Informants in Organizational Marketing Research: How Many, Who, and How to Aggregate Responses?

Gerrit H. van Bruggen, Gary L. Lilien, Manish Kacker

ERIM Report Series *Research in Management*	
ERIM Report Series reference number	ERS-2000-32-MKT
Publication status / version	draft / version August 2000
Number of pages	45
Email address first author	GBruggen@fbk.eur.nl
URL Location (elec. version)	http://www.eur.nl/WebDOC/doc/erim/erimrs20000822122555.pdf
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Erasmus Universiteit Rotterdam PoBox 1738 3000 DR Rotterdam, The Netherlands Phone: # 31-(0) 10-408 1182 Fax: # 31-(0) 10-408 9020 Email: info@erim.eur.nl Internet: www.erim.eur.nl

ERASMUS **R**ESEARCH **I**NSTITUTE OF **M**ANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	Organizational research frequently involves seeking judgmental data from multiple informants within organizations. Researchers are often faced with determining how many informants to survey, who those informants should be and (if more than one) how best to aggregate responses when disagreement exists between those responses. Using both recall and forecasting data from a laboratory study involving the MARKSTRAT simulation, we show that when there are multiple respondents who disagree, responses aggregated using confidence-based or competence-based weights outperform those with data-based weights, which in turn provide significant gains in estimation accuracy over simply averaging respondent reports. We then illustrate how these results can be used to determine the best number of respondents for a market research task as well as to provide an effective screening mechanism when seeking a single, best informant.	
Library of Congress Classification (LCC)	5001-6182	Business
	5410-5417.5	Marketing
	HF 5415.3	Market Survey
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 31 C 44	Marketing Statistical Decision Theory
	M 31	Marketing
European Business Schools Library Group (EBSLG)	85 A	Business General
	280 G 255 A	Managing the marketing function Decision theory (general)
	280 L	Marketing Research
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organisatiekunde: algemeen
	85.40 85.03	Marketing Methoden en technieken, operations research
	85.40	Marketing
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Marketing / Besliskunde	
	Marktonderzoek, Steekproef, Response	
Free keywords	Organizational Research, Marketing Research, Survey Research, Aggregation, Screening, Key Informants.	
Other information		

Informants in Organizational Marketing Research:

How Many, Who, and How to Aggregate Responses?

Gerrit H. van Bruggen*

Erasmus University Rotterdam

Gary L. Lilien

The Pennsylvania State University

Manish Kacker

Northwestern University

August 2000

*Contact address: Gerrit H. van Bruggen, Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, Phone: 31-(0)10-4082258, Fax: 31-(0)10-4089011, Email:GBRUGGEN@FAC.FBK.EUR.NL

The authors would like to acknowledge the support of Penn State's Institute for the Study of Business Markets

We thank Arvind Rangaswamy, John Rossiter and Ale Smidts for their insightful comments on an earlier draft of this paper. All errors remain our responsibility.

Informants in Organizational Marketing Research:

How Many, Who, and How to Aggregate Responses?

Abstract

Organizational research frequently involves seeking judgmental data from multiple informants within organizations. Researchers are often faced with determining how many informants to survey, who those informants should be and (if more than one) how best to aggregate responses when disagreement exists between those responses. Using both recall and forecasting data from a laboratory study involving the MARKSTRAT simulation, we show that when there are multiple respondents who disagree, responses aggregated using confidence-based or competence-based weights outperform those with data-based weights, which in turn provide significant gains in estimation accuracy over simply averaging respondent reports. We then illustrate how these results can be used to determine the best number of respondents for a market research task as well as to provide an effective screening mechanism when seeking a single, best informant.

Key Words: Organizational Research, Marketing Research, Survey Research, Aggregation, Screening, Key Informants.

1. Introduction

In a current research study involving organizational adoption of e-commerce technologies, we are collecting data from multiple informants in a number of organizations. While we were not surprised to find that the attitudes and perceptions of these informants differed about the innovativeness of their organization, their organizational culture and so on, we were challenged about how to address differences we found in their reports of the size of their organization (revenue, profit, number of employees), its historical growth rate, and other reports that should involve factual, ***recall*** data. For reasons of respondent anonymity, we could make no independent check of their responses.

In another research study involving the relative effectiveness of different new product generation procedures (also involving multiple respondents across different divisions of the same organization) we asked respondents to forecast the size of the business that would result from each (funded) product concept over the next five years. We are finding that these ***forecast data*** for the same business opportunity vary widely across respondents.

We are also involved in a consulting project for one of the world's largest telecommunications companies, involving responses about current and anticipated organizational needs for and adoption of telecommunications equipment. The sample involves respondents in over 2000 organizations, and we ask for their current telecommunication usage by type of service, by geographic area as well as data (as above) about the size and historical growth rate of the organization. Again, we find large differences in reports of both historical, factual data as well as in reports of forecasted growth and organizational needs (ie, both for ***recall and forecast data***).

In each of these situations, we face the problem of how to address the discrepancies in reports across informants from the same organizational unit to arrive at an overall, unit-level measure. Should we discard some responses and use only a single informant per organizational unit? If so, which informants should be discarded? Or should we consider all responses? If so, how should we aggregate these responses? With the benefit of hindsight, should we have used only a single respondent for each organizational unit? If so, what would have been the best way to screen for them?

As empirical, organizational research involves, almost by definition, multiple stakeholders, the data challenges we noted above pervade both academic research and market research practice (Reid and Plank, 2000). Hence, some fundamental questions that researchers face in collecting organizational data are (1) who to ask, (2) how many to ask, and (if more than one) (3) how to aggregate[1] the responses.

In this paper we address several issues. First, we ask, given that response disagreement exists, what is the best way to aggregate data from multiple informants into a single estimate? Second, we investigate the benefits in data quality improvement that we can achieve with such aggregation as we vary the number of informants. We find that responses aggregated using confidence-based or competence-based weights outperform those with data-based weights, which in turn provide significant gains in estimation accuracy over averaging respondent reports. Third, using the answers to the first two questions, we develop a procedure to determine the best number of respondents for a specific research study. Finally, we investigate if our findings can help develop a procedure to screen for a best *single* informant when a research design dictates that such a respondent be selected.

We proceed as follows. First, we develop a theoretical framework for analyzing informant issues. Next, we motivate and develop some alternative approaches for aggregating multiple-informant reports and investigate how they perform in two empirical tests. In the first empirical test (i.e., Study One) we measure the performance of the different weighting procedures on recall data. In Study Two we conduct similar analyses using forecasting data. After that, we derive results that show the value of increasing the number of respondents. Then we apply our findings to the selection of the best number of (multiple) respondents as well as to the problem of screening for a single respondent. We conclude with a discussion of the implications of our findings for academic and research communities and suggest additional research to assess the generalizability of our findings.

2. A Framework for Informant Issues

If one collects information through informants, two issues are important: 1) determining the number of informants and 2) developing a way to aggregate data if one collects data from multiple informants. We suggest a theoretical framework to address these two issues. In doing so, we assume that a true score exists for the organizational variable for which we want to develop a measure. The measured value of the variable (i.e., in our study the value of the informant's response) consists of two components:

Measured Value = True Score + Error

Where:

Error = Systematic Error + Random Error

We assume that the expected value of the random error is zero. The systematic error in an informant's response can result from, for example, the hierarchical or the functional position of the respondent.

Determining the Number of Informants We propose a series of consecutive questions (see Figure 1) that need to be answered in order to address the question of how many informants one would need to develop accurate measures at the organizational level.

Please Insert Figure 1

First, one must determine whether the variable of interest can be measured objectively: sales or number of employees at a location are examples of such variables. For such variables, an objective, true score exists. To measure subjective variables one will always expect real differences across informants. Examples of subjective variables include attitudinal variables like the evaluation of an organization's satisfaction with its channel partners and assessments of the nature and magnitude of channel conflicts; we will not address such questions here.

If variables can be measured objectively, the next question is whether data can be obtained from an existing source such as administrative records or an archive. If such records are available, it is probably advisable to collect data from those sources. If objective data are either not available or not directly accessible (Venkataraman and Ramanujam, 1987; Kumar, Stern and Anderson, 1993), as is often the case with historical or confidential data (Kacker, 1997), one will have to obtain proxy retrospective judgmental data through informants.[2]

Once one decides to collect primary, perceptual data (recollections, assessments or forecasts), a key question is whether to rely on reports from a single informant or to collect data from multiple informants. While it is clearly more convenient to rely on a single informant, several researchers have found that a multiple informants-based approach often yields data of far superior quality (Seidler 1974; Hogarth 1978; Hill 1982; Wilson and Lilien 1992). Consequently, researchers often recommend relying on multiple informants for the study of both intra-organizational (e.g., Silk and Kalwani, 1982; Wilson, Lilien and Wilson, 1991; Wilson and Lilien, 1992) and interorganizational phenomena (e.g., Philips, 1981; Bagozzi and Philips, 1982; John and Reve, 1982).

Multiple informants improve data validity (Philips, 1981) because researchers can use systematic differences amongst informants to correct for individual differences and biases in estimates provided by these informants (Wilson and Lilien, 1992; Anderson 1985; Anderson 1987). Philips (1981) notes that informant reports often exhibit less than 50% of the variance attributable to the trait factor under investigation, with random error and informant biases accounting for the rest of the variance. Random error in response may result from the fact that individuals who are asked to assume the role of a key informant and make complex judgments find it difficult to make those judgments accurately (Philips 1981). Informant biases result from differences in informants' organizational roles and perspectives (Seidler, 1974; Houston and Sudman, 1975). The use of single informants limits the researcher's ability to control for functional or response bias (Huber and Power 1985; Philips 1981). Indeed, even if data are collected from a *homogeneous* group of informants (i.e., informants with similar perspectives within the organizational unit), it may be beneficial to interview multiple informants, as individual informants may suffer from memory failures or response-distortions (Golden, 1992).

