

**On The Use of Regression and Verbal Protocol Analysis in Modeling Analysts'
Behavior in an Unstructured Task Environment: A Methodological Note**

M. J. Anderson

Michigan State University

G. S. Potter

Cornell University

Data availability: General data can be obtained from the authors upon request, including a copy of the instrument.

We would like to thank Michigan State University for financial support of this project.

ABSTRACT

We use two techniques—linear regression (LR) and verbal protocol analysis (VPA)—to model the decision behavior of four experienced financial analysts performing an unstructured valuation task. The research examines the structure and implied data use for each model type. We also assess the conclusion of prior research that the two techniques are complementary. This research also claims that these techniques are interchangeable in most decision contexts, in effect treating them as substitutes. Our results suggest that the techniques are complementary; however, each adds unique contributions to model development.

On The Use of Regression and Verbal Protocol Analysis in Modeling Analysts' Behavior in an Unstructured Task Environment: A Methodological Note

Recent judgment and decision making research in accounting has emphasized the interplay between subject characteristics and tasks (Johnson, 1983; Libby & Luft, 1993). In part, this emphasis stems from the contention that decision behavior cannot be improved without understanding the nature of the processes underlying the behavior (Libby, 1990; Hogarth, 1991). As stated by Hogarth (1991, p. 279): “You cannot improve a process unless you know what it is.” Both the decision task and the decision maker must be analyzed and understood (Hogarth, 1991, pp. 283-284).

A consequence of this emphasis has been increased use of relatively complex, “engaging” experimental tasks, and more experienced subjects (Johnson, 1983; Abdolmohammadi & Wright, 1987). We model the decision behavior of experienced financial analysts in a relatively complex (unstructured) task environment—the analysis of initial public offerings (IPOs).¹ This study extends Anderson (1988). Anderson (1988) presents a verbal protocol analysis (VPA) of the examination of an initial public offering prospectus by professional and student financial analysts. The study focused on the nature of the information search process and differences between the groups. Issues of interest included the time to decision, the number and types of information items addressed by the analysts, the structure of the search process, and the overall conclusion drawn about the case firm.

¹ A task's degree of structure is a measure of the adequacy of an individual's knowledge of how to solve it, given general human knowledge and capacities. The subject must extract information from the task and reformulate it in an appropriate way (Hayes & Simon, 1976; Smith, 1988). The definition implies that structure is not an independent characteristic of the task itself.

This study differs from Anderson (1988) in that the focus is on decision models and methodologies. VPA is used to develop formal computer (program) models of the decision behavior of each subject. The structure and performance of the VPA models are compared to linear regression (LR) models in a two step process. First, the VPA models are compared to LR models based on subjects' retrospective reports and a correlation analysis of the data set. Each VPA model is then recast as a LR model and compared to the best LR model from step 1.

We address an issue raised by Eihhom *et al.* (EKK) (1979) and Harte and Koele (1995). These authors suggest that linear modeling (LR) and VPA techniques are generally complementary, with differing strengths and weaknesses. However, EKK (1979) and Harte and Koele (1995) conclude that models and results from the techniques are often indistinguishable, in effect treating the techniques as substitutes. The generality of this conclusion across differing task environments is not addressed by EKK (1979) or Harte and Koele (1995). We argue, given the nature of the linear model, that it is unlikely that VPA and LR models will be consistent in unstructured task environments (Dawes & Corrigan, 1974). Many important tasks in which accounting data play a significant role (e.g. the valuation of initial public offerings) are relatively unstructured (Hogarth, 1991; Libby & Luft, 1993).

Our results support a negative response to the question of model convergence when the LR and VPA models are considered separately. LR and VPA models are substantively different, in terms of variance accounted for and face validity, in this task environment. Models developed using both regression and verbal protocol techniques

perform best. The VPA models play a significant role in determining the structure of the best performing models.

THEORY AND LITERATURE REVIEW

It has been demonstrated that task structure is a major determinant of observed decision behavior (Hayes & Simon, 1976; Einhorn & Hogarth, 1981). Further, the nature and structure of a task impacts directly the choice of subjects and analysis techniques in a given study. Issues relevant to the choice of analysis technique(s) include the efficiency and/or effectiveness of the technique, the nature of output information, and how the information can be used to assess the issue(s) of interest. We review relevant parts of the LR and VPA literatures, along with literature from finance dealing with the nature of the IPO environment.

EKK (1979) exploit the generality of the linear model. Extending the notion of vicarious functioning (Brunswik, 1952), EKK assert that it is not surprising that the LR model is effective in predicting human judgment. They note that the human decision maker *must* adapt to a probabilistic environment. Information must be collected and evaluated. This process appears to be linear in nature; invariably, judgments of subjects are more linearly predictable than the criterion variable as it exists in nature (EKK, 1979). The very robustness of such models suggests that they are capturing something fundamental about human judgments. EKK also note that information tends to be redundant in natural environments; search can be limited without large losses in

predictive accuracy. Problems may exist due to overconfidence and the inability to reliably make tradeoff among cues. They go on to demonstrate that verbal protocol data can be used to generate LR or VPA models which are structurally similar, and statistically indistinguishable. EKK (1979) essentially conclude that VPA and LR techniques are substitutes in their decision context.

Harte and Koele (1995) reexamine the issues raised by EKK (1979), addressing attribute weights, data reliability and model face validity. Harte and Koele (1995) use LR, VPA and a retrospective report technique to develop decision models which are statistically equivalent. They use median analysis to demonstrate that attribute weights (and hence structure) of decision models are equivalent across the techniques. Harte and Koele conclude that the results from use of the techniques are indistinguishable. Finally, Harte and Koele use the LR model as a performance bound for the other models, in effect assuming that no model could outperform the LR models. This use of the LR model implicitly assumes fixed data requirements and a fixed structure across the techniques.

Ford *et al.* (1989) take a different perspective than EKK (1979) to argue that what is fundamental in many decision contexts is not clear. In their view, the robustness of LR models suggests that the underlying processes do not matter. Differing processes may be consistent with a given LR model, with differing implications for decision making. Ford *et al.* argue further that the relationship between criterion and predictors in much of the research is typically linear *by design*.

