gigerenzer and Brighton, 2007). To the extent that inference and preference can be viewed as formally similar as well (Hogarth and Karelaia, 2005)—for example, the actual value of a stock five years from now can be viewed as its utility, which in turn can be a linear function of the stock’s attributes as in multi-attribute value theory—the lessons should be transferable to preference as well. But because we see that this formal perspective may miss important conceptual differences between inference and preference, such as the constructive nature of preferences (Slovic, 1995) or their basis on personal values (Cherry, 1979), and of course the repeated versus one-off difference, we do not press this perspective here. Rather, we discuss preference in the next subsection. In what follows in this subsection, we sometimes implicitly refer to forecasting and other times to inference.

The bias-variance decomposition is a mathematical fact that says that the prediction error of any model is the sum of two terms. The first term is called bias and it measures how well, on the average, the model agrees with the objectively correct decision (i.e., ground truth). Complex models, which usually have many parameters and mathematically flexible functional forms, tend to have less bias than simple models, which usually have fewer parameters and straightforward functional forms. At a first approximation, this is so because when many parameters can be tweaked within a flexible form, they can combine so that agreement between model prediction and ground truth can also increase. For example, MAUT can achieve low bias, whereas the fast and frugal tree of Figure 1 has zero parameters and has relatively high bias.

But this is not the whole story. There is a second term, which contributes to a model’s total prediction error. This term is called variance. Variance measures the variation of model predictions around the model’s average prediction. Unlike the bias term, when it comes to the variance term, model complexity is less of a blessing and more of a curse. Complex models tend to have higher variance than simple models because their more parameters and more mathematically flexible form allow them to generate more distinct predictions (this is again at a first approximation).

For example, one can intuit why simple models tend to have lower variance than more complex models for small training set sizes. The smaller the training set, the more likely it is that sampling error and natural variations in the instances which are included in the training set will lead to variation in the parameter estimates of a given model. This variation can be expected to have an influence on the more flexible, heavily parameterized models to a greater degree than on the simpler rules. In an extreme case, a highly parameterized version of MAUT has relatively high variance, whereas the fast and frugal decision tree of Figure 1 has zero variance because it has zero parameters. The same logic also suggests that models which combine or select between simple heuristics on the basis of parameter estimation (e.g., using coefficients for weighting the relative contribution of different heuristics; Schurz and Thorn, 2016) will likely suffer from the bias-variance tradeoff to a similar extent with more complex models.

Because a model’s total prediction error is the sum of its bias and variance, one can see that the final result can go either way: Either a simple or a more complex model can have higher predictive accuracy in a particular dataset, depending on whether an advantage in variance is larger than an advantage in bias in this dataset. It has been argued that in practice variance may be more critical than bias (Brighton and Gigerenzer, 2015). This claim is consistent with Green and Armstrong’s (2015) review of the forecasting literature which concluded that all predictive evidence-based forecasting methods are simple. Nikolopoulos and Petropoulos (2017) also suggest that, to date, (complex) forecasting methods might be focusing too much on minimizing bias and overfitting the training data. Optimality and robustness are in general two different things (Bertsimas and Shim, 2004).

Furthermore, and quite surprisingly, it has been discovered that simple rules may also achieve competitive bias in practice. This happens when there exists an attribute, or an alternative option, which dominates the others. An attribute dominates other attributes when it is statistically much more informative about the value of options than other attributes are. For instance, time since last purchase predicts future sales much more accurately than customer age does (Wübben and Wangenheim, 2008). It has been analytically shown that lexicographic models which decide based on a dominant attribute, incur zero bias (Martignon and Hoffrage, 2002; Katsikopoulos, 2011). An alternative option dominates other options when its attribute values are better or equal to the attribute values of the other options. In this case, most models—simple or otherwise—incur zero bias. More interestingly, less restrictive definitions of dominance exist, and they have been shown to lead to zero bias for lexicographic models and tallying (Baucells et al., 2008). These results hold when utility is an additive or multi-linear function of the attributes (Katsikopoulos, Egozcue, and Garcia, 2014).