The answer to the question of whether one needs multiple informants or whether a single informant report suffices depends on the magnitude of the error that can be expected in the informants' responses. If this error is near zero (i.e., the informant produces nearly perfectly accurate information) one would need only one single informant. If, as Philips (1981) and others have found, the error is substantial, multiple informants are needed. The question that needs to be addressed then is how many such informants are needed.

The minimum number of informants needed depends on the composition of the error. If only random error is expected (i.e., the systematic error is expected to be near zero) then the minimum number of informants is two. Furthermore, the more informants, the higher the expected accuracy of an aggregated measure. However, the additional number required for a given improvement in accuracy increases in proportion to the number of individuals and depends both on the individual accuracy and the average intercorrelation between the informants' responses (Hogarth 1978) – i.e., the higher the intercorrelation among existing informants the lower, the added value of an additional individual informant). Generally, there is an optimum number of informants, depending on the costs of obtaining independent judgments and on the costs of error in the final group judgment (Ferrell 1985).

If systematic error is also expected to be present in the informants' responses, the minimum number of informants depends on the cause of this systematic error. Systematic error in organizational research is associated with informants employed in different functional departments or at disparate hierarchical levels in the organization (Philips, 1981; Golden, 1992; Kardes, 1998). In such situations, heterogeneity in organizational position causes informants to have different perspectives. If systematic error is expected to be present, the minimum number of informants should equal the number of different informant perspectives in order to determine the impact of the informant's position on his/ her response. Again, increasing the number of informants will improve accuracy and the optimal number of informants will depend on the trade-off between accuracy and costs.

Approaches for Aggregating Multiple Informant's Responses: Data collected from multiple informants often reveal a surprising lack of agreement (e.g., Anderson and Narus, 1990), even if the informants share a similar background in terms of knowledge or organizational position. Phillips (1999) posits that this variation could result from differences in the cognitive processes used by informants. This variation leads to the "perceptual agreement problem" (Kumar, Stern, and Anderson 1993).

We focus on a situation (see Figure 1) in which data on objective variables (i.e., sales, profit, marketing expenditures etc.) are collected through informants. We do this where no a-priori, systematic error in the responses of the informants is expected (i.e., the informants occupy similar positions),[3] but where random error is expected, resulting in perceptual disagreement..

The next question then becomes how to aggregate the opinions of the different members of a group into a single group composite value? Two basic methods for aggregating individual informants' reports are in use: behavioral and mathematical. In behavioral aggregation, informants discuss the matter, work out their differences and agree upon a value (Ferrell 1985). This approach seems to solve the aggregation problem directly, however, behavioral aggregation requires considerable effort and (potentially impractical) coordination amongst respondents in the collection of the data. In addition, respondent requirements for anonymity and confidentiality may make it infeasible to apply. Finally, it is possible that any consensus reached is a poor indicator of perceptual agreement since the consensus may be affected by group properties and processes – e.g., power-dependence relations among informants, coalition formation, conformity pressure, and groupthink (Schwab and Heneman, 1986).

The effort and coordination required by the behavioral aggregation approach prompted Kumar, Stern and Anderson (1993) to propose a hybrid consensus-averaging approach. Here, they average responses to arrive at composite measures when there are only minor differences between informant reports. Where there is a major disagreement among knowledgeable informants, they suggest the consensual approach. They assessed the performance of their approach using multiple informant data (sales managers and fleet managers) in a major vehicle rental company and found significant differences between the initial individual reports of the two informant positions. The subsequent consensual responses were more highly correlated with responses of the hierarchically superior position (sales managers) than with the inferior position (fleet managers). This result confirms that the process used to arrive at these consensual responses reflects underlying power-dependence relations and conformity pressures faced by underlings – the reasons why Schwab and Heneman (1986) do not favor consensual approaches.

Mathematical aggregation is attractive because of its relative ease of use and simplicity (Ferrell 1985). A widely used example of mathematical aggregation is the simple averaging of the judgments of separate informants. However, when informants exhibit substantial disagreement, such aggregation rarely produces the most accurate values (James, 1982)

Respondents often consistently over or underestimate quantitative variables. Hence, aggregation by averaging *n* individual judgments will give a group judgment with a variance smaller than that of the individual estimates, but will not eliminate any consistent bias (Ferrell 1985; Rowe 1992). Under such circumstances, it is valuable to find the individual with the smallest error. If the most accurate response can be identified, this response should be the group judgment. If the accuracy of group members cannot be determined with certainty, a weighted average of the members that assigns higher weights to those more likely to be accurate gives results that fall between the performance of the equally weighted average and the best member approach. The result will be closer to the best-member approach if informants with more accurate responses can be identified reliably. The question then becomes: how can we determine an informant's accuracy?

An individual informant's accuracy can be identified by using other group members' perceptions or by using self-assessments. When using other group members' responses to assess an individual's accuracy, we can compare the response of the individual with the responses of the group as a whole. Using a "majority rules" guideline, one can define an individual's response inaccuracy as its deviation from the group's common response. The larger the group size, the more accurate the (unweighted) group mean will be and, consequently, the more accurate the deviation of an individual from this group mean will reflect the inaccuracy of this individual. Another way of using group input to assess an individual group member's accuracy is to let group members judge the accuracy of fellow group members. While the first approach only needs the already collected data, the second approach requires additional effort from informants, who should be well known to each other and competent in assessing their fellows (Rowe 1992).

Self-assessment of expertise, knowledge or confidence is an alternative to determine informants' accuracy. If respondents are biased about their ability, this approach can lead to overconfidence or underconfidence (Mahajan 1992). Rowe (1992) indicates that self-rated confidence might be an appropriate measure of expertise when subjects can actually evaluate their confidence in a specific problem area to which they are regularly exposed. We expect that simplicity (Fischhoff, Slovic, and Lichtenstein 1977) of and familiarity (Gigerenzer, Hoffrage, and Kleinbolting 1991) with the task will improve the accuracy of informants' expertise self-assessments.

We present formal operationalizations of these ideas in the section below and then apply them to two empircal studies in the sections that follow.

3. Data Aggregation Approaches

In line with the discussion above we apply three approaches to aggregate the scores of informants in our empirical studies: (a) an unweighted group mean (our reference value); (b) a value where weights are derived from the data (i.e., using group information), and (c) a value where the weights are derived from self-reported confidence scores.

Unweighted Group Mean. Our first (and benchmark) aggregation method entails computing the arithmetic mean of the individual scores of the group members. This is the simplest form of the "aggregation approach" (Kumar, Stern, and Anderson 1993). The value of the unweighted mean for variable *X* (X = 1,.., 8), $UNWMEAN_{xi}$, of group *i* can be computed as follows:

(1) $$UNWMEAN_{xi} = \frac{\left(\sum_{j=1}^{n_i} X_{ij}\right)}{n_i}$$

Where:

X_{ij} = the estimate of the value of variable *X* by informant *j* in group *i*.

n_i = number of informants in group *i*.

Data-Based Weighted Mean. The next aggregation method derives from the view that the extent of agreement between informants contains information that should be incorporated in the aggregate measure. For example, when two informants in a three-informant group provide similar estimates while the third informant provides a substantially different value, the estimates provided by the two agreeing informants might be weighted more heavily than that of the third. (This approach assumes that the true value is closer to the estimates provided by agreeing informants than to that of the deviating informant(s).) Developing such aggregated values addresses James' (1992) call to demonstrate perceptual agreement between informants before measurements are aggregated.

To develop such a data-based measure, we must compute weights for the estimates of each informant. We first compute $DIST_{xij}$, the absolute distance of informant *j*'s estimate of variable *X* from the unweighted, arithmetic mean of group *i* (which *j* belongs to):

(2) $$DIST_{xij} = \left| X_{ij} - UNWMEAN_{xi} \right|$$

The weight assigned to informant *j*'s estimate should be inversely related to its absolute distance from the unweighted mean for group *i*, relative to the distances of the other group members, so we compute the weight for informant *j*'s estimate of variable *X* ($WEIGHT_{xij}$) as follows:

(3) $$WEIGHT_{xij} = \left(\frac{\left(\sum_{j=1}^{n_i} DIST_{xij} \right)}{DIST_{xij}} \right)^{a}$$

In Equation 3 we have introduced a parameter α, with reference value of 1.0. It can be raised to increase the weight of observations that are close to the arithmetic mean, relative to the weights for observations that are further away; as α approaches zero, these weights will approach those associated with the unweighted mean.

Finally, we compute the weighted mean $WDMEAN_{xi}$ of variable X for each group *i*, where the estimates for each group member are weighted according to their distance from the unweighted group mean:

$$\text{(4)} \qquad \text{WDMEAN}_{xi} = \sum_{j=1}^{n_i} \left(\frac{\text{WEIGHT}_{xij}}{\left(\sum_{j=1}^{n_i} \text{WEIGHT}_{xij} \right)} * X_{ij} \right)$$

Confidence-Based Weighted Mean. Our third aggregation approach uses weights based on informants' confidence as reflected in their self-assessed confidence in the accuracy of each estimate they gave. Here, we weight estimates provided by more confident informants more heavily than those from less confident informants. WCMEAN_{xi}, the value of variable *X* for group *i* in which informant *j*'s estimate is weighted by his or her confidence CONF_{xij}, in the accuracy of that estimate is as follows:

$$\text{(5)} \qquad \text{WCMEAN}_{xi} = \sum_{j=1}^{n_i} \left(\frac{\text{CONF}_{xij}^{a}}{\left(\sum_{j=1}^{n_i} \text{CONF}_{xij}^{a} \right)} * X_{ij} \right)$$

Again, we introduce a parameter α (with a reference value of 1) that makes it possible to manipulate the weight assigned to more confident informants. As above, when α approaches zero, the estimate reduces to the arithmetic mean.

