



December 1994

Judgmental Decomposition: When Does It Work?

Donald G. MacGregor
Decision Research

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/marketing_papers

Recommended Citation

MacGregor, D. G., & Armstrong, J. S. (1994). Judgmental Decomposition: When Does It Work?. Retrieved from http://repository.upenn.edu/marketing_papers/132

Postprint version. Published in *International Journal of Forecasting*, Volume 10, Issue 4, December 1994, pages 495-506.
Publisher URL: [http://dx.doi.org/10.1016/0169-2070\(94\)90018-3](http://dx.doi.org/10.1016/0169-2070(94)90018-3)

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/marketing_papers/132
For more information, please contact libraryrepository@pobox.upenn.edu.

Judgmental Decomposition: When Does It Work?

Abstract

We hypothesized that multiplicative decomposition would improve accuracy only in certain conditions. In particular, we expected it to help for problems involving extreme and uncertain values. We first reanalyzed results from two published studies. Decomposition improved accuracy for nine problems that involved extreme and uncertain values, but for six problems with target values that were not extreme and uncertain, decomposition was not more accurate. Next, we conducted experiments involving 10 problems with 280 subjects making 1,078 estimates. As hypothesized, decomposition improved accuracy when the problem involved the estimation of extreme and uncertain values. Otherwise, decomposition often produced less accurate prediction.

Keywords

decision analysis, estimation, extreme values, forecasting, multiplicative

Comments

Postprint version. Published in *International Journal of Forecasting*, Volume 10, Issue 4, December 1994, pages 495-506.

Publisher URL: [http://dx.doi.org/10.1016/0169-2070\(94\)90018-3](http://dx.doi.org/10.1016/0169-2070(94)90018-3)

Judgmental Decomposition: When Does It Work?

Donald G. MacGregor
Decision Research, Eugene, OR

J. Scott Armstrong
The Wharton School, University of Pennsylvania, Philadelphia, PA

Abstract

We hypothesized that multiplicative decomposition would improve accuracy only in certain conditions. In particular, we expected it to help for problems involving extreme and uncertain values. We first reanalyzed results from two published studies. Decomposition improved accuracy for nine problems that involved extreme and uncertain values, but for six problems with target values that were not extreme and uncertain, decomposition was not more accurate. Next, we conducted experiments involving 10 problems with 280 subjects making 1078 estimates. As hypothesized, decomposition improved accuracy when the problem involved the estimation of extreme and uncertain values. Otherwise, decomposition often produced less accurate predictions.

Keywords: Decision Analysis; Estimation; Extreme Values; Forecasting; Multiplicative Decomposition; Uncertainty

I. Introduction

Consider the following question: What is the estimated yearly circulation of a proposed new magazine on raising exotic animals? People are likely to respond that they have no idea. But do they? What are they likely to say if asked whether the number was greater than 100 million? Would they say that it is less than 1000? Most likely, people would say that the true value is somewhere between these two values. Obviously, they know more than they think they do when first asked.

How well a person is able to forecast a quantity is related to the relevant information that they have at their disposal, either from information sources or from experts. It is also a function of whether they can break the problem into parts so that they can use their information effectively. Forecasters frequently break a problem into parts, make forecasts from each part, then recombine the separate forecasts to make a forecast of the target value. In 1968 Howard Raiffa (1968) claimed that such a procedure, *decomposition*, is 'the spirit of decision analysis.' Since then, research has seemed to support the view that decomposition is a useful strategy with wide applicability and little risk.

Prior literature on judgmental decomposition (Armstrong et al., 1975, and MacGregor et al., 1988) concluded that decomposition would be especially effective for problems involving uncertain values. However, we do not know much about the conditions under which judgmental decomposition is most useful. Armstrong et al. (1975) had suggested that the scale of the problem might make further study worthwhile, and our paper addresses that issue. In examining the problem, we reanalyzed results from two studies. In addition, we, conducted experiments with new subjects. We also examined alternative approaches for assessing uncertainty to determine whether they would yield different recommendations about when decomposition is appropriate.

2. Hypotheses

The basic idea behind decomposition is simple. Given a target quantity that is difficult to estimate, one breaks the problem down into subproblems that are easier to estimate. The difficulty lies in translating this idea into practice. For decomposition to be done successfully, certain conditions are desirable. First, the target value should

be *one* that is difficult to estimate. Second, estimation errors for each part should be less, relatively speaking, than the errors for estimating the target value. Third, estimation errors for the parts should not have strong positive correlations between one another. Negatively correlated errors are desirable so that one has offsetting errors. These conditions are not easy to specify in operational terms.