One may expect that dominant attributes and alternatives are rare. In fact, the opposite could be the case (Şimşek, 2013). Across 51 real datasets, it was found that dominant attributes exist in 93% of binary datasets (i.e., attributes had values of 1 or 0) and in 83% of the numeric datasets, and that dominant alternatives exist in 87% and 58% of binary and numeric datasets, respectively.

In sum, the conclusion of the theoretical work is that simple models tend to perform better than more complex models when (*i*) the information available is not of high quality or not ample enough to estimate the parameters of models reliably or (*ii*) there exists one attribute, or one alternative option, which dominates the others. On the other hand, when neither of conditions (*i*) or (*ii*) hold, more complex models tend to perform better than simple ones.

Condition (*i*) essentially says that a problem is difficult. Such difficulties may arise when a problem is dynamic or future developments are unpredictable. If (*i*) holds, an advantage in the variance component of the prediction error is much larger than the bias component, and simpler models have a very good chance of outperforming more complex models. An interesting interpretation of condition (*ii*) is that it says that the problem is easy, in the following sense. There either exists one alternative option which is better than all other options and the decision maker only needs to identify it; or there exists one attribute which is so important or informative that it suffices to only consult this attribute and, again, the decision maker only needs to identify it. If (*ii*) holds, as empirical research has shown that it could in practice, several simple models achieve zero bias, and thus can indeed outperform more complex models in predicting the future.

A corollary of the above analysis is that as one gathers more experience with a particular problem, the relative performance of simple and complex models might change. This is so because difficult problems might become easier—as when more high-quality information becomes available—and easy problems might become more difficult—as when new options are discovered which are neither dominating nor dominated by a previously dominating option.

**5.2. Strategic One-Off Decisions**

The theory available for explaining the relative performance of simple and more complex models for strategic decision making is less mathematical than the theory discussed just above for operational decision making, as well as richer and more nuanced than it. These differences are to be expected (for a convergent perspective, see also French, Maule, and Papamichail, 2009): First, it is not clear how to build a bias-variance-type of mathematical theory, like the one in Section 5.1., when decisions are not repeated. Second, the primary aim of preference modeling is not to predict—as is the case in inference/forecasting modeling—but to support decision making, a task which requires subtlety (Belton and Stewart, 2002).

Models supporting strategic decision making must confront the trade-off between creating more sophistication that addresses the discrepancies between model assumptions and real-world behaviour on the one hand, and retaining transparency, reflection, and insight on the other hand. In a multi-criteria decision analysis (MCDA) context, the analyst should be aware of the assumptions underlying the chosen MCDA model, put effort into verifying these assumptions, and most importantly inform the decision maker(s) of the effect these assumptions may have on the quality of decision outcomes.

Based on our own scanning of the literature on MCDA applications, we have found that neither laboratory studies nor case studies on the application of MCDA give authoritative and unambiguous guidance regarding the performance of MCDA tools in practice. Specific approaches or methods of MCDA are often strongly criticized on the basis of shortcomings or inappropriate assumptions, but typically by proposers of alternative approaches. It is no doubt true that no method can be free of criticism, but the practical impact of documented shortcomings in real world decision support is less well established. Consideration of practical impact is a critical question to pose, when alternative approaches addressing shortcomings tend often to involve more complex models—thus, hindering clarity of insight—or larger numbers of parameters which need to be elicited—thus, placing greater burdens on decision makers and introducing greater potential for error.

As such, it is not yet clear what are the conditions under which simple preference models—for example, those using rank-order centroid weights; adding single-attribute functions; or using one to three values per attribute—will satisfactorily approximate the full MCDA model. In our view, a promising avenue for developing such a theory would need, at the very least, to cover the following issues.

First, the decision makers’ *fully formed* preferences should be represented. By “fully formed” preferences, we mean preferences to which the decision maker(s) would agree after sufficient time and effort to explore all options (they could also be called “true” preferences). Although in practice true or fully formed preferences might not even exist, it is still meaningful to explore this set-up theoretically. Second, and very crucially, the potential costs of violating common MCDA assumptions, such as *additivity* and preferential *independence*, on satisfying these preferences, should be assessed. And finally it should be assessed whether such costs are acceptable to the decision maker(s).