In theory, given that the models are based on the same information set and behavior, VPA and statistical models should be similar. However, VPA generally involves a different type of analysis, which often involves the assessment of different

cues and greater amounts of data than does a LR model. The implications for a given VPA research study include reduced sample sizes, restricted variable ranges and reduced statistical power. Further, absent an error theory, it is unclear what value can be placed on the increased amount of data available (EKK, 1979).²

Increased amounts and different types of data also have implications for LR models in unstructured task environments. Johnson *et al.* (JMG) (1989) provide evidence that in situations where the number of decision variables is increased or variables are negatively related, LR models may be misspecified and misleading.³ Model misspecification generally implies degradation of performance (JMG, 1989).

How likely are these conditions to be operative in the initial public offering setting? Relationships among variables may be unclear given the often limited history of such firms. However, variables may be negatively related. For example, a firm may have a positive value with negative net income, or high sales or income growth with decreased cash balances. Relationships among variables may be volatile (Slovic, 1969). Firms tend to be relatively high risk.

Insights about the potential impact of such issues is provided by Selling and Shank (1989). They demonstrate that models developed from LR and VPA may provide both complementary and divergent information. Using student subjects assessing financial information, they find that the structure of VPA and LR models often differs. The results imply that VPA models have greater face validity than related LR models, but VPA models predict less well. Their task was relatively well structured.

² It is generally accepted that concurrent verbal protocol analysis can provide reliable information about predecision behavior (Anderson, 1985; Russo et al., 1989; Biggs et al., 1993; Harte & Koele, 1995).

³ EKK (1979) also allude to this result in their model discussion.

Larcker and Lessig (1983) also suggest that valuation models can be beneficially structured based on the interplay between verbal protocol data and regression models. They develop LR models of subjects making an investment decision. Subjects were provided general macroeconomic data and six pre-selected cues specific to each firm evaluated. Each subject also supplied retrospective protocols of their decisions after completing the experimental task. Model performance improved with use of the protocols. Interestingly, inter-correlations among many of the variables were negative. A majority of the subject models were in fact unreliable, consistent with JMG (1989). Given that Larcker and Lessig's (1983) task was relatively well-structured, the unreliability of the LR models is noteworthy (Dawes, 1988). Larcker and Lessig (1983) use a less stringent criterion to conclude that the LR models were reliable.

That different models may be derived in an unstructured task environment when using LR and VPA techniques is demonstrated by Biggs (1984). Bigg's subjects, financial analysts, were asked to assess "earning power." Most prepared a projective model based on future expected earnings. The projective component of the models did not appear in the data set provided to subjects; it is unclear how one would develop such a LR model absent, or even with, the verbal protocols.

Little work has been done in the IPO context at the individual level. Research to date has typically addressed financial analysis (e.g. Slovic, 1969; Bouwman, 1984); Anderson, 1988 is one of the few studies dealing specifically with IPOs. Anderson's (1988) work suggests that many variables may be assessed by analysts when evaluating an IPO. These included the firm's principal product, the nature of the customer base, the growth rate and volatility of firm sales and net income, backlogs, and the firm's

competition. Factors evaluated tended to be idiosyncratic, with differing cues used by each analyst. Analysis tended to be based on earnings, not balance sheet items.

Much of the literature concerned with IPOs comes from Finance and involves aggregate market effects around the IPO offer date. The studies mainly focus on two issues: (1) the Leland and Pyle signaling hypothesis—that selling shareholders signal the value of the offering firm by their willingness to hold an equity position (Leland & Pyle, 1977), and (2) its corollary, the underpricing hypothesis—that offering firms signal their quality by setting a selling price that is lower than warranted by the data Gain, 1996). Both hypotheses have been confirmed repeatedly in the literature (Downes & Heinkel, 1982; Ritter, 1984; Jain, 1996). These studies confirm the difficulty of assessing data relationships in the IPO context. The potential role of the signaling or underpricing hypothesis for individual analysis is unclear. Subjects in this study were provided with information on the equity position of selling shareholders.

Taken together, what do these studies imply about the interplay between LR and VPA models, particularly in the IPO setting? The EKK (1979), Harte and Koele (1995), and Larcker and Lessig (1983) results suggest that VPA and LR techniques may be substitutes (the Larcker and Lessig results suggest some complementarity). Selling and Shank (1989) provide evidence that the results from the two techniques may diverge, with the VPA models having greater validity, but predicting less well. Anderson (1988) demonstrates that VPA can be used to develop a dense “track” of decision behavior in the IPO context. Overall, the question of whether VPA and LR techniques are substitutes or complements in the IPO environment remains an unresolved issue. We address the following hypothesis stated in alternative form:

HI: Models based on verbal protocols will have smaller prediction errors and greater validity than LR models in an unstructured task (IPO) environment.

METHOD

Subjects

Four certified financial analysts (CFAs) were solicited for the experiment. Three held senior positions (vice presidents) with investment underwriting firms; the fourth held a senior investment position with a large bank. All had experience in the valuation and issuance of IPOs. The average age was approximately thirty-two years with an average 7.5 years of professional experience.

Design

Subjects were placed in a room with an experimenter, one at a time. They were told that they would be tape recorded during the session. After completing a short practice “think-aloud” exercise, they priced twenty case firms—ten while “thinking aloud” (verbal condition), and ten without concurrent verbalization (silent condition). The conditions (with or without verbalization) were counterbalanced, cases and case order were randomly assigned to condition, and the same ten cases were used in a given condition for all subjects. Finally, two cases were scaled by a numerical factor and repeated for reliability and validity analysis.

On average, three hours were used to complete the task. After completing the cases, subjects completed a debriefing questionnaire. All subjects were paid \$20 for participating, but each could earn up to \$30 additional by having the smallest cumulative

prediction error relative to the actual initial offering prices. Earnings ranged from \$20-\$50.⁴

Task

The pricing of initial public offerings (IPOs) of stock is an important task for investment underwriters and is reserved for the senior analysts in a firm (Halloran, 1979). No market price exists *a priori*, and the offering price must be developed by the issuer and the underwriting firm. The price reflects the capital needs of the client firm, tempered by the realities of the marketplace (Halloran, 1979). Factors generally viewed as important include the firm's product, net asset and earnings history, general and industry market conditions, market share, the proportion of shares retained by selling shareholders and backlogs. Accounting data play a central role in the task.