While there are clearly many other possible approaches, these models represent a range of possible aggregation procedures.

4. Study One: Aggregating Recall Data

Methodology: To assess the accuracy of these aggregation methods and measure the benefits of having different numbers of informants, we had to collect data in a realistic organizational setting where we could compare informants' estimates with the objective, "true" values. We used the environment of the MARKSTRAT simulation (Larréché and Gatignon, 1990), a computer-based, marketing strategy simulation that has been widely used by researchers to study decision-making (e.g., Hogarth and Makridakis, 1981; Glazer, Steckel, and Winer, 1989; Glazer and Weiss, 1993). In the simulation, groups of participants play over a number of periods and make strategic and tactical marketing decisions for different competing firms.

The informants in our study were 67 marketing students participating in a capstone marketing strategy course at a large Midwestern US university. The students formed two, three, or four-person groups to make decisions for one of five companies operating in one of four MARKSTRAT industries. Altogether, the 67 informants operated in 20 groups. The students made each decision after analyzing results from previous periods and reviewing market research studies. All groups received the same amount of time to evaluate their positions and make decisions, and all groups made decisions simultaneously.

After groups had played the game for a few periods, we asked each informant to complete a questionnaire *individually*. Amongst other questions, we asked them to recall the values of eight variables (the levels of marketing mix variables, such as advertising, price, and sales effort) for decisions they had just made as well as the size of the marketing budget they had available to spend for the next set of decisions. (See Appendix A for the eight relevant questions.) We also asked them to record their confidence in their responses to these questions on a nine-point scale, where 1 indicated "not certain at all" and 9 reflected "completely certain" about the accuracy of the estimate. To ensure involvement, to stimulate accuracy, and to discourage cooperation between group members, we awarded prizes to the individuals who came up with the most accurate estimates.

Two elements of the research context deserve mention. First, all group members were students who were not assigned any specific hierarchical positions or functional responsibilities. Consequently, they shared the same (homogeneous) viewpoint, with limited scope for hierarchical or functional bias. Second, the MARKSTRAT program provided the actual values of all variables, so we were able to explicitly assess the accuracy of the informants' reports.

Empirical Results: We applied the aggregation procedures described in the last section to the data produced by our experimental groups. In line with the literature, we use the Mean Absolute Percentage Error (MAPE), a dimensionless metric (Kennedy 1985), as our index of relative performance. The mean absolute percentage error $MAPE_i$ of group i for a specific aggregation approach (averaged over the eight MARKSTRAT variables) is computed as follows:

$$(6) \qquad MAPE_i = \left(\sum_{X=1}^{8} \left| \frac{\text{Estimated Value of } X_i - \text{Real Value of } X_i}{\text{Real Value of } X_i} \right| * 100\% \right) / 8$$

In Table 1 we present the MAPE for the three aggregation approaches.

Please Insert Table 1

The results in Table 1 show that weighting improves accuracy (F=15.45, p=.001) (and thus decreases MAPE). Further, confidence-based weighting does better than data-based weighting: compared to the unweighted mean, the confidence-based weighted mean improves accuracy by more than 20 percent (F=20.88, p<.001). For the data-based and confidence-based weighting procedures results in Table 1, we used the reference α value of 1. Next, we investigated whether accuracy could be improved by allowing the value of α to differ from its reference value.

For the data-based weighting procedure, we calculated the value of α in Equation 2 that minimized MAPE (in Equation 7) using the Solver module in Microsoft Excel. We computed an optimal value of α for each of the eight MARKSTRAT variables (*variable-specific* a). We also computed a single, optimal α that was restricted to have the same value for all eight MARKSTRAT variables (*uniform* a). The results in Table 2 show that the accuracy of the data-based weighting procedure can be improved by about 15% by this procedure (F=9.110, p=.007). The optimal uniform value of α was 25.70, while optimal variable-specific values of α ranged from 2.88 to 77.00. The difference between the MAPE of the uniform α approach and the MAPE of the variable-specific α approach is quite small (12.45/14.20 = 87.7% or 12.3% improvement in MAPE versus 12.35/14.20 which yields a 13.0% improvement). Apparently, most of the gain in MAPE derives from weighting the agreeing individuals most heavily (α substantially greater than 1) while MAPE is relatively insensitive to the actual value of the higher weight.

Please Insert Table 2

We optimized α in the confidence-based weighting approach in a similar manner. We calculated the value of α in Equation 5 that minimized the MAPE (in Equation 7), and computed both variable-specific values of α and a uniform, optimal α. Again we found that increasing the weights of the more confident informants leads to substantially more accurate aggregation results: MAPE for the unweighted group mean is 16.30; setting α equal to 1 gives a MAPE of 12.81; a single optimal α brings the MAPE down to 7.90 and using item-specific values for α yields a MAPE of 7.53. Thus, with confidence-based weights, MAPE can be reduced from 16.30 (for the unweighted group mean) to 7.53, yielding a reduction of over 50% ($F=19.876$, $p<.001$). As with the data based weights, most of this gain comes from increasing the value of α to well above 1, with only incremental improvements arising from item-specific tuning.

Based on the results of Table 1 and Table 2, we conclude (a) that applying a weighting procedure leads to considerably more accurate aggregation results than does using the arithmetic mean and (b) that confidence based weights do better than data based weights. To see how robust the latter finding is, we investigate how well the confidence based approach would do if we used a single (overall) confidence value as opposed to the item-specific values we have used in Table 2. Table 3 gives the results, where we used the respondents' overall confidence score, obtained by averaging specific confidence levels indicated for each variable. We find that MAPE increases by only 11% for both the uniform α (from 7.90 to 8.79) and variable-specific values of α (from 7.53 to 8.34). Thus, if we are asking for judgments on many items, it would be appropriate to seek only a single overall confidence judgement; the loss in accuracy is minimal while the reduced cognitive burden on informants will enhance the quality and quantity of responses.

Our results were similar when we used other measures of central tendency (median) or alternative functional forms involving more than one parameter (Little’s 1970 Adbudg s-shaped function, for example) for calculating weights for the data-based and confidence-based aggregation approaches.

Please Insert Table 3

5. Study Two: Aggregating Forecasting Data

The relatively strong performance of the confidence-based weighting procedure may seem surprising since these kind of self-assessments have not always been found to be very accurate (Larréché and Moinpour 1983). The relatively simple character of the task (i.e., straight recall) used in our study could help explain our findings in the context of the extant literature. In a second study we investigated whether the performance of the various aggregation methods would hold for a more complex task: forecasting. In such a setting, one might expect the self-assessed confidence-based weight to perform worse than for a recall task.

The informants in Study Two were 39 marketing students participating in a capstone marketing strategy course at a large Midwestern US university. The students formed two, three, or four-person groups that made decisions for one of five companies operating in one of four MARKSTRAT industries. The 39 informants operated in 13 groups. As before, students made each decision after analyzing results from previous periods and reviewing market research studies. All groups received the same amount of time to evaluate their positions and make decisions, and all groups made decisions simultaneously.

Again, we asked each informant to complete a questionnaire *individually* after groups had played the game for a few periods. This time, they were asked to make forecasts for the values of three variables: the brand awareness levels of the two brands they were responsible for managing and the marketing budget they expected to have available for the next period. This budget was a function of the profit their company would generate on the basis of the decisions they were making (See Appendix B for the relevant questions.) Again we asked them to record their confidence (in their responses to these questions) on a nine-point scale, where 1 indicated "not certain at all" and 9 reflected "completely certain" about the accuracy of the estimate. As in the previous study, we awarded prizes to individuals who came up with the most accurate values. Again, the MARKSTRAT program provided us with the actual values of the three forecasted variables, so we were able to explicitly assess the accuracy of the informant reports.

Empirical Results In Table 4, we present the results of applying the aggregation procedures to the forecasting data. Overall, these results are consistent with those of Study One – weighting improves accuracy, especially for confidence-based weights. Both for the variable-specific confidence scores and for the average confidence scores (averaged for the two brand awareness forecasts and the budget forecast) the values of the optimal α are higher in Study Two than in Study One. This suggests that for the more complex forecasting task, the self-stated confidence scores are more informative than for the recall task.

Please Insert Table 4

Besides replicating the analyses from Study One, we also applied three new types of weights:

1. a single overall confidence score that expressed the informants' confidence in all the forecasts they provided us with;

2. a recall competence score that reflected the informants' accuracy in recalling variables from the previous MARKSTRAT period (i.e., the same eight variables used in Study One); and
3. a forecasting competence score that reflected the informants' accuracy in providing us with forecasts on two other MARKSTRAT variables, i.e., the sales for Brand 1 and for Brand 2.

(Note that to calculate the competency-based weights we use the formulation in equations 2-4, with "actual value" replacing UNMEAN in those equations.)

The results in Table 4 show that using the overall *confidence* score as a weight produces less accurate aggregated values than using variable-specific confidence scores. Given the relatively low optimal α for this type of weight (compared to the average or individual-item confidence-based weights), we conclude that the overall confidence score is less informative than the other confidence scores. Evidently, the informants knew that they were less accurate on some variables than on others and expressed this in the item-specific confidence scores.