Traditionally, the term decomposition has been used to refer to the practice of breaking a problem into multiplicative elements. An additive breakdown is usually referred to as *disaggregation or segmentation*. Our paper is restricted to multiplicative decomposition and we use the term decomposition to refer to this.

Decomposition is often viewed as a safe strategy. Rather than putting all of one's eggs into a single basket, estimates are provided separately. Errors in one element may compensate for errors in another. However, when errors are positively correlated, they can be explosive. For example, if two components are in the same direction and are each equal to 20%, this would translate into an error of 44% in the target value ($1.2 \times 1.2 = 1.44$).

Target values with extreme values are likely to create difficulties for subjects unless these numbers are well known. For very large numbers, people might make estimates that are too small. Lacking good intuition, an estimator might assign a 'more reasonable number' to a quantity in question. We would expect the converse for very small numbers, such as 'one in 10 million.'

We hypothesized that decomposition would improve accuracy for problems with extreme values when subjects were highly uncertain about the target value. The reasoning is simply that large numbers are confusing to many people. With decomposition, the analyst might be able to avoid the extreme numbers associated with high uncertainty. Uncertainty is an important aspect of this hypothesis. Thus, we do not expect that decomposition would help to estimate well known numbers, such as the distance from the Earth to the sun (when most of the experts believe that the distance is about 93 million miles).

The operational definition of an extreme value is difficult to determine. To provide a simple measure of an extreme value, we initially defined it as any number having more than seven digits (equal to or greater than 10 million). Certainly, many people have difficulty grasping numbers of this magnitude. For example, a book has been written¹ with the sole purpose of helping people to understand the magnitude of one million. It consists of one million dots with comparisons at various points where examples are given (Hertzberg, 1970).¹ Psychologists also refer to the ability of the human mind to handle only seven things (plus or minus two).

The selection of the unit of measure causes problems. For example, one could change the units from miles to inches when asking someone to estimate the distance from New York to San Francisco. However, some important quantities are not amenable, either conceptually or computationally, to changes in scale.

We were also concerned about how best to assess uncertainty. In particular, would different approaches lead to different conclusions about when to use decomposition?

3. Reanalysis of prior studies

In an early study of judgmental decomposition, Armstrong et al. (1975) concluded that multiplicative decomposition typically improves accuracy and is unlikely to reduce accuracy. The study involved such problems as estimating the number of packs of Polaroid film that were used in the United States in 1970. The results also supported the hypothesis that decomposition is especially useful for problems where the estimator's perceived uncertainty about the true value is high. A subsequent study by MacGregor et al., (1988) also found that judgmental decomposition improves accuracy. That study used similar problems, for example, estimating the value of imported passenger cars sold in the U.S. the previous year.

¹ As an example of how difficult it is to think about extreme numbers, consider the following. A typographical error was made in Armstrong et al. (1975). The number of cards saying "Carefree Sugarless Gum" that were sent to a Philadelphia radio station was reported as 66.5 billion rather than the correct value, which was 66.5 million. We missed this in proofreading, and the number has subsequently been cited in other papers without any questions being raised.

Armstrong et al. (1975) examined uncertainty by asking 151 subjects to rank problems according to the confidence that they had in their ability to provide accurate answers. MacGregor et al. (1988) addressed the same issue by using the variability among 45 subjects in their estimates for each target value. Specifically, they focused on the interquartile range. The interquartile range represents the middle 50% of a distribution and is calculated as the difference between the point at the 75th percentile of the distribution (Q3) and the point at the 25th percentile (Q1); the median of the distribution is at the 50th percentile (Q2). We expected that problems with extreme unknown values would create uncertainty among estimators and would therefore show up in the interquartile range. We examined this hypothesis by comparing the number of digits in each of the 16 problems in MacGregor et al. with the interquartile range of error ratios. As expected, the number of digits was related to uncertainty. The correlation between the number of digits in the actual values for each problem and the corresponding interquartile range was about +0.75.

To examine whether decomposition improved accuracy for problems involving extreme unknown numbers, we split the MacGregor et al. data according to magnitude and disagreement. This yielded five problems where scale was not extreme (using seven or fewer digits gave a roughly equal breakdown of the problems) and where assessors were in agreement (we used an interquartile range with a log I o of 1.3 or less, which means that the ratio between the lowest quartile and the highest quartile is less than two). The five problems were the numbers of physicians, marriages, alcoholics, university employees and hospital employees. Six problems had extreme magnitude (over seven digits) and high disagreement among estimators (interquartile range of 1.75 or more, implying a ratio of 5.6 of the largest to smallest quartile). These problems involved the numbers of welfare cases, imported cars, alcohol dollars, mail handled by post offices, gasoline and cigarettes.

