



IIMK/WPS/109/ MKTG /2012/12

**Micro Modeling of Individual Level Information Seeking
Behavior: A Heterogeneity Specific Study**

*Mahuya Adhikary
Atanu Adhikari¹
Chaitali Dutta*

¹ (Corresponding Author), Visiting Assistant Professor at Indian Institute of Management Kozhikode,
Phone: 91-495- 2809241, email: atanu.adhikari@iimk.ac.in

Micro Modeling of Individual Level Information Seeking Behavior: A Heterogeneity Specific Study

The responses of individual information seekers are best studied in isolation. In the past researchers did not consider information users as individuals; viewing them instead in the context of their cultural system as a whole. This study emphasizes the importance of heterogeneity in information seekers, arguing that the behavior and responses of different seekers would vary based on their demographic and socio-graphic profile, the extent of their background knowledge, and the variations in their risk adversity and cost consciousness. In order to understand this aspect of the information seeker's behavior, information seeking behavior dimension was derived. The effect of this information seeking behavior dimension on the users was investigated. An in-depth questionnaire was administered to 307 respondents. Factor Analysis, Cluster Analysis, Multiple regression and Bayesian Regression were used on the responses in order to identify (i) information seeking behavior dimensions, (ii) different segments of information seekers, (iii) the effect of behavioral dimensions on the behavior of information seekers and (iv) the differences between the effect of these dimensions on users, both at an individual level and at the aggregate level. It was revealed that there is a significant difference in the behavior when the aggregate level estimates are considered and when individual level Bayesian estimates are made. These findings were discussed to understand the implications and to provide avenues for future research.

Introduction

The first phase of research in the area of information seeking behavior was mostly concerned with information centers and the information seeking process (Taylor, 1968; Robertson, 1977; Bettman, 1978; Belkin, 1980). It was in the 1980s that there was a change in the approach to information needs and also the study of users. Instead of considering users in the context of a system, Dervin and Nilan's (1968) alternative paradigm considered the individuality of the users along with their information needs, specifically in terms of their internal, psychological and cognitive states. The problems faced by the users were the primary concern in the second phase of research carried out in the area of information seeking (Taylor, 1985). The models used by Kulthau (1991), Krikelas (1983), Eisenberg and Berkowitz (1992) and Wilson (1981) considered the cognitive behavior and feelings of the information seeker. These models also took into consideration the different contexts and the environmental factors that influenced the information seeker's behavior. Johnson and Meischke (1993) first observed the critical behavioral factors of individual information seekers in health related information. They noted that the individual

information seeking became a critical element in determining health behavior. However, the actual needs of the user, behavioral details and the interaction of individuals with the information were considered in a very limited manner.

The responses of individual information seekers when looking for a particular piece of information in different contexts were best studied in isolation. Wilson's (1981) model elaborates the users' cognitive space, social/organizational environment, and the different contexts in which the need for information arises, in addition to the environmental factors. Wilson's revised model (1994) specified the barriers of information seeking and emphasized their role in information seeking. These barriers were explicitly specified by variables such as psychological, demographic, environmental and source characteristics. Two dynamic components were introduced to complete the model. These were stress/coping theory, social learning theory and the processing of information and its uses. Wilson (1999) described this model as a macro-behavior model; one that provided a rich source of hypotheses and also the potential for setting a research agenda that differed from the previous one. Ingwersen's (1996) model integrated ideas relating to information behavior and information needs with respect to IR system design. Hjørland (1997) believed in the "activity-theoretic" approach, which is founded on the ideas of John Dewey. Hjørland's view of subject retrieval is strongly user oriented. However, according to these researchers the user was not considered as an individual; rather he was deciphered in the context of his cultural system. Information needs were seen as secondary needs, and were believed to have derived from basic needs. There was absolutely no scope to consider the variations in an individual's information searching behavior.