We find that using the *forecasting competency* scores as weights leads to results that are as accurate as using the confidence scores, while using *recall competency* leads to less accurate results. Apparently, performance on a specific task (i.e., a forecasting task) is a fairly accurate predictor of accuracy on a similar task (i.e., forecasting other variables), while accuracy scores on a different task (i.e., recalling variables) produce less useful information.

Overall, Study Two shows that confidence-based weighting improves the accuracy of aggregated variables. Competence-based weighting can do as well as long as that competency is measured on task similar to the one under study.

6. How does adding respondents increase estimation accuracy?

We have seen thus far that we can increase estimation accuracy by weighting the responses of individual respondents. We now use these results to investigate the extent that the accuracy of these appropriately aggregated scores improves as we add respondents. For this purpose we use the recall data collected in Study One.

Method: In Study One we collected recall data from twenty groups. Two of these groups consisted of two informants, nine groups consisted of three informants, and nine consisted of four informants. To determine the extent to which accuracy increased by adding informants, we drew *random* samples of a varying number (1, 2, 3, or 4) of informants from each group. Next, we developed aggregated group scores using the three aggregation approaches described in the Section 3. For the data-based weighting approach, we used the optimal *uniform* value of α (= 25.70). For the confidence-based weighting approach, we used the single average confidence scores as weights with the optimal *uniform* value of α (= 13.18). Since all groups consisted of at least two informants, we were able to determine the accuracy of one-person and two-person groups for twenty groups. Since eighteen groups had at least three respondents and nine had four, our three and four person accuracy assessments relied on samples of eighteen and nine groups, respectively.

Empirical Results: Figure 2 shows that, from a reference value of MAPE of 23.45, increasing group size to 2 (with confidence-based weights as described above) improves MAPE to 17.85 or by 24% ($F=3.319$, $p=.084$); adding a third respondent lowers MAPE to 10.64 or by an additional 31% while adding the fourth respondent takes that value to 8.99 for an additional 7% improvement. The total gain from increasing the number of respondents from 1 to 4 and applying a simple (one item) confidence based weighting scheme is thus 1-8.99/23.45 or 62%! The improvements for the unweighted mean and the data based approaches, while directionally consistent, are not as dramatic. The biggest gain in accuracy comes from adding a third informant. This confirms the findings of Libby and Blashfield (1978) who found that the most dramatic gain in precision when aggregating data from multiple respondents came with three judges.

Please Insert Figure 2

Our results show that using confidence-based weights to aggregate information from informants gives more accurate results. If curves like Figure 2 were to be established for a specific marketing research study (perhaps through a pilot test), then the researcher could make an intelligent judgment of how many respondents to choose, balancing the increased accuracy (as seen in Figure 2) with the additional cost. Hence, depending on the purpose of the marketing research study, one might choose fewer organizations to study but select more (and more confident) respondents to get the best return on the marketing research investment. Indeed, our findings about the relationship between confidence (or competence) and accuracy raise a related question – can we use these findings to screen for a "best" respondent and, if so, how many respondents should we expect to have to screen? We outline a procedure to address these questions next.

7. Outline and illustration of a procedure for screening for a single informant

In most circumstances, marketing research practice dictates that a single, "best" respondent be used. Suppose we had a target value of (expected) accuracy for an estimate to be obtained from that respondent and we wanted a screening criterion to help select that respondent. Our results above suggest that we might consider using a self-stated confidence (or a measure of competence) score for this purpose.

Such a procedure would require that we knew (a) the relationship between confidence (or competence) and MAPE at the individual level; (b) the distribution of scores within the population and (c) the expected number of respondents to screen based on the results from (a) and (b). While we anticipate that the specific numerical results will be case and situation specific, we illustrate how such a procedure would work using our data collected in Study One.

The relationship between confidence and MAPE: In line with our suggestion above to combine confidence scores into an overall measure, we related average respondent MAPE to the respondent's confidence scores. We deleted eight respondents who reported essentially no confidence in their responses, leaving us with a sample size of 57. We ran a number of models (linear, quadratic, cubic, logistic) and found that the log model (Log CONF = b0 +b1*MAPE) fit best, with both parameters significant at the .001 level and an R^2 of 0.38. Figure 3 plots the resulting relationship.

Please Insert Figure 3

The distribution of population confidence: In order to determine the confidence distribution of our population, we constructed the empirical distribution of the confidence scores of our respondents. These are displayed in Figure 4.

Insert Figure 4

Calculating the expected screening sample size Given our results above, we can now determine the expected number of respondents we need to screen to get an expected confidence level. Note that if we specify a required MAPE, Figure 3 gives the necessary degree of confidence. Figure 4 then gives the likelihood that a randomly chosen individual will have that degree of confidence. If we call that value p, then how many individuals will we expect to have to screen, on average, before getting one with that confidence level or better? Note that if we assume that, in our market research study, we can sample randomly from the Figure 4 distribution, we have defined a geometric sampling process (i.e., the expected length of a series of Bernoulli trials until the first "hit") with probability parameter p; the expectation of that process is then 1/p. Figure 5 plots confidence versus expected sample size and Figure 6 integrates these findings, directly linking MAPE to expected screening sample size. In the present study, for example, the results in Figure 6 mean that if researchers find a MAPE of 40% acceptable, using a single informant is sufficient. For a MAPE of 20% one would need two informants on average and targets below 20% MAPE are essentially impossible to achieve for a single informant.

Insert Figures 5 and 6

We have now outlined a procedure that, given a desired MAPE and a population with an estimated distribution of confidence scores in providing estimates, tells how many individuals one would expect to screen to ensure that that the target MAPE was achieved. (Again, these are values that can be obtained in a marketing research pre-test, as we discuss next.)

8. Discussion

In this paper, we have addressed how to obtain more accurate judgments from informants in marketing research studies within organizations. In doing so, we make twofold contributions to extant marketing theory and practice: (a) we show how using multiple informants and aggregating multiple informant data using data-based distance weights or informant-based confidence or competence weights enhances the quality of data and (b) we develop and present simple procedures for marketing researchers and practitioners to incorporate our findings and insights into their data collection processes.

Our focal concern is how reports of multiple informants should be aggregated. Drawing on analyses of data collected from informants in the MARKSTRAT simulation, we were able to show that using multiple informants significantly enhances the quality of objective, recall and forecast data. The results in Figure 6 suggest that by using a single informant (after suitably screening for response-confidence) it will be very difficult to get a MAPE that is smaller than about 18%. Based on the same data, Figure 2 shows that aggregating results from multiple informants can cut that MAPE down to 9%. The quality of these results varies depending on the aggregation method: calculating unweighted, arithmetic means of informant reports was far inferior to techniques that weighted informant reports using self-assessed confidence, measured competence or data-based distance weights.

In our research setting, aggregation (even the computation of a simple unweighted mean) added accuracy to individual data. We demonstrated improvements in the accuracy of estimates, probably through a reduction in the error component of the individual-level data. This error component reflects perceptual biases or other cognitive limitations. Aggregation improves reliability by averaging those individual level errors and biases that are random (Rousseau 1985).

While previous research indicated weighting not to be effective in improving accuracy (e.g., Einhorn and Hogarth 1975; Fischer 1981) our results indicate that weighting is effective in enhancing the quality of recall or forecast measures based on multiple informant reports. This apparent contradiction may be due to the use of different types of weights in these studies. In our approach, reports provided by more confident and competent informants should be weighted more heavily than those from less confident and less competent informants. Note that in using competence-based weights, calibration of weights should be based on similar tasks.

To ensure that reports are from key informants, many researchers include questions in their survey instruments that assess informants' competency. The most effective technique for doing so entails using specific measures that assess the informant's knowledge of each major issue in the study (Kumar, Stern and Anderson, 1993). Our results show how using such competency measures as weights for informant reports can yield composite measures that are of superior quality to unweighted group means.

Our proposed weighting scheme is relatively simple to apply, especially if compared to more advanced hierarchical modeling approaches (e.g., Lipscomb, Parmigrani, and Hasselblad 1998). Indeed, to apply them, one needs only to

- obtain measures of reporting confidence for each group of responses; and

- include one or more questions on some responses in the study where *the answer is known* (to calibrate the procedure)

Results like the ones we produced in Figures 3-6, (that would allow informed selection of either the intra-organizational sample size or the expected number of responses to screen) could easily be obtained from an appropriately designed pilot test of a market research instrument.

Our study has a number of limitations. All our respondents reported on the same set of variables and were homogeneous in that there were no sources of functional or hierarchical bias. While that design improves the internal validity of our empirical analysis, further research should investigate how far these results generalize. In addition, our informants provided retrospective reports on observable phenomena as well as providing forecasts of these types of variables. Both tasks are much easier than making complex social judgments (Philips 1981; John and Reve 1982). It is likely that the nature and magnitude of the 'perceptual agreement problem' would become more significant in settings where respondents are required to provide complex, subjective responses.

In addition, our data were collected in a simulated setting. As with all such research, replications in other settings, both in the laboratory and in the field will be needed to better understand the realm of applicability of our findings. Indeed, it might well be that only our procedure (and not our specific empirical findings) has more general applicability.

On net, we are encouraged to have developed what appears to be both an easy to apply and relatively robust data aggregation procedure. It appears that this procedure can both justify the collection of data from an appropriate number of intra-organizational informants and help determine how best to screen for an appropriate single informant when reporting accuracy is critical to the research task.

References

Anderson, James C. (1985), "A Measurement Model to Assess Measure-Specific Factors in Multiple-Informant Research," *Journal of Marketing Research,* Vol. XXII, February, 86-92.