We estimated the average improvement for decomposition in MacGregor et al. in two steps. First, geometric mean estimates were calculated for the group of subjects who used the decomposed version (this being the computed full algorithm from Table 6 in MacGregor et al.) and for those who used the global version. These estimates were then compared with the actual values for each problem.

Table 1
Decomposition versus global errors: reanalysis of prior studies

Conditions	Number of problems	Median error ratios		
		Global	Decomposition	Error reduction
<i>Not extreme, low uncertainty</i>				
MacGregor et al.	5	1.8	2.3	-0.5
Armstrong et al.	1	5.4	2.3	2.1
<i>Extreme, high uncertainty</i>				
MacGregor et al.	6	99.3	3.0	96.3
Armstrong et al.	3	18.0	5.7	12.3

Decomposition errors were smaller than global errors for each of the six problems where disagreement (interquartile range) was high and the actual values were extreme. Subjects who made global estimates were in error by a factor of 99.3 (9930%) on average. In contrast, the error ratio for the decomposed version for the same six problems was 3.0, or 300%.² Thus, the median error was reduced by a factor of 96.3 (see bottom part of Table 1). For problems without extreme values and where disagreement was low, decomposition yielded less accurate estimates, as its error was 50% higher than that for the global approach. Table 1 summarizes these results.

We did a similar analysis for the problems in Armstrong et al. (1975). Here, the analysis was based on individuals rather than groups. Error ratios were calculated for each subject's estimate for each problem by comparing their estimates with the actual values. The median error ratio was then obtained for each problem. Decomposition produced substantial gains (1230% error reduction) for the three extreme problems with the highest

² We calculated the geometric means of the two error ratios in the middle of the distribution for the global and decompositional conditions.

uncertainty.³ Decomposition also provided a lesser improvement for the one problem that did not involve an extreme number. Table 1 summarizes these results as well.

Averaging across the two studies (weighting according to the number of questions), decomposition reduced error by a ratio of 68.3 for the nine problems involving extreme uncertain values. However, decomposition had no overall effect for the other six problems.

4. An experiment on the effects of extreme uncertain values

We conducted an experiment to provide further evidence on the effects of multiplicative decomposition when applied to problems with extreme uncertain numbers. This section describes the problems and the subjects.

4.1. Problems

We selected problems in which the magnitude of unknown numbers to be estimated varied. Our extreme problems had seven or more digits, ranging in value from 3 540 940 to 4 243 000 000. As noted earlier, this definition of *extreme* is somewhat arbitrary.⁴ Not extreme numbers in this set of problems had four digits or less, in order to provide a marked distinction from extreme numbers. Table 2 provides the 10 problems, along with, the correct answers taken from almanacs and fact books.

Table 2
Problems and magnitudes: versus actual versus estimated

Problem	Correct answer	Number of digits	
		Actual	Upper quartile estimate
<i>Not extreme magnitude</i>			
Circumference of a \$.50 coin in inches	3.71	1	1
U.S. Presidents	41	2	2
Argentine immigrants to U.S. (annually)	2,800	4	5
Bank failures in 1993	4,004	4	4
<i>Extreme magnitude</i>			
Area of U.S. in square miles	3,540,940	7	7
Circulation of TV Guide	16,900,000	8	6
Pairs of athletic shoes made per year	23,400,000	8	7
Auto accidents per year	24,100,000	8	8
Pairs of men's pants made per year	124,000,000	9	9
Bushels of wheat produced in world per year	4,243,000,000	10	10

All questions relate to the U.S. unless stated otherwise.

Because actual values would not be known to the subjects, we first determined whether it would be possible to identify problems that might involve extreme values. We reasoned that typical subjects would not do well at such estimates. Thus, we used the geometric mean of the upper quartile (top 25%) of the estimates. That is, if the upper quartile of subjects expected this to be an extreme number, then it was treated as such. By this measure, the expected number of digits was a good match of the actual number of digits, as shown in Table 2. The largest estimate for the small group was that Argentine immigrants would be a five-digit number, and the smallest estimate

³ We used the median error ratios across the new groups of subjects that were tested for the film and tobacco problems. Only two groups did the Contest problem, and here we used the geometric mean.

⁴ After analyzing the prior research (Table 1), we revised our definition of extreme for this study from 'more than seven digits' to 'seven or more digits.' Extremity could also be defined in terms of small numbers. An example would be, 'What is the chance that a person in the U.S. will die next year because of botulism?' (The answer is 1/100,000,000.)

for the extreme problems was that the Circulation of TV *Guide* would be a six-digit number, so the classification of the problems was the same.