Cases used in this study were based on disguised versions of actual IPOs and were three to five pages in length. Case firms were drawn from several industries including the computer, defense, biotechnology, and communications industries. Firm data provided included a description of its business, several years of sales, EPS, backlogs, and growth rates, along with similar data for a disguised competitor firm matched (using SIC and Moody's codes) on size and general line of business. Stock prices of the competitor firm and trend information on the general movement of the Dow Jones Industrial Average for the quarter prior to the offering were also supplied. Each of the analysts in the study specialized in one or more of the industries included. Subjects indicated that, generally, the information required to make a decision was available in the case materials.

⁴ Payments were not consistent with subjects' salaries. However, the objective was to make the activity more goal oriented. No subject refused any payment.

Model development

Both LR and VPA models were constructed for each subject. Two LR models were developed—one based on the subject’s debriefing questionnaire (RRM model) and one based on a correlation analysis of the data set (CRM model). The RRM model was constructed by regressing each analyst’s predicted prices on the variables listed on his debriefing questionnaire (see Table 1). The CRM model was comprised of the variables most highly correlated with each analyst’s predicted prices. These variables, which came from a review of the experimental materials, each analyst’s protocols and a search of the financial analysis literature, were used as regressors in each analyst’s model. Both the RRM and CRM models are constructed by minimizing mean square error. Note that predictors in the RRM models are a subset of the predictors in the CRM models. The LR models were generated using the same ten cases used in generating the VPA models.

VPA models were developed from each subject’s tape recorded concurrent verbal protocols. The protocols were transcribed and coded, using the operators shown in Table 2.⁵ Operators are mechanisms for producing changes in information, and are primarily concerned with the processes of discrimination, testing and comparisons (Newell & Simon, 1972; Anderson, 1984).

The coded protocols were used to construct a problem behavior graph (PBG) for each case for each subject (Newell & Simon, 1972; Anderson, 1984; Ericsson & Simon, 1984). The PBG consists of an interconnected network of nodes comprised of the inputs

⁵ Two coders were used. The Kappa coefficient (Cohen, 1960), which corrects for chance agreement, was used to assess inter-rater reliability. Coefficients ranged from 0.50 to 0.74. Raw agreement over the nine operators ranged from 0.62 to 0.85. Differences between coders were resolved by discussion and a consensus reached.

and outputs of operators. Essentially a tree representation, it is a mapping of the problem solving behavior of the analyst. Each PBG was then translated into a flowchart for purposes of devising a computer algorithm. The translation involved substituting if-then-else rules for the choices made by the subjects (Newell & Simon, 1972; Luger & Stubblefield, 1989).

The 10 flowchart models for each analyst were consolidated to produce a single “macro” flowchart. Operationally, the process involved retaining both repetitive and conditional problem solving activity. For example, analyst A3 projects future sales and an after-tax profit margin for each firm. An EPS number is then computed by dividing projected profits by the shares to be outstanding after the offering.⁶ However, if a firm had recent losses, the most recent period of *positive* earnings was used to derive an *achieved* relationship between sales, related costs and taxes. This relationship was used to derive a price per share. Both types of behavior were retained in the flowchart. The flowchart model was used to generate a computer program (VPA) model for each analyst (Newell & Simon, 1972). Each VPA model was validated using a split-half sample procedure. The flowchart model for analyst A3 is provided as Fig. 1.

The last stage of the modeling process involved pooling the twenty cases and re-estimating model parameters. Two additional LR models were constructed for each subject: (1) one based on the best fitting CRM model, and (2) another based on the predictors and structure implied by the VPA-derived computer model. These models permit an assessment of the effect of larger sample sizes on the LR models, and a further

⁶ A3’s behavior is typical in that all analysts used earnings based models (Hawkins & Campbell, 1978).

assessment of the extent to which the LR and VPA approaches are substitutes or complements in this task context.⁷

RESULTS AND DISCUSSION

Model evaluation metrics and data consistency tests

Models were evaluated using both deviation (prediction error) and correlation metrics. Deviation or prediction error was measured as the difference between model predictions and each analyst's predicted prices; both mean squared and mean absolute error metrics were used. Deviation measures were used for several reasons: (1) Regression models are generally assessed based on deviation measures (e.g. sums of squares), (2) subjects were rewarded based on a deviation measure—the difference between judged price and the actual initial offering price, and (3) deviation measures can be systematically related to other measures, such as correlation.

A correlation metric was also used because it provides a measure of the covariability of the analysts' predictions and model predictions. Our principal measure, Spearman's rho, is an indicator of the relative ranking, across the analysts, of differences between each analyst's predicted prices and his model predictions.⁸ This is important since one of the claims of VPA is that it better captures the decision process of problem solvers. One measure of "capture" is the extent to which model predictions move with the output of the analysts.

⁷ One is generally better off with more data in either case. The data requirements of the LR approach are *relatively* greater.

⁸ The Larcker and Lessig criterion was the number of subject judgements correctly predicted by their models. We use a continuous criterion variable.

Our analysis focuses on the models and observed differences in prediction performance (both deviation and correlation). Panel A of Table 3 depicts mean, deviation and correlation results between the actual and predicted prices for each analyst for the twenty cases. The correlation between predicted and actual initial offering prices (IOP) range from 34 to 71%, and is significant for three of the four analysts. The final two rows of Panel A provide mean square and mean absolute error measures. While the correlation results indicate that A1's judgments were most highly related to the IOPs, analyst A2's prediction errors are the smallest, suggesting greater accuracy.

Panel B of Table 3 displays the prediction errors of the analysts by verbalization condition (silent vs speak aloud). The Mann Whitney test ($N = 20$) by analyst indicated that the method of analysis did not lead to a difference in the predictive accuracy of the analysts.

Model assessment

The RRM and CRM models of the analysts from the ten verbal cases are presented in Tables 4 and 5. Table 4 displays the RRM results of the analysts based on their questionnaire responses. For analysts A1 and A3, the models are not statistically significant. The models contain few statistically significant variables.⁹ Inferences about the underlying process or economic significance of the analysts' models are difficult to draw.

⁹ The expectation was no more than three predictors in a LR model (Dawes & Corrigan, 1974). More than three predictors *decreased* predictive validity.

Table 5 presents the CRM models. For three of the analysts, the cues differ from the RRM model. This is somewhat consistent with the literature that suggests that retrospective reports of decision makers tend to lack validity (Russo *et al.*, 1989).