Although other approaches are possible, simulation studies are one promising avenue for assessing and evaluating the impact of invalid modelling assumptions as well as errors made during preference elicitation. To be most useful, these studies need to focus on better understanding the conditions under which the impact of invalid assumptions are large, rather than averaging over simulated decision contexts, as it is typically done (Durbach and Stewart, 2012).

**6. Conclusion: An Aid to Help Determine When to Use Simple Models**

Based on the discussion in Section 5, here we put forth an aid to help determine if a simple model should be used, or not, for a particular type of decision problem (Table 1). It should be emphasized that this aid is not prescriptive but aims at promoting conversation among the stakeholders, aiming at reflection and insight.

|  |  |  |  |
| --- | --- | --- | --- |
| **Kinds of decisions**  | **Examples of simple models** | **Types of problems for which simple models should be used** | **Strength of evidence for use of simple models** |
| Repeated operational decisions(e.g., inference and forecasting) | Lexicographic models,fast and frugal trees, tallying, and exponential smoothing | Difficult problems(e.g., low-quality or scant information, dynamic or unpredictable situations) | Moderate |
| Easy problems(e.g., problems with dominant attributes or dominant options) | Strong |
| Strategic one-off decisions(e.g., preference/multi-criteria analysis) | Rank-order centroid weights; single-attribute functions; and use of one to three values per attribute | Problems where assumptions (e.g., additivity and independence) are satisfied | Moderate |
| Problems where assumptions (e.g., additivity and independence) are violated but costs are acceptable | Limited |

Table 1**:** An aid to help determine if a simple model be used, or not, for a particular type of decision problem, based on the discussion in Section 5. This aid is not prescriptive, but aims at promoting conversation among the stakeholders.

In this article, we analysed and synthesized evidence and argumentation for and against the use of simple models for inference, forecasting, and strategic decision making. The evidence does suggest that simple decision models give approximately the same, or even better, results as full models in “most” instances. Certainly anecdotal experiences suggest that in many cases a very small number of alternatives might nearly dominate, and/or that the effective decision choices are limited to a sufficiently small range of outcomes that constant trade-offs apply, both properties favouring simple models. Is there a role for decision aid or support then? Yes. In favourable cases, almost any sensible model or heuristic will work well; the real value of decision aid or support is precisely in cases when such favourable conditions do not apply. The challenge to the analyst is in knowing when these conditions do or do not apply. An appeal to averaging out is also not valid; care has to be taken in each case that effective consideration is given to the demands of all interests and stakeholders. Perhaps the main lesson to be gleaned is the following: It may be most effective to include a preliminary step in the decision process in order to identify what kind of decision problem is at hand. This seems to be consistent with Phillips’ (1984) principle of *requisite modelling*, where “requisite” means just the right level of simplification.

**References**

Abdellaoui, M. (2000). Parameter-Free Elicitation of Utilities and Probability Weighting Functions. *Management Science*, *46*(11), 1497–1512.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*(10), 1659–1674.

Ahn, B. S. (2011). Compatible weighting method with rank order centroid: Maximum entropy ordered weighted averaging approach. *European Journal of Operational Research*, *212*(3), 552–559.

Ahn, B. S., & Park, K. S. (2008). Comparing methods for multiattribute decision making with ordinal weights. *Computers and Operations Research*, *35*(5), 1660–1670.

Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K. V., Kothiyal, A., Murphy, E., and Neumann, T. (2014). Taking uncertainty seriously: Simplicity versus complexity in financial regulation, *Financial Stability Working Paper* No. 28, Bank of England, London, UK.

Arkes, H. R., Gigerenzer, G., and Hertwig, R. (2016). How bad is incoherence? *Decision*, *3*, 20-39.

Armstrong, J. S., Green, K. C., and Graefe, A. (2015). Golden rule of forecasting: Be conservative, *Journal of Business Research*, *68*, 1717-1731.

Barron, F. H., & Barrett, B. E. (1996). Decision quality using ranked attribute weights. *Management Science*, *42*(11), 1515–1523.