To determine whether the large target values were uncertain, we examined the interquartile ranges. The smallest of these ranges for the group of problems having extreme values indicated that the upper quartile mean was more than 10 times as large as the lower quartile mean.

For each problem we constructed a global version and a decomposed version. Table 3 summarizes the full set of 10 decomposed algorithms. For the sake of brevity, only the algorithm steps requiring subjects to make component estimates are provided; intermediate arithmetic steps are omitted. We also asked subjects to rate their knowledge about each target value, their expected accuracy and the probability that their answer would be within 10% of the true value.

Table 3
Abbreviated descriptions of algorithms for the ten estimation problems

<p><i>Argentine immigrants</i></p> <ul style="list-style-type: none"> • Population in the U.S. • Proportion of population that immigrated to U.S. • Percentage of U.S immigrants from Argentina 	<p><i>Men's pants</i></p> <ul style="list-style-type: none"> • Number of men in the U.S. • Number of pairs of pants the average man buys each year • Number of women in the U.S. • Number of pairs of men's pants the average woman buys each year • Proportion of men's pants manufactured in the U.S. that are sold to U.S. customers
<p><i>Circulation of TV Guide</i></p> <ul style="list-style-type: none"> • Households in the U.S. • Proportion of households with a TV • Proportion of households with a TV receiving TV Guide 	<p><i>Athletic shoes</i></p> <ul style="list-style-type: none"> • Population of the U.S. • Proportion of the population that wears athletic shoes • Pairs of athletic shoes each wearer buys per year • Proportion of athletic shoes manufactured in U.S. that are sold to U.S. customers
<p><i>Circumference of 50¢ coin</i></p> <ul style="list-style-type: none"> • Diameter in inches of a 50¢ coin • Number of pieces of string the length of the diameter needed to wrap around circumference 	<p><i>Auto accidents</i></p> <ul style="list-style-type: none"> • Number of people in the U.S. of driving age • Proportion of people of driving age who drive • Number of accidents the average driver has per year
<p><i>Bushels of wheat</i></p> <ul style="list-style-type: none"> • Population of the world • Number of bushels of wheat consumed per person per year • Proportion of wheat wasted per year 	<p><i>Area of U.S.</i></p> <ul style="list-style-type: none"> • Distance in miles from San Francisco, CA to Washington, D.C. • Distance in miles from San Diego, CA to Seattle, WA • Proportion of the U.S. that would fit into a rectangle with an area equal to the product of the above dimensions.
<p><i>Bank failures in 1933</i></p> <ul style="list-style-type: none"> • Current population of the U.S. • Population of U.S. in 1933 as a proportion of current population • Number of customers for a typical bank • Proportion of banks failed in 1933 	
<p><i>U.S. presidents</i></p> <ul style="list-style-type: none"> • Number of years U.S. has had presidents • Number of years the average president holds office 	

For some of the problems, such as *Athletic shoes*, one of the components involved an extreme value. However, we were reasonably confident that subjects would know this value. Also, data on these values are readily available so that one could insert the known value.

4.2. Subjects

Subjects for the experiment were individuals who answered advertisements in the University of Oregon daily newspaper. The advertisements called for participation in judgment and decision-making tasks. Two hundred and eighty individuals participated in the experiment, which was conducted in two sessions. Subjects were randomly assigned to either the global or the decomposition treatment. In the first session, the problems *\$.50 coin*, *U.S.*

presidents, Argentine immigrants, Bank failures, Circulation of TV Guide and Bushels of wheat were administered. Those subjects assigned to the global treatment received all six problems. Because of time constraints, subjects assigned to the decomposition condition received half of the problems. In the second session, the remaining four problems were administered. Again, subjects in the global condition received all four of the remaining problems, while decomposition subjects received half of the problems.

5. Results

As had been done in previous studies of judgmental forecasting (Armstrong et al., 1975, and MacGregor et al., 1988), we used the error ratio as an index of accuracy. The error ratio is computed as the ratio of the individual's estimated value to the correct answer, or the reverse, such that the result is greater than or equal to 1.0. Estimates for a given problem were summarized across subjects by computing the geometric mean of the error ratios.