Hypothesis

The cross validation prediction performances of the models are presented in Table 6. These results directly address the research question of our study. Each analyst's derived model is used to predict his assessed prices for the remaining ten cases.

For the prediction error measures, differences among the models are not statistically significant. For both pairwise (Mann Whitney U; $N = 20$) and overall tests across models (Kruskal Wallace; $N = 30$) for each analyst, in only one instance was a difference (between the RRM and VPA models of A3) significant. The data do suggest the power of the LR procedures. The RRM (analysts A1 and A3) and CRM (analyst A3) models are statistically insignificant; yet, they predict as well as the VPA model. Overall, the hypothesis of no difference in prediction performance cannot be rejected.

The correlation analysis in Panel B provides a validity measure. Results suggest performance differences. Price predictions for the VPA models are significantly correlated with the analysts' predicted prices for each analyst. In only two cases do the price predictions of the LR model significantly correlate with those of the related analyst. For one of those cases (analyst A2), the magnitude of the LR model correlation exceeds that of the VPA model. However, as noted, the VPA model correlation for A2 is significant. Taken together, the prediction error and correlation data suggest that, while the LR models are predicting amounts that are relatively close to the analysts' prices, the

predictions are not moving systematically with the analysts' predictions. The results suggest greater predictive validity for the VPA models.¹⁰

The prediction performance analysis does not shed light on model structure and cue selection by the analysts. As can be inferred from Tables 1, 4 and 5, and Fig. 1, many, often differing cues are potential candidates for use by subjects. Closer analysis of the flowchart of analyst A3 (Fig. 1) is instructive. A3 uses an earnings capitalization model, but also employs heuristic rules (e.g. the sales of Arms going public should generally double on a yearly basis, dependent on firm age and size). To further examine the descriptive validity of the VPA models, the cues from each VPA model were recast as a LR model in the combination suggested by the protocols. The resulting regression models are reported in Table 7. The per share pricing of the analysts implied a multiplicative model, necessitating a log transformation. If the VPA models were perfect representations of the decision process, then the shares coefficient should equal -1 , the constant coefficient 0, and the rest of the coefficients should equal 1.

Analysis of the LR model for A3 does suggest face validity. The coefficients are significantly different from zero in the direction suggested by the protocols, and also make economic sense. Only some of the coefficients are close to 1 or -1 (suggestive of error). The conditional component from A3's flowchart model is not statistically significant. These points can also be made for analyst A2.

The VPA-derived LR models imply that analysts A1 and A4 do not use an earnings model in a strict sense. A4's model implies use of a modified price-earnings and growth algorithm. A1 appears to use a model based on a price-earnings ratio adapted

¹⁰ The test-retest reliability (repeated cases) of the analysts *as a group* is mixed. For two analysts, the correlation is 0.95; for the other two -0.48 .

from that of the competitor firm, combined with a price heuristic when a firm had negative or low earnings. Three analysts—A1, A2 and A4—vary between an earnings model and a levels model (e.g. book value) depending on the profitability of the firm. Note that, consistent with Anderson (1988), all analysts used some variation of an earnings-based model.

Several of the predictors in analysts' models have negative coefficients. This is in part structural, given the per share pricing techniques of the analysts. However, this may help to explain the failure to find a significant RRM model for A3. This result is consistent with the findings of JMG (1989).

Table 8 repeats the best fitting RRM/CRM model for each analyst based on twenty data points. Relative to the models shown in Table 7, note the decline in adjusted R^2 s (consistent with the correlation analysis in Table 6). An additional predictor variable, competitor PE ratio, becomes significant in the model of Analyst A2. Each model is now statistically significant. The results are consistent with the point that the LR models are sensitive to sample size.

The VPA-derived LR models differ substantively from the correlation LR models developed for each subject. For example, the VPA-derived model of Ai implied the use of a P/ E ratio. The CRM model (Table 8) implies that market and industry trend predictors were used. A similar situation exists for A4. The models of A2 and A3 imply some correspondence between the VPA-derived and correlation LR model predictors. Each CRM/RRM model contains at least one predictor that is consistent with the VPA-derived model. The VPA-derived models generally contain more predictors for a given

analyst, making comparisons difficult.¹¹ Results are not consistent with Harte and Koele (1995) or EKK (1979).

The models shown in Tables 7 and 8 also suggest different data combination rules for the analysts. Taken together, these differences in the models provides support for the claim that the LR and VPA approaches are complementary. The combination of methods improves model validity.

Several points can be made. First, the structure of the LR models (Table 7) suggests that not all steps in the VPA flowcharts explain decision variance, implying error, as suggested by EKK (1979). Second, regression techniques can be used to refine a VPA model—and to assist in building an error theory of the problem solving process. For example, the LR model of (Table 7) suggests that much of the related VPA model is *statistically* irrelevant. However, EKK (1979), in their analysis and discussion, imply that error variance is fixed for a given set of models and contexts. Our analysis suggests that, since the derived models are *not* indistinguishable, the amount of error variance is not fixed. VPA-derived LR models imply substantially less error variance than do the RRM/CRM LR models in this task environment. Third, we argue that the VPA-derived models provide greater face validity. The models in Table 7 are generally comprised of more economically and statistically relevant variables (Bernard, 1994; Frankel & Lee, 1995). Further, the correlation analysis suggests that VPA model predictions tend to “track” with analysts’ predictions. Fourth, the PBGs and models are consistent with Anderson (1988) in that all subjects used a directed search procedure, constructing variations on an earnings capitalization model. Fifth, the prediction performance of the LR models (Table

¹¹ The models and data provide some support for the view that an experienced analyst may not be aware of what he or she was doing immediately after doing it (Nisbett & Wilson, 1977; Ericsson & Simon, 1980; Russo *et al*, 1989).

6) support the claim of Ford *et al.* (1989) that LR models are consistent with many different underlying processes. Even statistically insignificant LR models perform well, suggesting compatibility with virtually any process. Finally, in a setting with low “Ns,” dense data and potentially negative intercorrelations, regression models may not present a stable picture of data inter-relationships, consistent with JMG (1989).