Barron, F. H., & Barrett, B. E. (1996). The efficacy of SMARTER — Simple Multi-Attribute Rating Technique Extended to Ranking. *Acta Psychologica*, *93*(1-3), 23–36.

Baucells, M., Carrasco J. A., and Hogarth, R. M. (2008). Cumulative dominance and heuristic performance in binary multi-attribute choice, *Operations Research*, *56*, 1289-1304.

Bell, D. E., Raiffa, H., and Tversky, A. (1988). Descriptive, normative, and prescriptive interactions in decision making, In D. E. Bell, H. Raiffa, and A. Tversky (Eds.), *Decision Making: Descriptive, Normative, and Prescriptive Interactions* (pp. 9-32). New York, NY: Cambridge University Press.

Belton, V., & Stewart, T. J. (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Boston: Kluwer Academic Publishers.

Bertsimas, D., and Sim, M. (2004). The price of robustness. *Operations Research*, *52*(1), 35–53.

Bleichrodt, H., Doctor, J. N., Filko, M., & Wakker, P. P. (2011). Utility independence of multiattribute utility theory is equivalent to standard sequence invariance of conjoint measurement. *Journal of Mathematical Psychology*, *55*(6), 451–456.

Bleichrodt, H., Pinto, J. L., & Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science*, 1498–1514.

Bleichrodt, H., Pinto, J. L., & Wakker, P. P. (2001). Using descriptive findings of prospect theory to improve the prescriptive use of expected utility. *Management Science*, *47*, 1498–1514.

Bleichrodt, H., Schmidt, U., & Zank, H. (2009). Additive Utility in Prospect Theory. *Management Science*, *55*(5), 863–873.

Breiman L, Friedman J. H., Olshen R. A., and C. J. Stone. (1984). *Classification and Regression Trees*, London, UK: Chapman and Hall.

Brighton, H., and Gigerenzer, G. (2015). The bias bias, *Journal of Business Research*, *68*, 1772-1784.

Clausing, D., and Katsikopoulos, K. V. (2008). Rationality in systems engineering: Beyond calculation or political action, *Systems Engineering*, *11*, 309-328.

Cherry, C. (1979). Human communication: Values, choice, and courage in a world of change, In C. R. Bell (Ed.) *Uncertain Outcomes* (pp. 79-92). New York, NY: Spectrum.

Corrente, S., Figueira, J. R., & Greco, S. (2014). The SMAA-PROMETHEE method. *European Journal of Operational Research*, *239*(2), 514–522.

Cutting, J. E. (2000). Accuracy, scope, and flexibility of models. *Journal of Mathematical Psychology*, *44*, 3–19.

Czerlinski, J., Gigerenzer, G., and Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, and the ABC Research Group, *Simple Heuristics that Make us Smart* (pp. 97-118). New York, NY: Oxford University Press.

Durbach, I. N. (2014). Outranking under uncertainty using scenarios. *European Journal of Operational Research*, *232*(1), 98–108.

Durbach, I. N., & Calder, J. M. (2016). Modelling uncertainty in stochastic multicriteria acceptability analysis. *Omega 64*, 13–23.

Durbach, I. N., & Stewart, T. J. (2009). Using expected values to simplify decision making under uncertainty. *Omega*, *37*(2), 312–330.

Durbach, I. N., & Stewart, T. J. (2012). A comparison of simplified value function approaches for treating uncertainty in multi-criteria decision analysis. *Omega*, *40*(4), 456–464.

Edwards, W., and Von Winterfeldt, D. (1986). *Decision Analysis and Behavioral Research*. Cambridge, UK: Cambridge University Press.

Fader, P. S., Hardie, B. G., and Huang, C. Y. (2004). A dynamic changepoint model for new product sales forecasting. *Marketing Science*, *23*(1), 50–65.

Fernandez, J. P., Katsikopoulos, K. V., and Shubitizde, F. (2010). Simple geometric heuristics for the detection of unexploded ordnance, *Working Paper*, Dartmouth College, Hanover, NH.

Fildes, R., and Petropoulos, F. (2015). Is there a golden rule? *Journal of Business Research*, *68*, 1742-1745.