We had hypothesized that decomposition would improve accuracy for problems having extreme uncertain values. The results, shown in Table 4, were consistent with this hypothesis. We summarized the problems into two groups: extreme problems (correct answer greater than 3,540,940) and not extreme problems (correct answer 4,004 or less). Accuracy was superior for decomposition in five of the six extreme problems, with an error reduction that ranged from a factor of 4.10 (*Athletic shoes*) to 91.47 (*Auto accidents*). Only the *Circulation of TV Guide* problem suffered a decrease in accuracy with decomposition. This decrease was modest compared to the gains in accuracy for the other five extreme problems, and this decrease was not statistically significant. Across all six problems, the median error was reduced by a factor of 19.78, approximately a 20-fold improvement in accuracy. Following Winer's method of adding is (as described in Rosenthal, 1978), these results were statistically significant at $p < 0.001$ using a one-tail test.⁵

Table 4
Error ratios for global versus decomposed estimates (for individuals)

Problems	Sample size		Error ratios (geometric means)		Error reduction	<i>t</i> -test
	Global	Decomp	Global	Decomp		
<i>Not extreme</i>						
\$.50 coin	64	62	1.82	1.41	0.41	4.07**
U.S. presidents	64	63	1.23	1.35	- 0.12	-1.55
Argentine immigrants	65	54	4.89	46.77	-41.88	-5.85**
Bank failures	64	57	10.45	19.50	- 9.03	-1.69
Median					- 4.58	
Combined experiments (<i>z</i> -test)						-2.49*
<i>Extreme</i>						
Area of U.S.	30	30	33.88	1.70	32.18	6.00**
<i>Circulation of TV Guide</i>	64	60	7.76	10.96	- 3.20	-1.11
Athletic shoes	31	32	19.95	15.85	4.10	0.47
Auto accidents	31	30	93.33	1.86	91.47	8.07**
Men's pants	31	31	17.38	10.00	7.38	1.01
Bushel of wheat	61	62	45.71	6.92	38.79	4.57**
Median					19.78	
Combined experiments (<i>z</i> -test)						4.37**

*Significant at $p < 0.05$

**Significant at $p < 0.001$

By contrast, accuracy for not extreme problems was reduced with decomposition. Error Auction values for three of the four not extreme problems were negative, indicating a superiority of global estimation over

⁵ Because the ratios involved some extreme values, the *t*-tests were done on the logs of the error ratios rather than on the ratios themselves.

decomposition. Decomposition increased the median error of these problems by 458%, an increase that is statistically significant at $p < 0.05$. The test of the not extreme values was two-tailed because we had no directional hypothesis. Our analysis overstates the statistical significance; the various estimates are not completely independent of one another.

5.1. Uncertainty of estimation

Whether decomposition is appropriate depends on some measure of uncertainty. We propose that analysts first determine whether the problem is subject to much uncertainty. If so, decomposition may be appropriate, especially if one can structure the problem to avoid extreme certain values.

Otherwise, global estimates should be used. Uncertainty decreases the degree to which an estimate from various assessors exhibits a lower variance or a reduced range. Table 5 shows the interquartile ranges for the global and decomposed estimates. The entries consist of the logs of Q1 and Q3 as well as their differences. Q1 corresponds to the 25th percentile of the distribution, while Q3 corresponds to the 75th percentile. If decomposition reduces uncertainty, then a lower Q3-Q1 difference should result. Computed in this way, the differences in Table 5 can be interpreted as the number of digits by which the estimates of Q1 and Q3 differed.

Table 5
Analysis of interquartile ranges

Problems	Global			Decomposed		
	Log Q3	Log Q1	Differences (Q3-Q1)	Log Q3	Log Q1	Differences (Q3-Q1)
Not extreme						
\$.50 coin	0.48	0.20	0.28	0.63	0.42	0.21
U.S. presidents	1.71	1.59	0.12	1.70	1.53	0.17
Argentine immigrants	4.30	3.30	1.00	5.74	3.60	2.14
Bank failures	3.70	2.08	1.62	4.92	3.18	1.74
Extreme						
Area of U.S.	6.30	4.00	2.30	6.76	6.39	0.37
Circulation of TV Guide	7.54	6.18	1.36	7.80	5.95	1.85
Athletic shoes	8.00	6.00	2.00	8.84	7.75	1.09
Auto accidents	6.18	5.00	1.18	7.80	6.08	1.72
Men's pants	8.00	6.00	2.00	9.53	8.07	1.46
Bushels of wheat	10.48	7.18	3.30	10.54	9.65	0.89

For not extreme problems, the interquartile ranges are higher for the decomposed estimates than the global estimates for three of the four problems. For one problem, *Argentine immigrants*, the interquartile range for the decomposed version was higher than that for the global (2.14 versus 1.00). This occurred even though each part had the same interquartile range as the target value. This problem did not, then, meet the condition that the parts are easier to forecast than the target value, nor were the errors independent. Thus, it is not surprising that decomposition was not helpful for this problem.