Anderson, James C. (1987), "An Approach for Confirmatory Measurement and Structural Equation Modeling of Organizational Properties," *Management Science,* Vol. 33, No. 4, April, 525-541.

Anderson, James C. and James A. Narus (1990), "A Model of Distributor Firm and Manufacturer Firm Working Relationships," *Journal of Marketing*, Vol. 54, 42-58.

Bagozzi, Richard P. and Lynn W. Philips (1982), "Representing and Testing Organizational Theories: A Holistic Construal," *Administrative Science Quarterly*, Vol. 27, 459-489.

Bagozzi, Richard P., Youjae Yi, and Lynn W. Philips (1991), "Assessing Construct Validity in Organizational Research," *Administrative Science Quarterly*, Vol. 36, 421-458.

Einhorn, Hillel J. and Robin M. Hogarth (1975), "Unit Weighting Schemes for Decision Making," *Organizational Behavior and Humand Performance,* 13, 171-192.

Ferrell, William R. (1985), "Combining Individual Judgments," in: G. Wright (ed.), *Behavioural Decision Making,* New York: Plenum Press, 111-145.

Fischer, Gregory W. (1981), "When Oracles Fail- A Comparison of Four Procedures for Aggregating Subjective Probability Forecasts," *Organizational Behavior and Human Performance,* 28, 96-110.

Fischhoff, B., P. Slovic, and S. Lichtenstein (1977), "Knowing with certainty: The appropriateness of extreme confidence," *Journal of Experimental Psychology: Human Perception and Performance,* 3, 552-564.

Gigerenzer, Gerd, Ulrich Hoffrage, and Heinz Kleinbolting (1991). "Probabilistic mental models: A

Brunswikian theory of confidence," *Psychological Review,* 98, 4, 506-528.

Glazer, Rashi, Joel Steckel and Russell Winer (1992), "Locally Rational Decision Making: The Distracting Effect of Information on Managerial Performance," *Management Science,* Vol. 38, 212-226.

Glazer, Rashi and Allen M. Weiss (1993), "Marketing in Turbulent Environments: Decision Processes and the Time-Sensitivity of Information," *Journal of Marketing Research*, Vol. 30 (November), 509-521.

Golden, B. R. (1992), "The Past is the Past – Or is it? The Use of Retrospective Accounts as Indicators of Past Strategy," *Academy of Management Journal*, Vol. 35: 848-860.

Hill, Gayle W. (1982). "Group Versus Individual Performance: Are N+1 Heads Better Than One?," *Psychological Bulletin*, 91, 3, 517-539.

Hogarth, Robin M. (1978), "A Note on Aggregating Opinions," *Organizational Behavior and Human Performance,* 21, 40-46.

Hogarth, Robin M. and Spiro Makridakis (1981), "The Value of Decision Making in a Complex Environment: An Experimental Approach," *Management Science*, Vol. 27, 93-107.

Houston, Michael J. and Seymour Sudman (1975), "A Methodological Assessment of the Use of Key Informants," *Social Science Research*, No. 4, 151-164.

Huber, George P. and Daniel J. Power (1985), "Retrospective Reports of Strategic-Level Managers: Guidelines for Increasing their Accuracy," *Strategic Management Journal,* Vol. 6, 171-180.

James, Lawrence R. (1982), "Aggregation Bias in Estimates of Perceptual Agreement," *Journal of Applied Psychology,* Vol. 67, No. 2, 219-229.

John, George and Torger Reve (1982), "The Reliability and Validity of Key Informant Data from Dyadic

Relationships in Marketing Channels," *Journal of Marketing Research,* Vol. XIX, November, 517-524.

Kacker, Manish (1997), "Growing a Franchise System: Structure and Strategy," unpublished Ph.D. dissertation, Northwestern University.

Kardes, Frank R. (1998), *Consumer Behavior and Managerial Decision Making*, Addison-Wesley.

Kennedy, Peter (1985), *A Guide to Econometrics,* 2nd edition, Oxford: Basil Blackwell.

Kumar, Nirmalya, Louis W. Stern and James C. Anderson (1993), "Conducting Interorganizational Research Using Key Informants," *Academy of Management Journal,* 36, 6, 1633-1651.

Larréché, Jean Claude and Hubert Gatignon (1990), *MARKSTRAT2*: *A Marketing Simulation Game.* Palo Alto, CA: The Scientific Press.

Larréché, Jean-Claude and Reza Moinpour (1983), "Managerial Judgment in Marketing: The Concept of Expertise," *Journal of Marketing Research,* Vol. XX, 110-121.

Libby, Robert and Roger K. Blashfield (1978), "Performance of a Composite as a Function of theNumber of Judges," *Organizational Behavior and Human Performance,* 21, 121-129.

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

Lipscomb, Joseph, Giovanni Parmigiani, and Vic Hasselblad (1998), "Combining Expert Judgment by Hierarchical Modeling: An Application to Physician Staffing," *Management Science,* Vol. 44, February, 149-161.

Mahajan, Jayashree (1992), "The Overconfidence Effect in Marketing Management Predictions," *Journal of Marketing Research,* Vol. XXIX, 329-342.

Philips, Lynn W. (1981), "Assessing Measurement Error in Key Informant Reports: A Methodological

Note on Organizational Analysis in Marketing," *Journal of Marketing Research*, Vol. XVIII, November, 395-415.

Phillips, Joan M. (1999), "How Informants Report About Interorganizational Relationships," ISBM Working Paper #23-1999.

Reid, David A and Richard E. Plank (2000) "Business Marketing Comes of Age: A Comprehensive Review of the Literature," *Journal of Business-to-Business Marketing*, Vol 7 No. 2/3, 9-185.

Rousseau, Denise M. (1985), "Issues of Level in Organizational Research: Multi-Level and Cross-Level Perspectives," *Research in Organizational Behavior*, Vol. 7, 1-37.

Rowe, Gene (1992), "Perspectives on Expertise in the Aggregation of Judgments," in: George Wright and Fergus Bolger (eds.) *Expertise and Decision Support*, New York: Plenum Press, 155-180.

Schwab, Donald P. and Herbert G. Heneman (1986), "Assessment of a Consensus-Based Multiple Information Source Job Evaluation System," *Journal of Applied Psychology*, Vol. 71, No. 2, 354-356.

Seidler, John (1974), "On Using Informants: A Technique for Collecting Quantitative Data and Controlling for Measurement Error in Organization Analysis," *American Sociological Review*, Vol. 39 (December), 816-831.

Silk, Alvin J. and Manohar U. Kalwani (1982), "Measuring Influence in Organizational Purchase Decisions," *Journal of Marketing Research*, Vol. 19 (May), 165-181.

Venkatraman, N. and Vasudevan Ramanujam (1987), "Measurement of Business Economic Performance: An Examination of Method Convergence," *Journal of Management*, 13, 1, 109-122.

Wilson, Elizabeth J. and Gary L. Lilien (1992), "Using Single Informants to Study Group Choice: An Examination of Research Practice in Organizational Buying," *Marketing Letters*, Vol. 3, No. 3, 297-305.

Wilson, Elizabeth J., Gary L. Lilien and David T. Wilson (1991), "Developing and Testing a Contingency Paradigm of Group Choice in Organizational Buying," *Journal of Marketing Research*, Vol. 28 (November), 452-466.

Table 1: Mean Absolute Percentage Errors (MAPE) of Three Aggregation Procedures (Study 1: Recall Data)

Aggregation Procedure	MAPE
1. Unweighted Group Mean	16.30 (10.08)
2. Data-Based Weighted Mean	14.20 (10.22)
3. Confidence-Based Weighted Mean	12.81 (8.37)

Table 2: Mean Absolute Percentage Errors (MAPE) of the Weighted Aggregation Procedures for Different Values of a **(Study 1: Recall Data)**

Aggregation Procedure	α = 1	Uniform α (optimized across MARKSTRAT variables)	Variable Specific α (optimized per MARKSTRAT variable)
Data-Based Weighted Mean	14.20 (10.22)	12.45 (10.76) (α = 25.70)	12.35 (10.71) (α range = 2.88–77.00)
Confidence-Based Weighted Mean	12.81 (8.37)	7.90 (6.24) (α = 12.87)	7.53 (6.07) (α range = 5.46–25.80)

Table 3: Mean Absolute Percentage Errors (MAPE) of the Weighted Aggregation Procedures for Variable-Specific Confidence Weights and Overall Confidence Weights (Study 1: Recall Data

Aggregation Procedure	α = 1	Uniform α (optimized across MARKSTRAT variables)	Variable Specific α (optimized per MARKSTRAT variable)
Confidence-Based Weighted Mean			
Using item-specific confidence score	12.81 (8.37)	7.90 (6.24) (α = 12.87)	7.53 (6.07) (α range =5.46–25.80)
Using single, average confidence score	14.69 (8.77)	8.79 (7.16) (α = 13.18)	8.34 (6.79) (α range = 5.05–165)

Table 4:
Mean Absolute Percentage Errors (MAPE) of Aggregation Procedures Applied to Forecasting Data (Study 2)