French, S. (2013). Cynefin, statistics and decision analysis, *Journal of the Operational Research Society*, *64*, 547-561.

French, S., Maule, J., and Papamichail, N. (2009). *Decision Behaviour, Analysis and Support*. Cambridge, UK: Cambridge University Press.

Friedman, J. H. (1997). On bias, variance, 0/1-loss, and the curse-of- dimensionality, *Data Mining and Knowledge Discovery*, *1*, 55-77.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma, *Neural Computation*, *4*, 1-58.

Gigerenzer, G., and Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences, *Topics in Cognitive Science*, *1*, 107-143.

Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality, *Psychological review*, *103*, 650-669.

Goodwin, P. (2011). High on complexity, low on evidence: Are advanced forecasting methods always as good as they seem? *Foresight*, *23*, 10-12.

Graefe, A. (2015). Improving forecasts using equally weighted predictors, *Journal of Business Research*, *68*, 1792-1799.

Green, K. C., and Armstrong, J. S. (2015). Simple vs. complex forecasting: The evidence, *Journal of Business Research*, *68*, 1678-1685.

Hardie, B. G., Fader, P. S., & Wisniewski, M. (1998). An empirical comparison of new product trial forecasting models. *Journal of Forecasting*, *17*(34), 209-229.

Hogarth, R. M., and Karelaia, N. (2005). Simple models for multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes? *Management Science*, *51*, 1860-1872.

Hokkanen, J., Lahdelma, R., Miettinen, K., & Salminen, P. (1998). Determining the implementation order of a general plan by using a multicriteria method. *Journal of Multi-Criteria Decision Analysis*, 7(5), 273–284.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439-454.

Kadziński, M., & Tervonen, T. (2013). Stochastic ordinal regression for multiple criteria sorting problems. *Decision Support Systems*, *55*(1), 55–66.

Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice, *Decision Analysis*, *8*, 10-29.

Katsikopoulos, K. V. (2013). Why do simple heuristics perform well in choices with binary attributes? *Decision Analysis*, *10*, 327-340.

Katsikopoulos, K. V., Egozcue, M., and Garcia, L. F. (2014). Cumulative dominance in multi-attribute choice: Benefits and limits, *EURO Journal on Decision Processes*, *1*, 153-163.

Katsikopoulos, K. V., Schooler, L. J., and Hertwig, R. (2010). The robust beauty of ordinary information, *Psychological Review*, *117*, 1259–1266.

Katsikopoulos, K. V., and Syntetos, A. A. (2016). Bias-variance trade-offs in demand forecasting, *Foresight*, *40*, 12-19.

Keeney, R. L., and Raiffa, H. (1976). *Decision-making with Multiple Objectives: Preferences and Value Tradeoffs*, New York, NY: John Wiley and Sons.

Keller, N., and Katsikopoulos, K. V. (2016). On the role of psychological heuristics in operational research; and a demonstration in military stability operations, *European Journal of Operational Research*, *249*, 1063-1073.

Kirkwood, C. (1992). Estimating the impact of uncertainty on deterministic multiattribute evaluation. *Management Science*, *38*(6), 819–826.

Kirkwood, C. W., & Corner, J. L. (1993). The Effectiveness of Partial Information about Attribute Weights for Ranking Alternatives in Multiattribute Decision Making. *Organizational Behavior and Human Decision Processes*.

Kolassa, S. (2016). Sometimes it’s better to be simple than correct, *Foresight*, *4*0, 20-26.

Laeven, L., and Valencia, F. (2010). Resolution of banking crises: The good, the bad and the ugly”, *IMF Working Paper* 10/146.

Lahdelma, R., & Salminen, P. (2001). SMAA-2: Stochasitc multicriteria acceptability analysis for group decision making. *Operations Research*, *49*(3), 444–454.

Lovie, A. D., and Lovie, P. (1986). The flat maximum effect and linear scoring models for prediction, *Journal of Forecasting*, *5*, 159-168.

Makridakis, S., Hibon, M., and Moser, C. (1979). Accuracy of forecasting: An empirical investigation, *Journal of the Royal Statistical Society*: *Series A*, 97-145.