For extreme problems, the range for the decomposed estimate was less than that for the global, except for the *Auto accidents* and *Circulation of TV Guide* problems. In other words, decomposition often improved confidence for difficult problems when the agreement among assessors' estimates was used to gauge confidence. Furthermore, the differences between the global and decomposed ranges for the four problems with improvements were substantial, being typically greater than one digit. Although the number of problems is not sufficient to assess the relationship between the interquartile ranges and errors, this result is consistent with that found in the seven problems examined by Aschenbrenner and Kasubek (1978).

A tenet of decomposition states that the parts of a problem are more tractable than the whole. This means that uncertainty in the estimates of a problem's components should be lower than that for the global estimate. We computed the interquartile ranges for each of the components of the six problems in Table 6. The parts were easier

to estimate than the target value for three problems: *50 ¢ coin*, *U.S. presidents* and *Bushels of wheat*. The first two of these had target values that were easy to assess directly, whereas *Bushels of wheat* had an extreme value that was difficult to measure. The *Bushels of wheat* problem met all conditions for decomposition. As expected, decomposition was successful for this problem. Conversely, decomposition was less accurate for four of the other five questions.

Table 6
Assessments of subjective confidence

Problems	Mean knowledge ratings ^a		Mean accuracy ratings ^a		Mean probability ratings that estimate is within 10% of true answer	
	Global	Decomposition	Global	Decomposition	Global	Decomposition
	U.S. Presidents	6.16	5.46	6.02	5.38	64.4
\$.50 coin	5.50	4.45	5.61	4.45	54.9	55.8
Circulation/TV <i>Guide</i>	3.40	2.35	3.46	2.58	32.1	24.6
<i>Bank failures</i>	3.19	2.17	3.06	2.20	28.0	18.9
Argentinie immigrants	2.15	1.81	2.38	2.32	24.4	16.8
Bushels of wheat	2.24	2.02	2.16	2.27	18.9	19.9

^a High scores imply greater knowledge and greater perceived accuracy (scale from 1 to 10).

5.2. Subjective confidence ratings

A second source of uncertainty estimates is the subjective confidence that forecasters have in their knowledge about a problem. We addressed three questions with respect to subjective uncertainty. (1) Do alternative measures of uncertainty yield similar recommendations? If yes, then we could use the least expensive approach to assessing uncertainty. (2) Are judges more confident when they make decomposed estimates or global estimates? (3) Does decomposition lead subjects to become better calibrated about their confidence?

As the simplest and least expensive approach, we asked subjects to provide judgments of their knowledge about each target value, and the degree to which they thought their estimate would be accurate. Self-ratings of knowledge and accuracy were obtained from the subjects before they made their estimates by using the following scales.

“Before you begin, indicate on the scale below *how much you think you know about the topic*”
(1= know very little; 10 = know a great deal).

“How *accurately* do you think you will be able to estimate this quantity?”
(1= low accuracy; 10 = high accuracy).

Judgments were obtained for a subset of six problems. Table 6 shows alternative assessments of accuracy for these problems.

After subjects had estimated the value for each of the six problems, we asked them to indicate the probability that their estimate was within 10% of the correct answer. These results are also presented in Table 6. Finally, we calculated the interquartile ranges of the global estimate for each problem, shown in the last column of Table 6.

With the exception of the interquartile range, the different approaches to subjective confidence produced similar results. The intercorrelations among the three measures across the six problems were all over 0.99. Given the close correspondence among the three measures, they were expected to be of roughly equal value in deciding when to use decomposition.

We applied the same procedures to subjects who received the decomposed versions of the problems. Across all six problems, subjects had higher self-ratings of problem knowledge in the global condition than in the decomposition condition. Because subjects in the decomposition condition received more than one estimation problem, their self-ratings of problem knowledge may have been influenced by the difficulties they experienced