These observations are consistent with, but offer a differing perspective than, the results of Campbell and Fiske (1959) and Payne *et al.* (1978). These authors suggest using multiple techniques in order to obtain convergent validation of results. Our results suggest that VPA and LR techniques are complements, not substitutes. Each provides enough *different* information that use of both techniques should be encouraged in modeling decision behavior in this task environment.

Our data support the contention that verbal protocols provide insights into subjects' pre-decision behavior that would be difficult to otherwise obtain about this task. The potential effect of this in the IPO context is illustrated by the following example. Downes and Heinkel (DH) (1982) provide an earnings capitalization model in their test of the signaling content of the percentage of shares retained by selling shareholders during an initial offering. The EPS number of the DH model is based on the shares outstanding prior to the offering. Our protocol models suggest that our analysts use the shares outstanding *after* the offering when they are computing an EPS number. While this apparent discrepancy is likely to be correlated with the percentage of shares retained by prior owners, it does point to the potential for modeling of this type to add to the precision of both our models and our theories. For example, the above point suggests an alteration in the computation of expected return per share. The analysts in our study did

not presume (as is implied by the DH (1982) formulation) that the proceeds of the offering would provide an additional level of return to the issuing company.

Limitations and future directions

This study examines the use of regression and verbal protocol techniques in the modeling of decision behavior in the pricing of initial public offerings. The sample size used was small, was not randomly selected, and subjects may not be representative of the population of CFAs. Further, our task is clearly an abstraction of that done by analysts “on the job.” Nonetheless, we believe that the study is an important step in understanding such tasks. The study supports the joint use of verbal protocol analysis and regression techniques in less structured task environments.

Table 1. Subjects' retrospective rank ordering of cues used during task.

A1	A2
1. Earnings potential	1. Profitability
2. Profitability	2. Market trend
3. Earnings trend	3. Earnings variability
4. Sales growth	4. Competitor stock price
5. Financial strengths	5. Price/book value ratio
6. Industry position	6. Price/sales per share
7. Product type	
A3	A4
1. Earnings growth	1. Market trend
2. Sales growth	2. Proprietary product
3. Margin trend [profit]	3. Size of market
4. Recent net income	4. Expected PE/EPS
5. Competitive strengths	5. Level of market capitalization
6. Comparative results	6. Degree of leverage
7. Industry (firm) trends	7. Maturity of firm
8. Return on investment	8. Industry trend
9. Balance sheet	9. Sales growth rate

Table 2. Operators used in search and evaluation activity.

Operator type (name)	Definition (or meaning)	Activity (type of manipulation)	Locus of standard for evaluation
Examine	Search and evaluation	Discovery and evaluation	
Tag 1	Measure of "goodness" (good, bad, indifferent)	Comparison (NQ)	Internal
Tag 2	Inputting of text	Paraphrasing of text	
Read	Inputting of text	Input, process text	
CI	Measure of size (large, small, average)	Comparison (Q)	Internal
Compute	Calculation	Calculation	—
Increase	Relative size (greater or less than)	Comparison (Q)	External
Trend	Increase over time (increase, decrease)	Comparison (Q)	External
Formulate hypothesis or relation	Because, due to	Linkage tentative conclusion	—
Formulate problem	Overall impression	Summarization	Either

Standard refers to the basis used for a given comparison—either the subject's own internal judgment or some external index could be used. Q: quantitative; NQ: nonquantitative.

Table 3. Relationship among analysts' (A_j) Predicted Price (APP_y) and actual stock offering price (IOP_t)

Panel A: Summary data on analysts' predicted prices (APP_y)				
	Analyst			
	A_1	A_2	A_3	A_4
Mean APP_j (mean $IOP=15.98$)	22.55	15.10	17.85	17.05
Correlation APP_y and IOP_t	0.7120*	0.6370*	0.5280*	0.3480
MAE_j^1	6.88	5.23	6.40	6.93
MSE_j^2	67.11	46.13	60.88	69.92
*Pearson correlation; significant beyond 0.05 $N=20$; $i=1..20$; $j=1..4$				
Panel B: Mean prediction errors by method of analysis (silent/verbal)				
	Analyst			
	A_1	A_2	A_3	A_4
MAE_j : Ten verbal cases**	6.92	5.28	7.63	6.07
MAE_j : Ten silent cases	6.82	5.18	5.18	7.77
MSE_j : Ten verbal cases**	79.11	38.93	74.93	43.78
MSE_j : Ten silent cases	55.11	53.33	46.83	96.13

Statistics (Panel A): $N = 20$

**Mann Whitney U; no differences significant at 0.05.

¹ MAE =mean absolute error. MAE_j is computed as the mean of $|APP_y - IOP_t|$.

² MSE =mean squared error. MSE_j is computed as the mean of $(APP_y - IOP_t)^2$.

Table 4. Regression fit of analyst's predicted prices (APP_{ij}) with predictors from retrospective reports (RRM)

	Analyst			
	A ₁	A ₂	A ₃	A ₄
Intercept	18.93*	8.22*	20.80*	13.86*
	[3.65]	[2.69]	[3.14]	[2.68]
Projected EPS	6.22			
	[3.17]			
Book value/share		2.19*		
		[0.65]		
Proprietary product				11.81*
				[4.90]
Stock market trend			-6.34	
			[3.75]	
Adj R^2	0.24	0.53	0.17	0.35
$p(F)$	0.09	0.01	0.13	0.04
N	10	10	10	10

Analyst's RRM using the ten verbal cases. Variables included ≤ 3 in each model.

Table 5. Regression fit of analyst’s predicted prices (APP_{ij}) with most highly correlated variables (CRM)

	Analyst			
	A ₁	A ₂	A ₃	A ₄
Intercept	16.58*	8.22*	20.85*	20.53*
	[1.45]	[2.69]	[3.28]	[2.70]
Stock market trend	3.78*			
	[1.40]			
Industry trend	7.41*			
	[1.84]			
Book value/share		2.19*		
		[0.65]		
Profit margin trend				-9.67*
				[3.23]
Competitor P/E			-0.02	
			[0.014]	
Adj R²	0.65	0.53	0.12	0.47
p(F)	0.01	0.01	0.17	0.02
N	10	10	10	10

Analyst’s CRM using the ten verbal cases. The variables are chosen based on (1) their degree of correlation with the APP_i and (2) lack of correlation with other predictor variables (if two predictor variables are highly correlated ($r > 0.70$) one of the variables is replaced).