Aggregation Procedure	MAPE
1. Unweighted Group Mean	26.88 (32.17)
2. Data-Based Weighting	
$\alpha = 1$	25.57 (32.29)
uniform optimized α (= 5.47)	23.82 (32.00)
variable specific optimized α (ranges from 5.47 to 9.76)	23.80 (31.90)
3. Confidence-Based Weighted Mean	
<u>Variable specific confidence</u>	
$\alpha = 1$	22.96 (29.30)
uniform optimized α (= 53.34)	16.40 (17.63)
variable specific optimized α (ranges from 4.08 to 109.40)	16.21 (17.66)
<u>Average confidence</u>	
$\alpha = 1$	
uniform optimized α (= 20.91)	21.59 (24.40)
variable specific optimized α (ranges from 1.83 to 73.83)	15.75 (16.27)
	15.60 (16.39)
<u>Overall confidence</u>	
$\alpha = 1$	
uniform optimized α (= 6.04)	
variable specific optimized α (ranges from 3.93 to 7.86)	23.95 (29.07)
	19.19 (16.44)
	18.48 (17.23)
4. Competence-Based Weighted Mean	
<u>Recall Competence</u>	
$\alpha = 1$	
uniform optimized α (= 2.82)	22.81 (19.08)
variable specific optimized α (ranges from 1.99 to 12.40)	20.77 (18.26)
	20.41 (18.38)
<u>Forecasting Competence</u>	
$\alpha = 1$	
uniform optimized α (= 5.64)	19.02 (16.38)
variable specific optimized α (ranges from 3.77 to 7.57)	15.34 (16.80)
	15.11 (16.93)

Figure 1: A Framework for Determining the Number of Informants

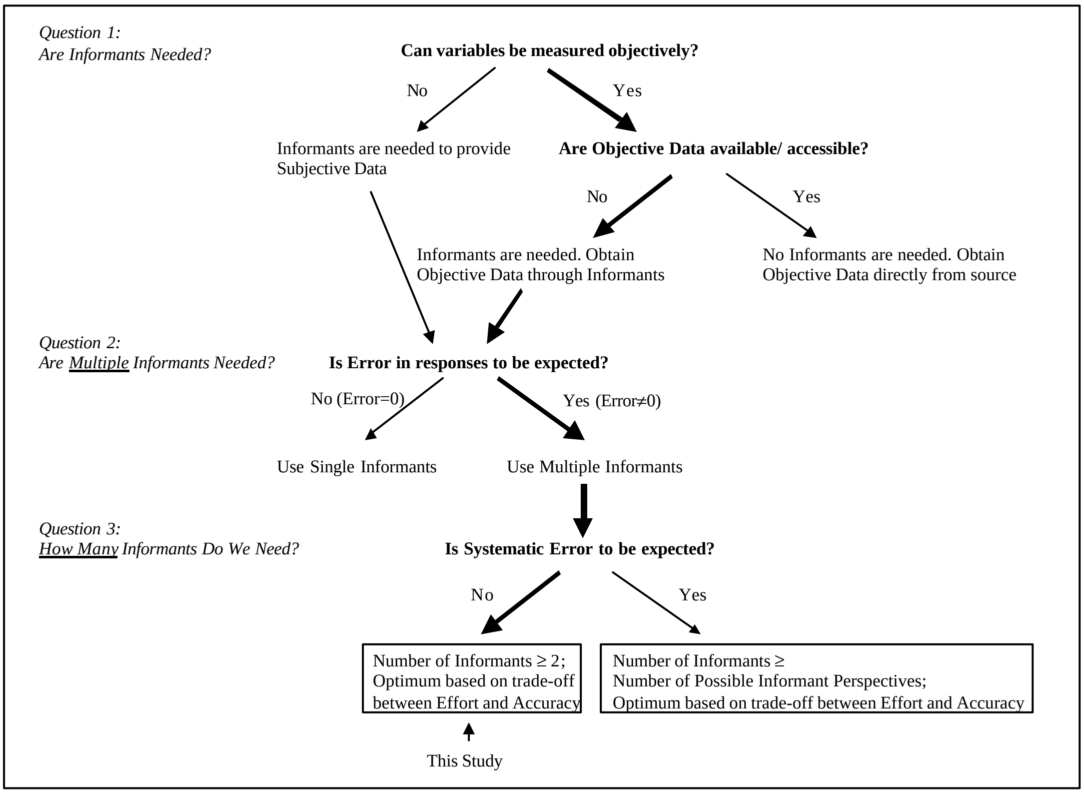


Figure 2: Mean Absolute Percentage for Different Aggregation Approaches in Study 1 (showing that confidence-based weights outperform data based weights which in turn outperform simple averages).

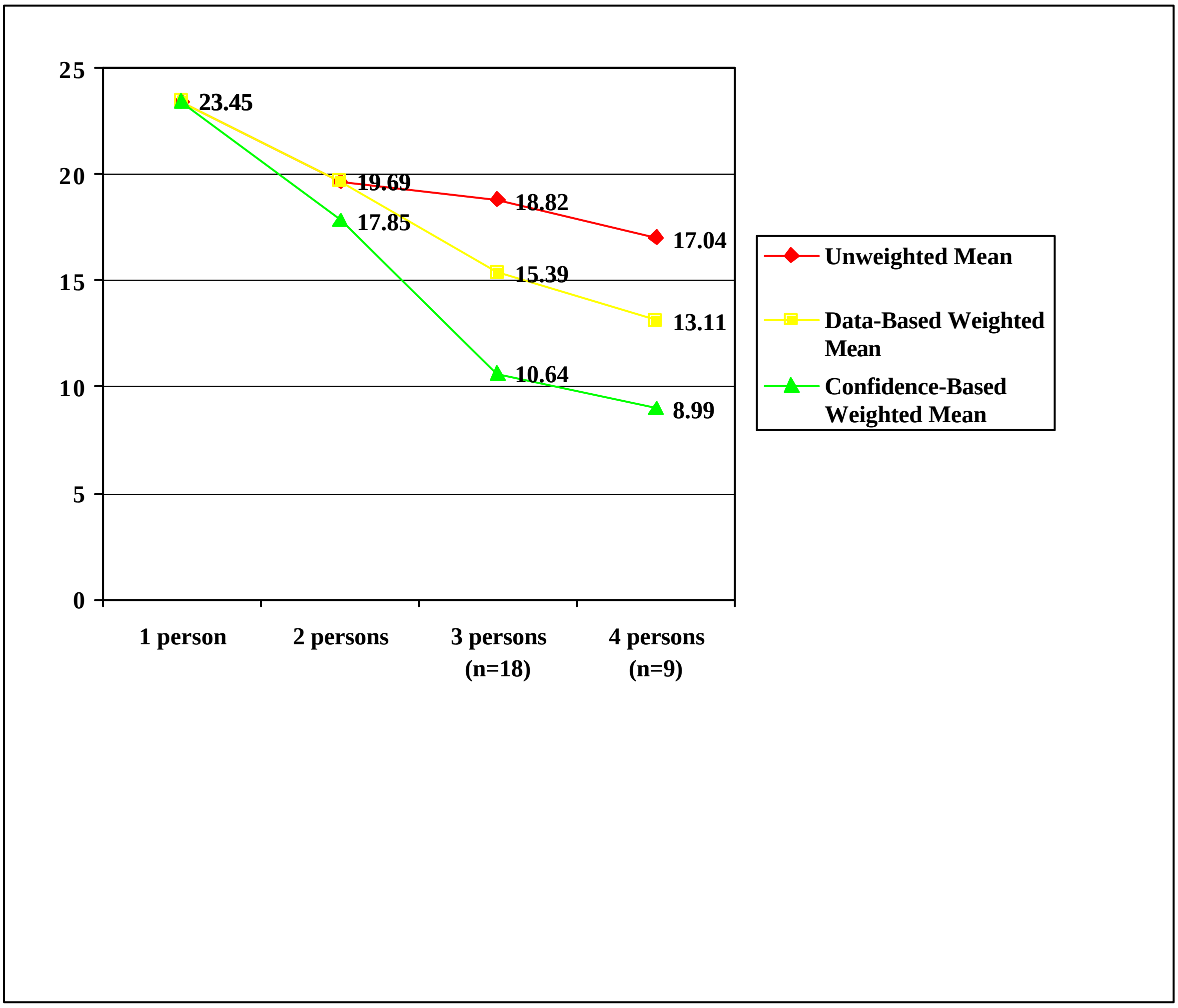


Figure 3: How MAPE decreases with increasing respondent confidence. (for Study 1 Data)

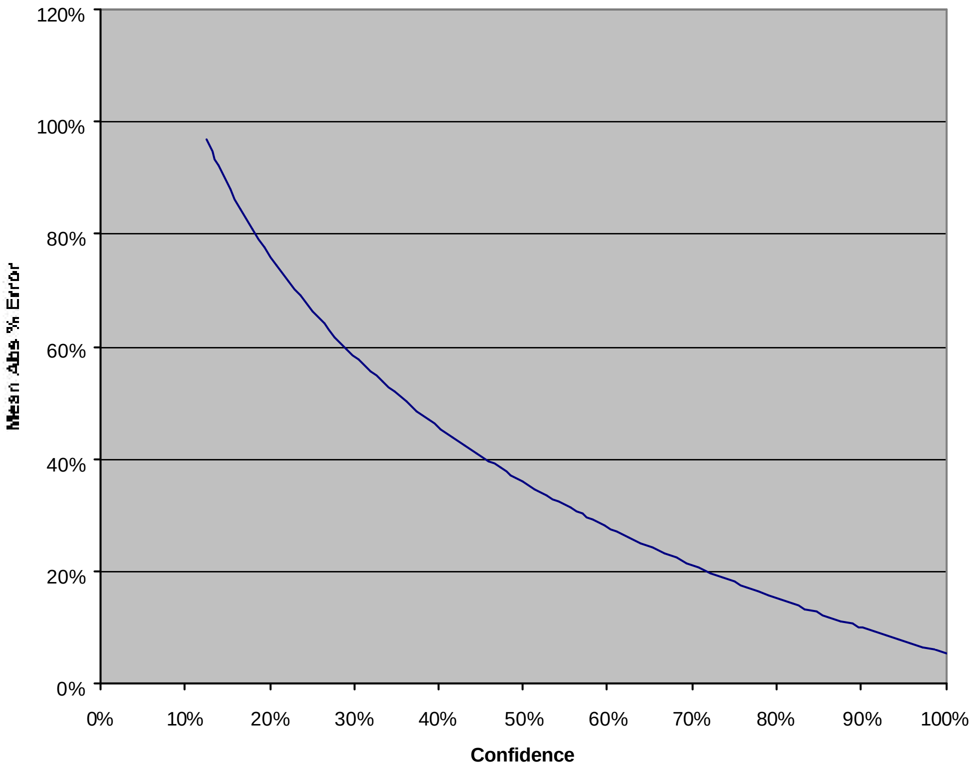


Figure 4: Empirical distribution of degree of confidence across the Study 1 sample respondent population.