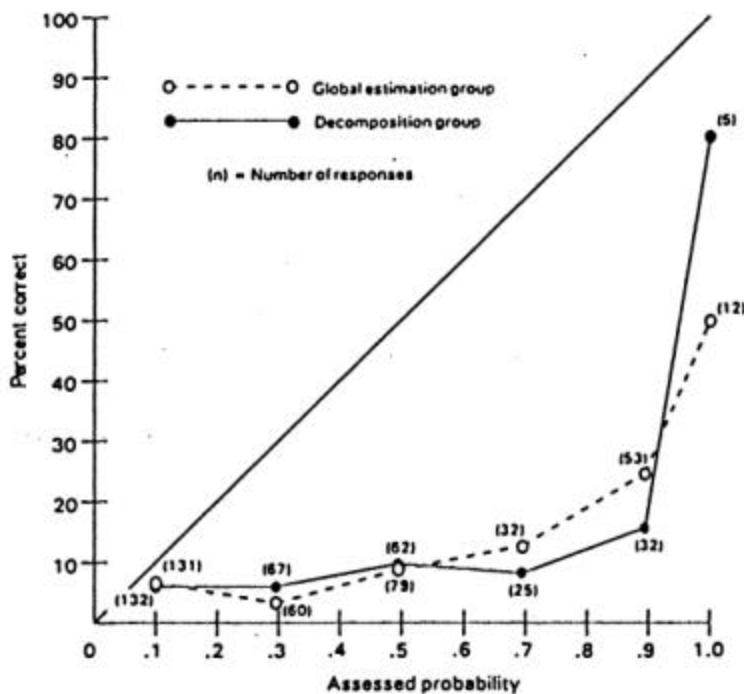
with the complexity of the problem. This was also the case for self-ratings of accuracy, except for the *Bushels of wheat* problem. Similar results were obtained when we asked the questions about confidence after subjects had completed their estimates. In other words, the different assessments each led to the conclusion that subjects in the decomposition condition thought that the problems were more difficult than did subjects in the global estimation condition. These results agree with the findings of Sniezek et al. (1990), who had concluded that the increased processing (for decomposed problems) leads to a reduction in confidence. In retrospect, it might have been better for us to have asked for estimates of the difficulty for each of the parts. Henrion et al. (1993) did this, and their subjects reported that the *components* were easier to estimate than the global value.

Are subjects better calibrated when they use decomposition? Probability assessments are said to be externally calibrated if, for a given probability assessment (e.g., 0.6), exactly that proportion (e.g., 60%) turn out to be correct. We summarized the calibration results for global and decomposed estimates, across all ten problems. Mean probability assessments were generally higher than the proportion correct for both approaches, indicating overconfidence. On average, those making global estimates expected 38.9% of their answers to be within 10% of the true value, but only 10.9% were that accurate. Those using the decomposed approach expected 32.6% of their estimates to be within 10% of the true value, but only 9.0% were that accurate. In effect, decomposition reduced overconfidence from 28.0% in the global case to 23.6% for decomposition, with the largest reduction occurring in those situations where subjects felt most confident, as shown in Fig. 1.

5.3. Limitations

Two of the four problems in the not extreme version (*Argentine immigrants* and *Bank failures*) involved elements with extreme values. Because each of the components had an element dealing with the U.S. population, we assumed that the subjects would be familiar with these values. To examine this assumption, we analyzed the population estimates for each of the problems. The median population estimate for the *Argentine immigrants* problem was in error by a factor of 1.97 from the actual, while for *Bank failures* it was in error by a factor of 1.42. For both problems, errors for the U.S. population component were less than errors for the global quantities. Nevertheless, we were surprised at the difficulty individuals had with estimating this value. In practical problems, of course, one could simply use the actual value. In their study of decomposition, Henrion et al. (1993) gave the U.S. population value to the subjects.

Fig. 1. Calibration of probability assessments that estimated answer is within 10% of true answer.



The issue of 'how extreme is extreme' has not been resolved. We proposed a definition based on the number of digits (six or seven), but we did not examine alternatives. Nor did we resolve the issue of how to specify the unit of measure.

We expect that other conditions might affect decisions on when to use decomposition. For example, question type may have some importance. We do not know the extent to which our problem selection may have affected findings.

6. Discussion

Despite the improved accuracy it afforded, decomposition did not increase subjects' confidence in the accuracy of their estimates. However, the interquartile estimates were smaller for the decomposed estimates and confidence in the accuracy of estimates was slightly more appropriate.

Perceived uncertainty measures are easy to obtain. As shown in Table 6, self-assessments of uncertainty provided similar rankings of the relative uncertainty for the problems. The interquartile ranges provided somewhat different information than the self-assessments. Interquartile ranges of the estimates are not expensive, but they do require a pretest.

The present study addresses the issue of whether estimates by individuals can be improved when no other data are available. However, we expect that other situational characteristics or estimation-aiding strategies would also affect the usefulness of decomposition. For example, a forecaster could decompose a problem to use different sources of information or different experts. For some parts of the problem, known values may exist. Alternative decomposition methods could be used to produce an estimate, and resulting values for a quantity could be resolved in light of one another. MacGregor and Liechtenstein (1991) attempted such an approach and found that subjects tended to resolve estimates by applying an averaging model. Revised estimates generally fell between two estimates of a target quantity, where each judgmental estimate was produced by a different method.