Table 6. Comparative cross validation predictive accuracy of the RRM, CRM, and VPA models—ten silent cases Panel A

	Analyst							
	A ₁		A ₂		A ₃		A ₄	
	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE
CRM	5.85	49.47	5.19	45.55	5.90	54.82	7.13	78.44
RRM	6.25	66.36	5.19	45.55	9.88	132.84	4.63	36.67
VPA	6.64	91.63	6.60	75.70	5.21	50.77	6.36	67.33

Predictive accuracy: Analyst's predicted price (APP_{ij})—model's predicted price (MPP_{ij}). Statistical tests: Mann Whitney U (paired) and Kruskal Wallis (overall).

Panel B: Spearman Correlations of Analysts' Predicted Prices (APP_{ij}) and models predicted price (MPP_{ij})

	Analyst			
	A ₁	A ₂	A ₃	A ₄
CRM	0.177	0.74*	-0.213	0.39
RRM	0.04	0.74*	-0.34	0.65*
VPA	0.80*	0.54*	0.96*	0.76*

CRM: Correlation regression model. RRM: Retrospective regression model. VPA: Verbal protocol model. MAE: mean absolute error; $AE_{ij}=|APP_{ij}-MPP_{ij}|$. MSE: mean square error; $SE_{ij}=(APP_{ij}-MPP_{ij})^2$. *Significant beyond 0.05 (two tail). Note: Panel A significant difference only for Analyst A₃, for VPA vs RRM models.

Table 7. Regression of analysts' protocol variables with analysts' predicted prices, APP_{ij} (standard errors in brackets) $\ln(APP_{ij}) = \beta_{0j} + \beta_{1j} \ln X_{1ij} + \beta_{2ij} \text{fty} \ln X_{2ij} \dots + \beta_{nj} \ln X_{nij} + e_{ij}$

	Analyst			
	Λ_1	Λ_2	Λ_3	Λ_4
Constant	-1.69 [0.01]	1.11* [0.11]	-0.49 [0.91]	0.14 [0.53]
$\ln(\text{MARGIN}_{jt}) \cdot D1_{jt}$	-0.05 [0.22]	0.47* [0.22]	0.28* [0.10]	
$\ln(\text{SALES}_{jt}) \cdot D1_{jt}$	0.07 [0.12]	0.76* [0.16]	0.79* [0.09]	
$\ln(\text{GROWTH}_{jt}) \cdot D1_{jt}$	-0.54 [0.49]		1.45* [0.41]	1.04* [0.28]
$\ln(\text{SHARES}_{jt}) \cdot D1_{jt}$	0.10 [0.17]	-0.80* [0.19]	-0.75* [0.11]	
$\ln(\text{EPS}_{t4}) \cdot D1_{t4}$				0.04 [0.24]
$\ln(\text{PE}_{jt}) \cdot D1_{jt}$	1.27* [0.21]	0.54* [0.15]	0.56* [0.140]	0.86* [0.19]
$\ln(\text{BOOK PER SH}_{t4}) \cdot D2_{t4}$				0.70* [0.29]
$\ln(\text{SALES PER SH}_{t2}) \cdot D2_{t2}$		0.94* [0.13]		
$\ln(20) \cdot D2_{t1}$	1.55* [0.34]			
Adj. R^2	0.63	0.92	0.79	0.62
$p(F)$	0.00	0.00	0.00	0.00
n	20	20	20	20

*Coefficient significant beyond 0.05. The variables defined below are computed using the procedures suggested by the verbal protocols of the analyst. The measurement of a variable is likely to vary across analysts. For instance, analysts used different earnings numbers (before (before extraordinary, before tax and extraordinary) and different number of time periods when estimating a firm's profit margin.

Margin_{jt} =the margin (net income/sales) computed for firm t by analyst j using j 's computational procedures.

Sales_{jt} =the annualized sales as estimated by analyst j for firm t .

Growth_{jt} =1 plus the projected growth in sales for firm t as estimated by analyst j .

Shares_{jt} =the number of shares of firm t projected to be outstanding after the offering.

EPS_{jt} =the reported (or annualized) earnings per share of firm t .

PE_{jt} =the price earnings ratio for firm t as projected by analyst j .

Book per sh =(stockholders equity _{t} /shares outstanding before offering _{t}).

Sales per sh =(sales _{t} /shares outstanding before offering _{t}).

$D1_{t1}$ =1 and $D2_{t1}$ =0 if analyst 1 uses an earnings model. $D1_{t1}$ =0 and $D2_{t1}$ =1 if analyst 1 does not use an earnings model.

Analyst 1 uses an earning model if projected $\text{EPS} > \$0.20$.

$D1_{t2}$ =1 and $D2_{t2}$ =0 if analyst 2 uses an earnings model. $D1_{t2}$ =0 and $D2_{t2}$ =1 if analyst 2 does not use an earnings model.

Analyst 2 uses an earnings model if projected $\text{EPS} > \$0.25$.

$D1_{t3}$ =1. Analyst 3 always uses an earnings model.

$D1_{t4}$ =1 and $D2_{t4}$ =0 if the most current twelve months earnings per share for firm t is positive, otherwise $D1_{t4}$ =0 and $D2_{t4}$ =1.

Table 8. Regression fit of analyst's predicted prices (APP_{ij}) with subject's best-fitting RRM or CRM models [standard errors in brackets]

	Analyst			
	A ₁	A ₂	A ₃	A ₄
Intercept	18.70* [1.66]	11.09* [2.09]	19.53* [2.01]	12.27* [2.31]
Competitor PE		-0.014* [0.006]	-0.017* [0.008]	
Proprietary product				-9.50* [3.45]
Book value		1.91* [0.49]		
Stock market trend	3.34* [1.51]			
Industry trend	5.49* [1.88]			
Adj R^2	0.45	0.56	0.17	00.26
$p(F)$	0.01	0.00	0.04	0.01
n	20	20	20	20

* Coefficient significant beyond 0.05.