This study emphasizes the importance of information seeking behavior in the context of personal behavior. Heterogeneity among the information seekers suggests that different seekers would behave differently when seeking information. We argue that previous studies considering prespecified demographic and psychographic variables can capture Intel individual difference in very limited way. It is because information users with same demographic profile and psychographic profile may have different level of understanding and varied a priory knowledge on the information they search for. Despite belonging to similar kind of group in terms of demographic and psychographic variables, they may have different level of resource constraints, personality traits etc that would influence their information seeking behaviour. The

objective of this study is to explore if there exists considerable difference of the effect of information seeking behaviour dimensions on information seekers action if inter-individual heterogeneity is considered. To do that, we first find out information seeking behavioral dimensions of the people who use information. We then, group the users according to these behavioral dimensions into segments, and then investigate impact of these behavioral dimensions on these segments considering individual level heterogeneity through hierarchical Bayes. We then compare these estimates with aggregate level estimates to find out the improvement of prediction if Intel individual heterogeneity is considered. This paper addresses research questions namely (i) what are the underlying dimensions of information seeking behavior, (ii) how the users are segmented according to these dimensions and how members of several groups differ on these dimensions as well as on their demographic profile, (iii) what is the effect of these information seeking behavior dimensions on actual information seeking activities, and (iv) a comparison of the effects of individual level analysis and an aggregate level analysis. The purpose of this comparison is to show that in most instances individual level analysis provides better results than aggregate level analysis. To our knowledge, there has been no prior study in the field of information science literature on this issue. There does exist a school of research literature which takes into consideration aggregate level analysis. The researchers who subscribe to this research model believe that information users from a particular population are influenced and motivated in a similar manner (Brown 2007, Buente 2008). In contrast to this, our research focuses on the individual level behavior. Previous researchers were more interested in aggregate effects and paid no attention to individual differences. They considered heterogeneity an unimportant statistical parametric problem; but in the application context, the individual differences are considered significant as there are differences in individual's outcome behavior. Information management policies often attempt to respond to individual differences. Information practices require flexible models of heterogeneity in order to exploit users' differences in individual level estimates. The Bayesian approach of following a continuous model of heterogeneity is used in this research to examine individual differences in the information seekers. The reasoning behind our project is the belief that every individual is unique. The differences in the demographic and socio-graphic background will have an impact on several information seeking dimensions, causing varied effect on the behavior of the individual. In other words, we would like to show

how different customers respond differently to the same information seeking context. As part of this research project we studied the information seeking behavior of management students from various management schools, each belonging to different levels. We used Factor analysis and cluster analysis to find the behavioral dimensions, and then segmented the respondents on certain behavioral variables. We then applied Individual Level Bayesian (IBE) estimate in estimating segment level coefficients and then compared them with the aggregate level coefficients.

The organization of the paper is as follows. In next section we provide a review of the already existing literature on information seeking behavior and individual level heterogeneity, followed by the description of the methodology used in finding behavioral dimensions and segmenting information users. In next section we provide a description of hierarchical Bayes methodology in estimating parameters considering individual level information, followed by a discussion of the data and presentation of the findings. The conclusion presents a summary of the findings and discusses their implications. The paper ends with a list of future research topics in this area.

Review of the literature

It is traditionally believed that the main characteristics of information seeking are concurrent, continuous and cumulative, and that these occur in various cyclic loops. Traditional paradigms of information retrieval tend to simplify the information seeking process by presenting different models. Robertson (Robertson, 1977) presented models in which the information seeker entered a query and answers that matched it were given. This model described the retrieval process, but failed to take into consideration the actual user needs, behaviors, situations and gaps in knowledge. Wilson (1981) proposed a model which elaborated the concepts of information need, information seeking, information exchange and information use. The model also explained the barriers that prevent the completion of a search process. This model was revised in 1994 to incorporate Elis's (1989) specific empirical model on information seeking. The revised model explicitly specified the barriers and emphasized their roles in information seeking. The barriers were identified by the variables, and divided into the psychological, demographic, environmental and source characteristics. Solomon (1997) revealed that information seeking behavior rarely followed the ideal order of selecting, gathering, processing of information. Information seeking behavior was better characterized by expressions such as non-linear, cycling, recycling, and false

starts. Several developments of studies on human information behavior during the post-war period led to a shift of paradigm in information science research (Dervin & Nilan, 1986). The new perspectives in research reversed traditional models of system design, placing the user, rather than the system, at the center of attention. Sonnenwald (1999) perceived information seeking as an activity that occurred in specific contexts and situations. He proposed an 'information horizon' in specific contexts and situations. Foster (2004) offered a non-linear model of information seeking behavior which contrasts with the earlier linear behavioral model of information searching strategies or processes. This model was based on the findings of a study of interdisciplinary information seeking behavior and revealed information seeking as possessing nonlinear, dynamic, holistic and flowing characteristics. All these research projects, however, were more concerned with the information seeking processes and systems.