Although our approach to decomposition was harmful for problems that did not involve extreme - uncertain numbers, there might be alternative approaches that are successful. For example, decomposition might restructure a problem so that it is easier for subjects to think about.

Decomposition tended to reduce estimators' confidence levels, perhaps because of the increased processing involved. This reduction in overconfidence and the improvements in accuracy produced modest gains in calibration.

7. Conclusions

The theory behind decomposition is simple. What is difficult is how to translate the theory into operational terms. We examined some operational procedures for identifying conditions under which decomposition should improve accuracy.

Extreme uncertain values are difficult for subjects to estimate. We hypothesized that decomposition to remove extreme values would improve estimation accuracy. This study examined nine "extreme value-high uncertainty" problems from two prior studies. Decomposition proved useful for each of these nine problems, and the typical gain in accuracy was substantial (error ratio was reduced by 96.3 for the study with six problems, and by 12.3 for the study with three problems). In the present study, involving six problems with extreme values, the error ratio was reduced by a factor of almost 20.⁶ Decomposition failed for one extreme problem because it was not successful in producing more accurate estimates of the parts.

⁶ The results from Hora et al. (1993) also are consistent with our hypothesis. They found that decomposition was more accurate than global estimates for three quantities whose true values had at least eight digits (e.g., What were the sales for Long's Drug Stores in Hawaii in 1986?).

Decomposition was risky for problems that did not involve extreme and uncertain values. For six such problems from two prior studies, decomposition had little overall effect on accuracy. However, for four such problems in the current study, decomposition yielded less accurate estimates by an average error ratio of 458%.

Based on the limited evidence to date, we suggest the following procedure for judgmental decomposition. First, assess whether the target value is subject to much uncertainty by using either a knowledge rating or an accuracy rating. If the problem is an important one, obtain interquartile ranges. For those items rated above the midpoint on uncertainty (or above 10 on the interquartile range), conduct a pretest with 20 subjects to determine whether the target quantity is likely to be extreme. If the upper quartile geometric mean has seven or more digits, decomposition should be considered. For these problems, compare the interquartile ranges for the target value against those for the components and for the recomposed value. If the ranges are less for the global approach, use the global approach. Otherwise use decomposition.

The current study suggests that decomposition has more limited value than previously thought. It improved accuracy only when the situation involved uncertain and extreme quantities. Furthermore, decomposed elements needed to be easier to estimate than the global. For problems that did not concern extreme values with high uncertainty or where estimates of the parts were not more accurate than that of the target value, decomposition produced less accurate estimates.

Acknowledgements

This research was supported in part by the National Science Foundation under Contract SES-9013069 to Decision Research. Fred Collopy, George Loewenstein, Robin Hogarth and unidentified referees provided helpful comments on early drafts. Jennifer L. Armstrong, Suzanne Berman, Gina Bloom, Vanessa Lacoss, Phan Lam and Leisha Mullican provided editorial assistance.

References

- Armstrong, J.S., W.B. Denniston and M.M. Gordon (1975), "The use of the decomposition principle in making judgments," *Organizational Behavior and Human Performance*, 14, 257-263.
- Aschenbrenner, K.M. and W. Kasubck (1978), "Challenging the Cushing syndrome: Multiattribute evaluation of cortisone drugs," *Organizational Behavior and Human Performance*, 22, 216-234.
- Henrion, M., G.W. Fischer and T. Mullin (1993), "Divide and conquer? Effects of decomposition on the accuracy and calibration of subjective probability distributions," *Organizational Behavior and Human Performance*, 55, 207-227.
- Hertzberg, H. (1970), *One Million*. Simon and Schuster, New York.
- Hora, S.C., N.G. Dodd and J.A. Hora (1993), "The use of decomposition in probability assessments of continuous variables," *Journal of Behavioral Decision Making*, 6, 133-147.
- MacGregor, D.G. and S. Lichtenstein (1991), "Problem structuring aids for quantitative estimation," *Journal of Behavioral Decision Making*, 4, 101-116.
- MacGregor, D.G., S. Lichtenstein and P. Slovic (1988), "Structuring knowledge retrieval: An analysis of decomposed quantitative judgments," *Organizational Behavior and Human Decision Processes*, 42, 303-323.
- Raiffa, H. (1968), *Decision Analysis*. Princeton University Press, Princeton, New Jersey.
- Rosenthal, R. (1978), "Combining results of independent studies," *Psychological Bulletin*, 85, 185-193.

Sniczek, J.A., P.W. Paese and F.S. Switzer, III (1990), "The effect of choosing on confidence in choice," *Organizational Behavior and Human Decision Processes*, 46, 264-282.