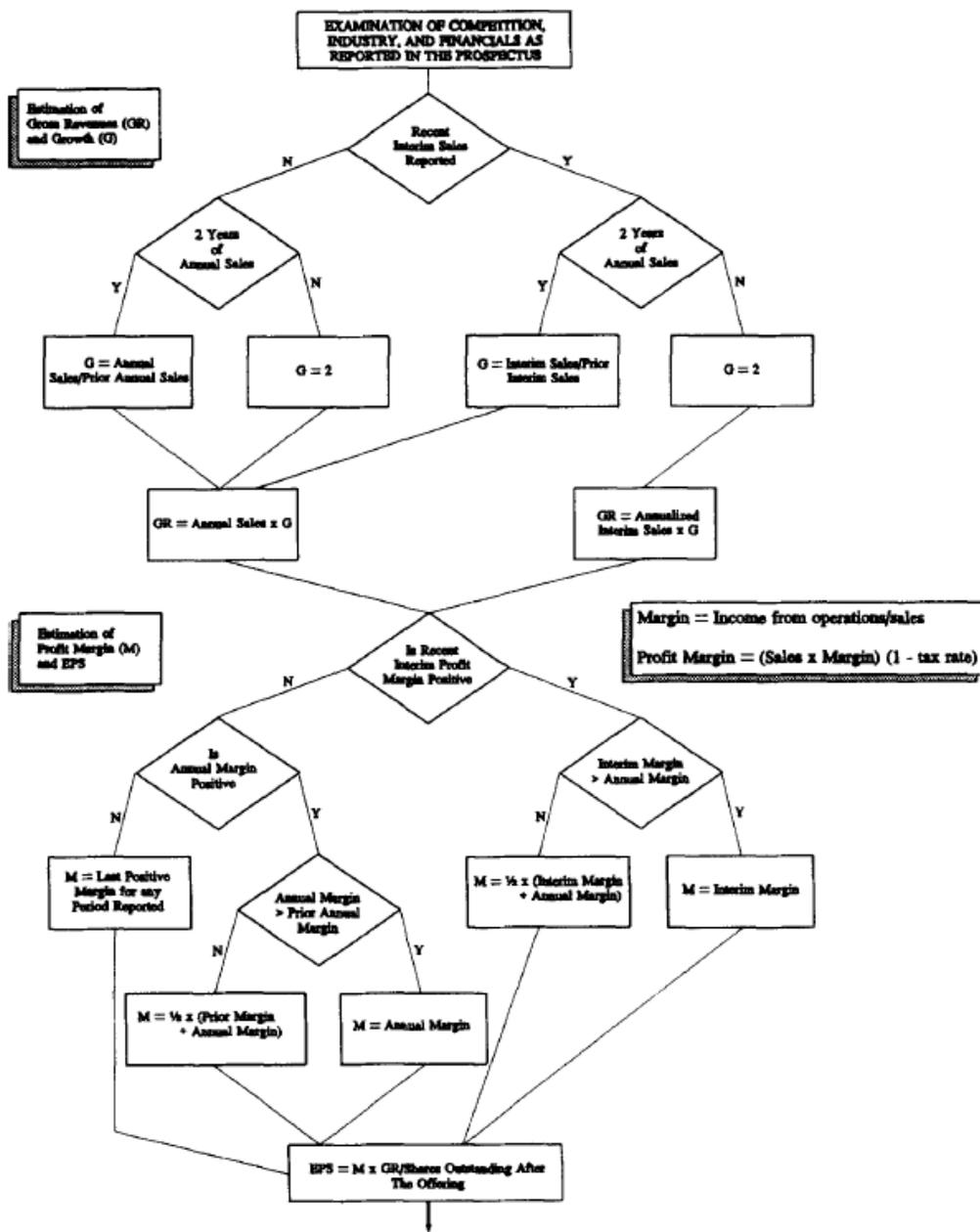
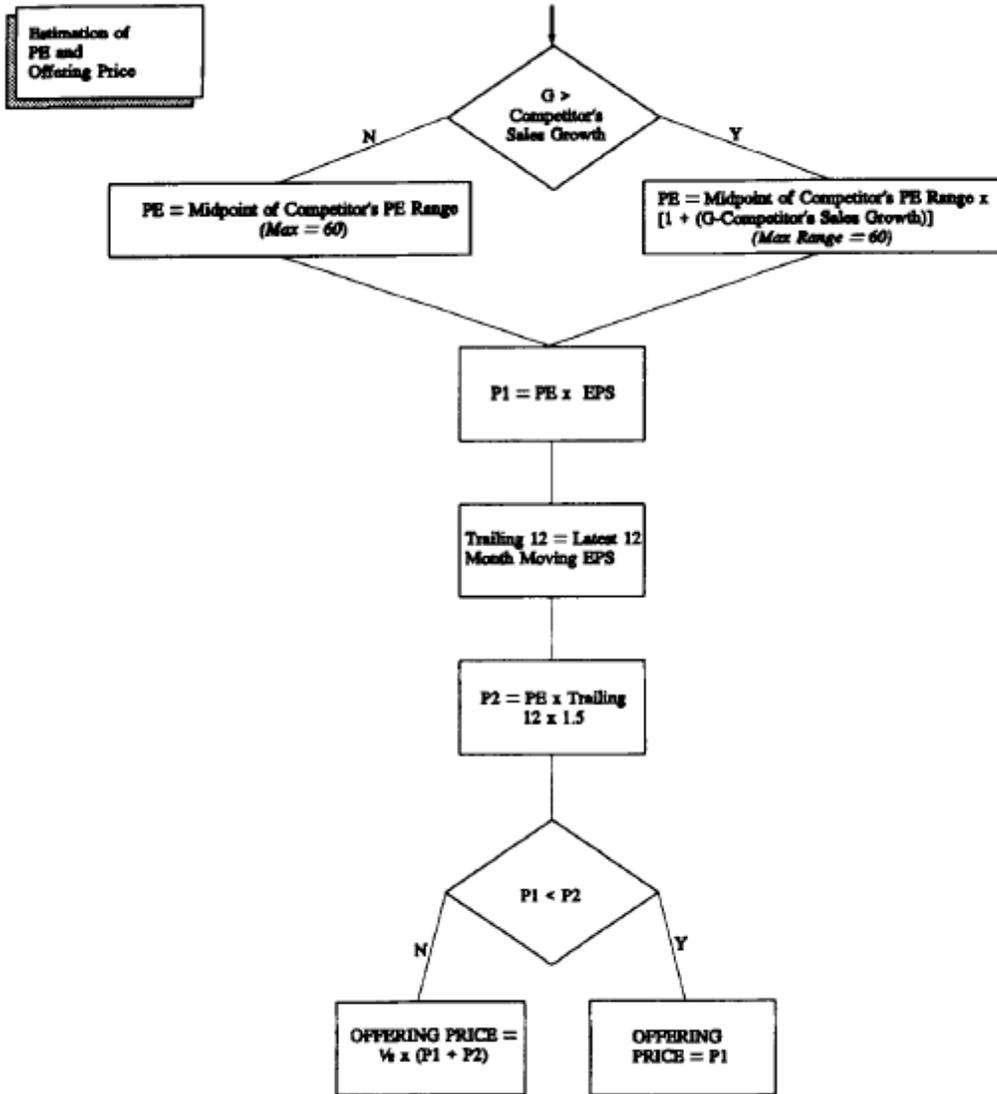