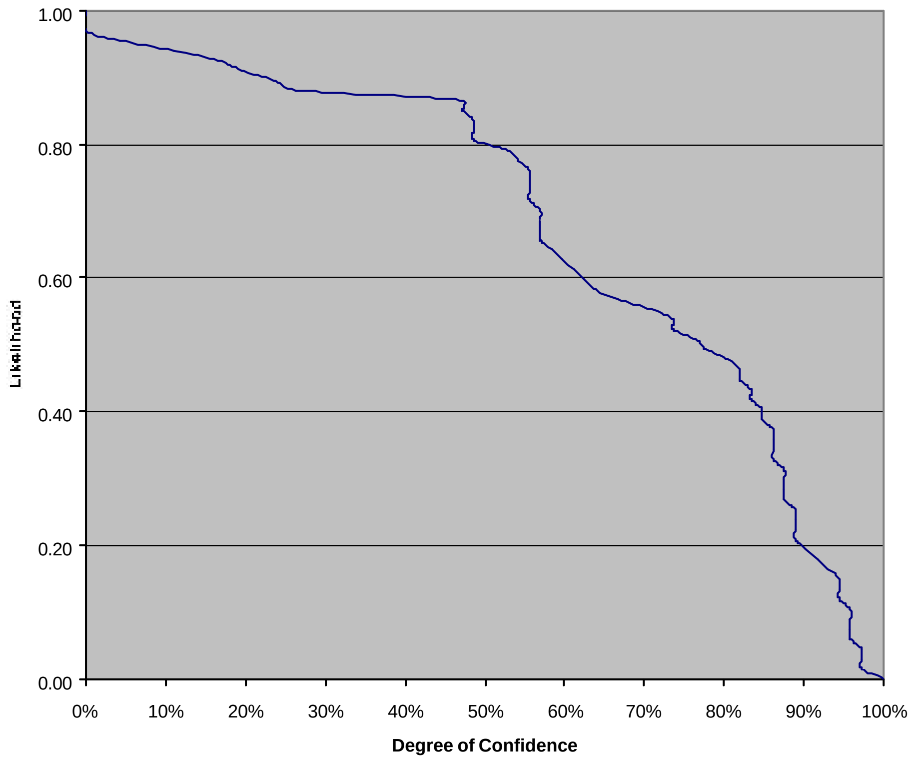


Figure 5: Expected sample size needed, from Figure 3 results, to screen for a respondent with a target degree of confidence.

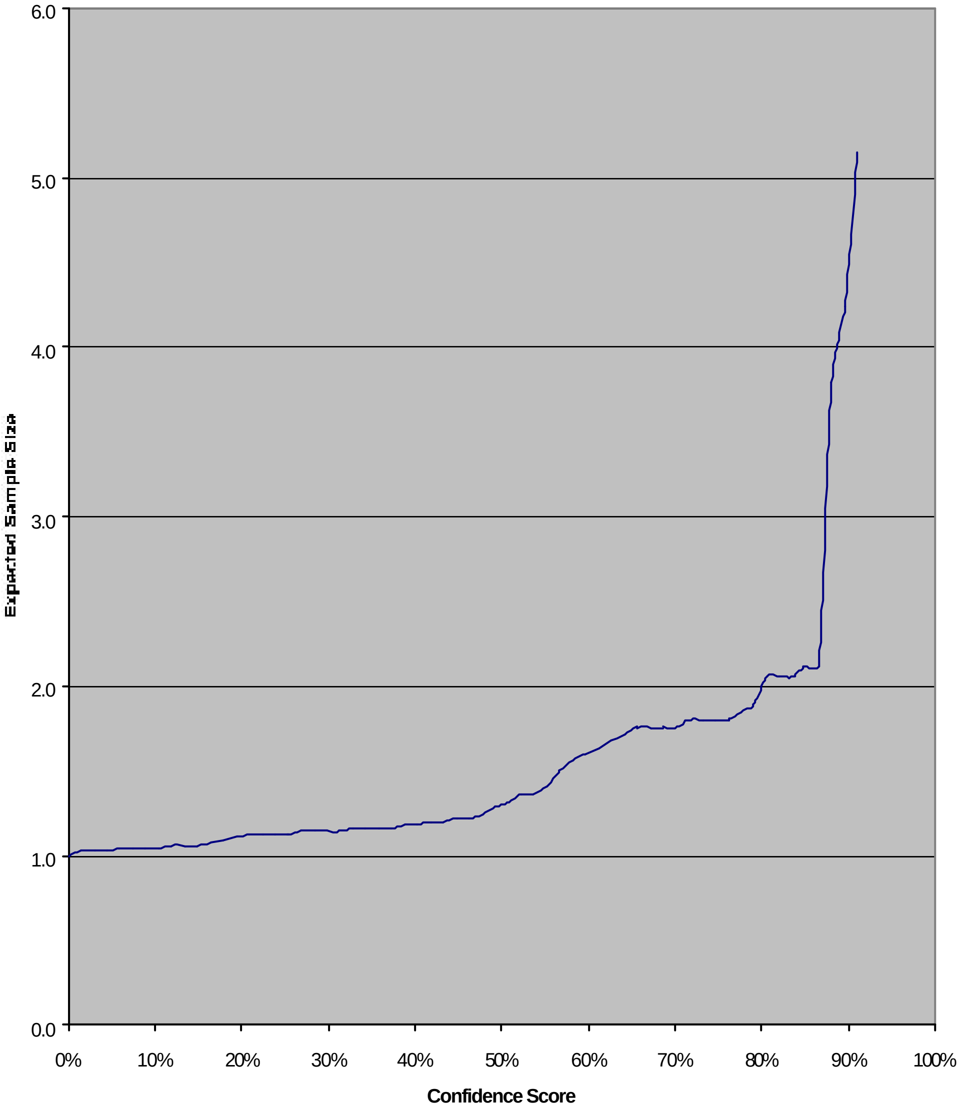
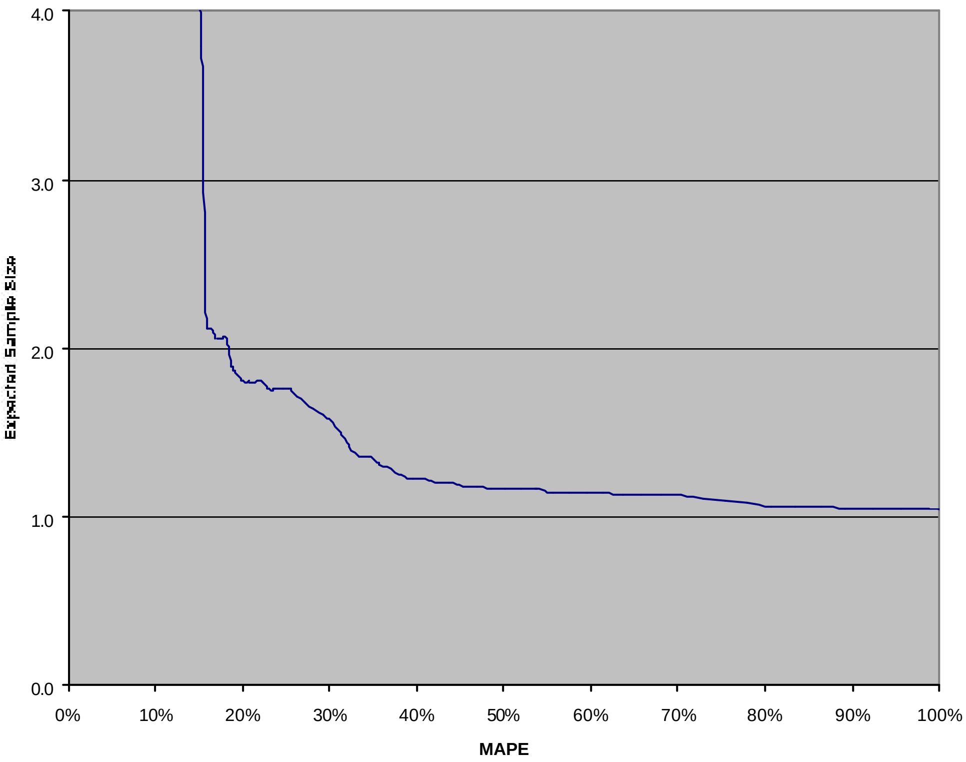


Figure 6: Expected sample size needed, from Figures 3, 4 and 5 to screen for a single respondent to achieve a target MAPE.