Makridakis, S., and Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International journal of forecasting*, *16*, 451-476.

Makridakis, S., Hogarth, R. M., and Gaba, A. (2009). Forecasting and uncertainty in the economics and business world, *International Journal of Forecasting*, *25*, 794-812.

Makridakis, S., Wheelwright, S. G., and Hyndman, R. J. (1998). *Forecasting*: *Methods and Applications*, New York, NY: John Wiley and Sons.

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

Martignon, L., Katsikopoulos, K. V., and Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics, *Journal of Mathematical Psychology*, *52*, 352-361.

Nikolopoulos, K., Goodwin, P., Patelis, A., and Assimakopoulos, V. (2007). Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches, *European Journal of Operational Research*, *180*, 354-368.

Nikolopoulos, K. and Petropoulos, F. (2017). Forecasting for big data: Does suboptimality matter? *Computers and Operations Research*, http://dx.doi.org/10.1016/j.cor.2017.05.007

O’Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., … Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts’ Probabilities*. John Wiley & Sons.

Phillips, L. D. (1984). A theory of requisite decision models. *Acta Psychologica*, *56*(1), 29-48.

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega*, *64*, 126-130.

Rosenhead, J., and Mingers, J. (Eds.) (2001). *Rational Analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity, Uncertainty and Conflict*, New York, NY: John Wiley and Sons.

Roy, B. (1985). *Methodologie Multicritere d’Aide à la Decision*. Paris: Economica.

Saaty, T. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Schurz, G., and Thorn, P. D. (2016). The revenge of ecological rationality: Strategy-selection by meta-induction within changing environments. *Minds and Machines*, *26*, 31-59.

Şimşek, Ö. (2013). Linear decision rule as aspiration for simple decision heuristics, *Advances in Neural Information Processing Systems*, *26*, 2904-2912.

Şimşek, Ö., and Buckmann, M. (2015). Learning from small samples: An analysis of simple decision heuristics, *Advances in Neural Information Processing Systems*, *28*, 3141-3149.

Slovic, P. (1995). The construction of preferences, *American Psychologist*, *50*, 364-371.

Smith, J. E., and von Winterfeldt, D. (2004). Anniversary article: Decision analysis in *Management Science*, *Management Science*, *50*, 561-574.

Stewart, T. J. (1995). Simplified Approaches for Multicriteria Decision Making under Uncertainty. *Journal of Multi-Criteria Decision Analysis*, *4*, 246–258.

Stewart, T. J. (1996). Robustness of Additive Value Function Methods in MCDM. *Journal of Multi-Criteria Decision Analysis*, *5*, 301–309.

Syntetos, A. A., Boylan, J. E., and Croston, J. D. (2005). On the categorization of demand patterns, *Journal of the Operational Research Society*, *56*, 495-503.

Tetlock, P. E. (2006). *Expert Political Judgment*: *How Good Is It? How Can We Know*? Princeton, NJ: Princeton University Press.

Thomas, L. C., Edelman, D. B., and Crook, J. N. (2002). *Credit Scoring And Its Applications*, Philadelphia, PA: Society for Industrial and Applied Mathematics.

Todd, P. M. 2007. How much information do we need? *European Journal of Operational Research*, *177*, 1317-1332.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

Vincke, P. (1999). Outranking Approach. In T. Gal, T. Stewart, & T. Hanne (Eds.), *Multicriteria Decision Making: Advances in {MCDM} Models, Algorithms, Theory, and Applications*. Dordrecht: Kluwer Academic Publishers.

Wübben, M., and Wangenheim, F. V. (2008). Instant customer base analysis: Managerial heuristics often “get it right”, *Journal of Marketing*, *72*, 82-93.

1. Fader et al. (2004) demonstrate the superiority of their approach using two of the 19 datasets that were available to them (Hardie, Fader, and Wisniewski, 1998). The most likely reason for this selection would seem to be that their approach requires a fairly large number of repurchases in the calibration period, requiring a lengthier trial or more frequently bought products than in the remaining 17 datasets (Fader et al., 2004, p. 61). [↑](#footnote-ref-1)