Figure 1. The flowchart model developed from the verbal protocols of Analyst A, is depicted above. As can be seen, the model is quite detailed, with numerous feedback loops and contingent decision points.

(Fig. 1 con't)



REFERENCES

- Abdolmohammadi, M., & Wright, A. (1987). An examination of the effects of experience and task complexity on audit judgments. *The Accounting Review*, 62, 1-13.
- Anderson, M. J. (1984). The process tracing paradigm: a general view. In S. Moriarity & E. Joyce (Eds.), *Decision making and accounting: current research* (pp. 158-172). Norman, OK: University of Oklahoma Printing Service.
- Anderson, M. J. (1985). Some evidence on the effect of verbalization on process: a methodological note. *Journal of Accounting Research*, 23, 843-853.
- Anderson, M. J. (1988). A comparative analysis of information search and evaluation behavior of professional and non-professional financial analysts. *Accounting, Organizations and Society*, 13, 431-446.
- Bernard, V. L. (1994). Accounting-based valuation methods, determinants of market-to-book ratios, and implications for financial statement analysis. Working paper, University of Michigan.
- Biggs, S. F. (1984). Financial analysts' information search in the assessment of corporate earning power. *Accounting, Organizations and Society*, 9, 313-334.
- Biggs, S. F., Rosman, A. J., & Sergenian, G. K. (1993). Methodological issues in judgment and decision-making research: concurrent verbal protocol validity and simultaneous traces of process data. *The Journal of Behavioral Decision Making*, 6, 187-206.
- Brunswik, E. (1952) *Conceptual framework of psychology*. Chicago: University of Chicago Press.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement*, 20, 37-46.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-106.
- Dawes, R. M. (1988). *Rational choice in an uncertain world*. Orlando, FL: Harcourt, Brace, Jovanovich.
- Downes, D. H., & Heinkel, R. (1982). Signaling and the valuation of new issues. *Journal of Finance*, 37, 1-10.
- Einhorn, H. J., & Hogarth, R. M. (1980). Behavioral decision theory: processes of judgment and choice. In M. Rosenweig & L. Porter (Eds.), *Annual review of psychology* (pp. 53-88). Palo Alto: Annual Review, 1981.
- Einhorn, H. J., Kleinmutz, D., & Kleinmutz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 86, 465-485.
- Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological review*, 87, 215-251.
- Ericsson, K., & Simon, H. (1984). *Protocol Analysis*. Cambridge, MA: The MIT Press.
- Ford, J. K., Schmitt, N., Schechtman, S. L, Hults, B. M., & Doherty, M. L. (1989).
Process tracing methods: contributions, problems, and neglected research

- questions. *Organizational Behavior and Human Decision Processes*, 1989, 43, 75-117.
- Frankel, R., & Lee, C. M. C. (1995). Accounting valuation, market expectation, and the book-to-market effect. Working paper, University of Michigan.
- Halloran, M. J. (1979). *Going public*. New York: Dain Bosworth.
- Harte, J. M., & Koele, P. (1995). A comparison of different methods for the elicitation of attribute weights: structural modeling, process tracing and self reports. *Organizational Behavior and Human Decision Processes*, 1995, 64, 49-64.
- Hawkins, D. F., & Campbell, W. J. (1978). *Equity valuation: models, analysis and implications*. New York: Financial Executives Research Foundation.
- Hayes, J. R., & Simon, H. (1976). Understanding complex task instructions. In D. Klahr (Ed.), *Cognition and instruction* (pp. 271-295). Hillsdale: Lawrence Erlbaum.
- Hogarth, R. M. (1991). A perspective on cognitive research in auditing. *The Accounting Review*, 66, 277-290.
- Jain, B. A. (1996). Is underpricing a signal of firm quality?. *American Business Review*, 14, 38-45.
- Johnson, P. E. (1983). What kind of expert should a system be?. *The Journal of Medicine and Philosophy*, 8, 77-97.
- Larcker, D. F., & Lessig, V. P. (1983). An examination of the linear and retrospective process tracing approaches to judgment modeling. *The Accounting Review*, 58, 58-77.
- Leland, H. E., & Pyle, D. H. (1977). Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance*, 32, 371-387.

- Lev, B., & Thiagarajan, S. (1990). Financial information analysis. Working paper, University of California, Berkeley.
- Luger, G. F., & Stubblefield, W. A. (1989). *Artificial Intelligence and the Design of Expert Systems*. Redwood City, CA: Benjamin/Cummings.
- Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: verbal reports on mental processes. *Psychological Review*, 84, 231-277.
- Payne, J. W., Braunstein, M. L., & Carroll, J. S. (1978). Exploring predecisional behavior: an alternative approach to decision research. *Organizational Behavior and Human Performance*, 22, 17-44.
- Porcano, T. M. (1981). A comparison of information needs and sources of the investment community. *Akron Business and Economic Review*, 12, 43-52.
- Ritter, J. R. (1984). Signaling and the valuation of unseasoned new issues: a comment. *Journal of Finance*, 39, 1231-1237.
- Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory and cognition*, 17, 759-769.
- Selling, T., & Shank, J. (1989). Linear versus process tracing approaches to judgment modelling: a new perspective on cue importance. *Accounting, Organizations and Society*, 14, 65-77.
- Slovic, P. (1969). Analyzing the expert judge: a descriptive study of a stockbroker's decision process. *Journal of Applied Psychology*, 255-263.

Smith, G. F. (1988). Towards a heuristic theory of problem structuring. *Management Science*, 34, 1489-1506.