APPENDIX A: INSTRUCTIONS AND QUESTIONS FOR SUBJECTS – STUDY ONE

Perceived Marketing Mix-Performance Relationship Study

We are studying the perceived relationships between marketing-mix expenditures and performance. As part of this study, we would like you to provide us with your estimates of the following marketing-mix variables and performance indicators for your MARKSTRAT company.

We would like to obtain your estimates. Please do not use MARKSTRAT materials and do not speak with your MARKSTRAT teammates to assist in completing this survey. Also, please do not skip over to the next page without completely answering all questions on the page you are on. Even if you do have difficulties in recalling the exact figures, please provide us with your best estimates.

Please give us the following information with respect to your firm's decisions in the current period i.e. (period 3). In addition, please indicate how certain you are about the accuracy of each of your estimates, using a 1-9 scale where 1 indicates 'not certain at all' and 9 indicates 'completely certain'.

		CERTAINTY (1-9)
Brand 1:__________________		
Advertising ($k)	_____________	_____________
Retail price ($)	_____________	_____________
Brand 2:__________________		
Advertising ($k)	_____________	_____________
Retail price ($)	_____________	_____________
Overall:		
Salesforce		
Channel 1	_____________	_____________
Channel 2	_____________	_____________
Channel 3	_____________	_____________
Budget ($k)	_____________	_____________

APPENDIX B: INSTRUCTIONS AND QUESTIONS FOR SUBJECTS – STUDY TWO

Perceived Marketing Mix-Performance Relationship Study

We are studying the perceived relationships between marketing-mix expenditures and performance. As part of this study, we would like you to provide us with your estimates of marketing-mix variables and performance indicators for your MARKSTRAT company.

We would like to obtain your estimates. Please do not use MARKSTRAT materials and do not speak with your MARKSTRAT teammates to assist in completing this survey. Also, please do not skip over to the next page without completely answering all questions on the page you are on. Even if you do have difficulties in recalling the exact figures, please provide us with your best estimates.

Please give us the following information with respect to your firm's most recent decisions i.e., the decisions you have submitted today. In addition, please indicate how certain you are about the accuracy of each of your estimates, using a 1-9 scale where 1 indicates 'not certain at all' and 9 indicates 'completely certain'.

CERTAINTY (1-9)

Brand 1:__________________

Advertising ($k) _____________ _____________

Retail price ($) _____________ _____________

Brand 2:__________________

Advertising ($k) _____________ _____________

Retail price ($) _____________ _____________

Overall:

a) Salesforce

Channel 1 _____________ _____________

Channel 2 _____________ _____________

Channel 3 _____________ _____________

b) Budget ($k) ____________ ____________

Please give us the following information with respect to your expectations about what your firm's performance will be, as a consequence of the decisions you have submitted today. In addition, please indicate how certain you are about the accuracy of each of your estimates, using a 1-9 scale where 1 indicates 'not certain at all' and 9 indicates 'completely certain'.

CERTAINTY
(1-9)

Brand 1: ________________

Expected Sales (units) ___________ ___________

Expected Brand Awareness (%) ___________ ___________

Brand 2: ________________

Expected Sales (units) ___________ ___________

Expected Brand Awareness (%) ___________ ___________

Expected Budget for Next Period ($k) ____________ ___________

Please indicate how certain you are, overall, about your recollection of your current decisions and your forecast of the likely outcomes of your current decisions. Please use a 1-9 scale where 1 indicates 'not certain at all' and 9 indicates 'completely certain'.

CERTAINTY
(1-9)

Recall of current decisions ____________

Forecast of outcomes from current decisions ____________

Endnotes

[1] Several labels have been used to refer to this issue – e.g., aggregation, synthesizing, opinion pooling, merging, compromising, and consensus building (Lipscomb, Parmigiani, and Hasselblad 1998). We will use the term *aggregation* exclusively here.

[2] In spite of problems associated with these retrospective informant reports (Huber and Power 1985), they have been used extensively in research in marketing.

[3] In using the latent trait approach, structural equation techniques are used to model reports from multiple informants to reflect latent constructs (Kumar, Stern, and Anderson 1993). The latent trait approach is primarily applicable in situations where variation in informants' backgrounds or roles exist and is expected a priori to affect responses and thus lead to structural error (e.g., members of different departments in cross-functional groups; informants at different hierarchical levels in the organization). However, this approach can result in ill-fitting models and poor estimates in the absence of perceptual agreement among the multiple informants (Anderson and Narus, 1990; Bagozzi, Yi and Philips, 1991). Since we focus on situations in which we do not a-priori expect non-zero structural error we leave do not consider the latent-trait approach here.

ERASMUS **R**ESEARCH **I**NSTITUTE OF **M**ANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

Publications in the Report Series Research* in Management

Impact of the Employee Communication and Perceived External Prestige on Organizational Identification
Ale Smidts, Cees B.M. van Riel & Ad Th.H. Pruyn
ERS-2000-01-MKT

Critical Complexities, from marginal paradigms to learning networks
Slawomir Magala
ERS-2000-02-ORG

Forecasting Market Shares from Models for Sales
Dennis Fok & Philip Hans Franses
ERS-2000-03-MKT

A Greedy Heuristic for a Three-Level Multi-Period Single-Sourcing Problem
H. Edwin Romeijn & Dolores Romero Morales
ERS-2000-04-LIS

Integer Constraints for Train Series Connections
Rob A. Zuidwijk & Leo G. Kroon
ERS-2000-05-LIS

Competitive Exception Learning Using Fuzzy Frequency Distribution
W-M. van den Bergh & J. van den Berg
ERS-2000-06-LIS

Start-Up Capital: Differences Between Male and Female Entrepreneurs, 'Does Gender Matter?'
Ingrid Verheul & Roy Thurik
ERS-2000-07-STR

The Effect of Relational Constructs on Relationship Performance: Does Duration Matter?
Peter C. Verhoef, Philip Hans Franses & Janny C. Hoekstra
ERS-2000-08-MKT

Marketing Cooperatives and Financial Structure: a Transaction Costs Economics Analysis
George W.J. Hendrikse & Cees P. Veerman
ERS-2000-09-ORG

---

* ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Decision Making in Marketing Management
F&A Financial Decision Making and Accounting
STR Strategic Renewal and the Dynamics of Firms, Networks and Industries

A Marketing Co-operative as a System of Attributes: A case study of VTN/The Greenery International BV,
Jos Bijman, George Hendrikse & Cees Veerman
ERS-2000-10-ORG

Evaluating Style Analysis
Frans A. De Roon, Theo E. Nijman & Jenke R. Ter Horst
ERS-2000-11-F&A

From Skews to a Skewed-t: Modelling option-implied returns by a skewed Student-t
Cyriel de Jong & Ronald Huisman
ERS-2000-12-F&A

Marketing Co-operatives: An Incomplete Contracting Perspective
George W.J. Hendrikse & Cees P. Veerman
ERS-2000-13– ORG

Models and Algorithms for Integration of Vehicle and Crew Scheduling
Richard Freling, Dennis Huisman & Albert P.M. Wagelmans
ERS-2000-14-LIS

Ownership Structure in Agrifood Chains: The Marketing Cooperative
George W.J. Hendrikse & W.J.J. (Jos) Bijman
ERS-2000-15-ORG

Managing Knowledge in a Distributed Decision Making Context: The Way Forward for Decision Support Systems
Sajda Qureshi & Vlatka Hlupic
ERS-2000-16-LIS

Organizational Change and Vested Interests
George W.J. Hendrikse
ERS-2000-17-ORG

Strategies, Uncertainty and Performance of Small Business Startups
Marco van Gelderen, Michael Frese & Roy Thurik
ERS-2000-18-STR

Creation of Managerial Capabilities through Managerial Knowledge Integration: a Competence-Based Perspective
Frans A.J. van den Bosch & Raymond van Wijk
ERS-2000-19-STR

Adaptiveness in Virtual Teams: Organisational Challenges and Research Direction
Sajda Qureshi & Doug Vogel
ERS-2000-20-LIS

Currency Hedging for International Stock Portfolios: A General Approach
Frans A. de Roon, Theo E. Nijman & Bas J.M. Werker
ERS-2000-21-F&A

Transition Processes towards Internal Networks: Differential Paces of Change and Effects on Knowledge Flows at Rabobank Group
Raymond A. van Wijk & Frans A.J. van den Bosch
ERS-2000-22-STR

Assessment of Sustainable Development: a Novel Approach using Fuzzy Set Theory
A.M.G. Cornelissen, J. van den Berg, W.J. Koops, M. Grossman & H.M.J. Udo
ERS-2000-23-LIS

Creating the N-Form Corporation as a Managerial Competence
Raymond vanWijk & Frans A.J. van den Bosch
ERS-2000-24-STR

Competition and Market Dynamics on the Russian Deposits Market
Piet-Hein Admiraal & Martin A. Carree
ERS-2000-25-STR

Interest and Hazard Rates of Russian Saving Banks
Martin A. Carree
ERS-2000-26-STR

The Evolution of the Russian Saving Bank Sector during the Transition Era
Martin A. Carree
ERS-2000-27-STR

Is Polder-Type Governance Good for You? Laissez-Faire Intervention, Wage Restraint, And Dutch Steel
Hans Schenk
ERS-2000-28-ORG

Foundations of a Theory of Social Forms
László Pólos, Michael T. Hannan & Glenn R. Carroll
ERS-2000-29-ORG

Reasoning with partial Knowledge
László Pólos & Michael T. Hannan
ERS-2000-30-ORG

Applying an Integrated Approach to Vehicle and Crew Scheduling in Practice
Richard Freling, Dennis Huisman & Albert P.M. Wagelmans
ERS-2000-31-LIS