In behavioral sciences various researches was carried out in order to identify the information seeking behavior of different social groups of people. According to Maslow (1943) the hierarchy of human need varied with every individual, since the satisfaction of each basic need is never the same from person to person. McKenzie and Carey (2000) introduced Davies and Harre's social positioning theory as a framework to identify the link between an individual's health status and social factors which affected his information seeking behavior. The Positioning theory was useful for studying the ways in which individuals or a group of people seek information in different situations. McKenzie's research was based on information use and search behavior of an individual information user. Howard (1993) in his 'multiple intelligences' theory stated that an individual manifests varying levels of different intelligences. Research was carried out to understand the literature searching experiences and skills of clinical information users in different medical centers (Wessel, Tannery & Epstein 2006, Johnson and Meischke 1993, Ippolite and Mathios 1990). In another study (Nicholas et al. 2006) users of digital journal library were analyzed with a focus on their different behavioral characteristics. In this study, users were grouped based on some predetermined characteristics namely their occupation, place of work, geographical location etc. In the study of information science behavior, research (Haines et al. 2010) revealed that information centers were not the only source of information for the users. The researchers concluded that information specialists had to be accustomed to the demands of the users in order to make their

resources and services valuable to them. Hence, individual specifications and variations in services based on personal differences of the users play a significant role.

The difficulty in considering individual level heterogeneity in information science research lies in estimating parameters keeping in mind individual level information. Hierarchical Bayes estimation procedure solves this problem to a great extent. Several hierarchical Bayes literature on random effect model (Rossi, Allenby and McCulloch, 2005; Allenby and Ginter, 1995; Lenk et al, 1996) suggest a method that helps in estimating individual specific parameters using aggregate level information under limited data. In a model of this kind individuals are considered as independent conditional on aggregate level parameters. This means the prior induced in individual level are not independent prior. Individual parameters are considered as drawn from the whole population. Ainslie and Rossi (1998) attempted to incorporate demographic information in their research in order to study the differences in individual specific parameters.

This research analyses information seeking behavior considering individual level heterogeneity. This study proposes that users are distinctive about their information seeking behavior during the process of seeking information.

Methodology

The present study examined four aspects of information seeking behavior as mentioned in the introductory section. To address the first question, we used Principal Component Factor Analysis to derive underlying behavioral dimensions. Factor analysis is used in behavioral research because of its ability to group measured variables based on their internal correlation. If a group of measured variables share a strong correlation with each other, they are the indicators of an underlying dimension. SPSS software was used to conduct Factor analysis.

The second point addressed in the study is the segmentation of respondents using Cluster analysis. Cluster analysis divides respondents into a number of homogeneous groups. The variability among the members of a group is minimum while that of members between the groups is maximum. This technique can be used to determine the number of homogeneous groups, the size of these groups and their demographic profiles.

The third question has been addressed using multiple regression analysis. Several dependent variables that depict the effect of the users' action on behavioral dimensions based on factor analysis have been regressed. The estimated regression

coefficient shows the effect of the behavioral dimensions on the perceived actions of users.

As regards the fourth question we argue that heterogeneity in preferences results in providing differentiated information to different individuals and to different segments of information seekers. Differences in sensitivities are the basis for targeting information seekers. As the preferences of information users' regarding the sources of information and the methods used for the search becomes diverse, it becomes less efficient to consider the information at the aggregate level. We used hierarchical Bayes for this purpose. The advantage of Bayesian methodology is that it produces estimates of individual units across all parameters. These estimates are of particular interest for the information service providers who are interested in pursuing service differentiation strategies. These strategies include designing services which are offered to specific groups of users with specific needs. The classic method of capturing heterogeneity cannot provide information since it gives aggregate level estimates. The disaggregate nature of many service decisions is based on the assumption that there are individual differences and on the availability of limited number of data points per user. This study successfully uses data crossways for an individual user while estimating the individual's posterior parameter vector.

Natter and Feurstein (2002) pointed out that only demographic variables may not be adequate to capture individual level heterogeneity. They believed that one required additional information about the consumer and needed to take into consideration expectation values with reference to the distribution of heterogeneity. Thus it is imperative that the methodology of estimating users' information seeking behavior consider individual level heterogeneity through hierarchical Bayes. Bayesian methodology is proposed in this research to estimate individual parameters with a particular distribution as opposed to the classical method of parameter estimation where population parameters are constant.

Data analysis and Findings

Peer evaluation of the instrument was carried out (Devellis 2003) before the pilot testing of the instrument. Twelve items were marked as ambiguous due to lack of clarity, proper design of instrument, ease of reading and writing, and layout of the page. After incorporating the comments of the peers, the pilot test was carried out (Devellis 2003). The first pilot test was carried out with 30 respondents from the same population, and the measure of internal consistency, i.e. Cronbach alpha value found

to be 0.64. Eight more items were included, three excluded, and a second pilot test was carried out, which gave an alpha value near 0.78.

The final questionnaire was sent to a total of 563 management and doctoral students in six major cities in India. 366 responses were received, out of which the data of 307 respondents was suitable for analysis. These respondents shared their experiences on the different behavioral characteristics of the MBA students at the time of seeking information for various purposes. They also agreed that information searching behavior varies depending on individual variation. Every student has some specialty that is unique to him or her. Thus the information seeking behavior also varies from one individual to another.

Identifying the information seeking behavioral dimensions from the behavioral variables:

Information seeking behavior dimensions are behavioral traits that are inside the mind of the users. . We collected data of indicator variables on the seven point Likert scales, which is a continuous scale. Then we checked the correlation between each of the measured variables. Sampling adequacy was measured by the Kaiser-Meyer-Olkin (KMO) statistics. This method reveals, based on the inter-item correlation of the measured variables, if they are likely to have underlying dimensions and can form factor or not. According to researchers (Kaiser & Cerny, 1979), only when the KMO value is more than 0.5 can we proceed with factor analysis. In this research the KMO value was found to be 0.607 compliant with the required value. We also carried out Bartlett's test of sphericity, which indicates whether the correlation matrix is an identity matrix, that is, if the variables are unrelated. The value of the test variables was less than 1% (i.e. 0.01), indicating that the correlation matrix is not an identity matrix and that factor analysis can be executed.

The Measure of Sampling Adequacy (MSA) for the respective item was checked. MSA values that exceed 0.5 indicate a good fit with the structure of the other variables. In this research, all but one variable had a value above 0.5. The one variable had a value 0.487. This value was considered as being not significantly less than 0.5 and therefore retained in the analysis. . Communality value of each variable was checked and it was ensured that all variables had a communality value that exceeded the specified value of 0.5. Factors having Eigen values greater than 1.0 were the only ones extracted and rotated to get a final solution. Four factors were extracted from the variables. Cumulative variance for the set of variables is 72.32%.

Table 1: Underlying dimensions and their corresponding variables

Identification of Factors (Underlying Dimension)
Specific document search behavior
Mode of search behavior
Related document search behavior
New knowledge seeking behavior

Table 1 shows four underlying dimensions of the information seeking behavior of individuals. These dimensions reveal how the users behave while they search for information. The dimensions also reveal information about their mode of search and their new knowledge seeking behavior. These four dimensions also reveal the different types of information seeking behavior of the individuals.

Cronbach alpha (Cronbach 1951) measure is used to measure the reliability of the factors. Cronbach's alpha generally increases as the inter-correlations among test items increase, thus earning it the reputation of providing an internal consistency estimate of reliability of the factors. Cronbach (1951) opined that an Alpha value greater than 0.6 indicates that the factors are internally consistent and that they show high reliability of the variables under consideration. In this research the alpha values of the above four dimensions are 0.702, 0.810, 0.707 and 0.584.

Estimating effects of information seeking behavior dimensions on several dependent variables by multiple regression analysis:

Next, the effects of underlying dimensions of information seeking behavior on usage of web based sources is estimated using linear multiple regressions. The objective of multiple regressions is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. The effect of independent (explanatory) variables on a dependent variable is then estimated. It is a method to learn how the average value of dependent variable changes for unit change of one independent variable given that other independent variables remains constant. For example, information users will have different information seeking behavior as defined in the four factors above. Different users will have different values of the four factors which would have an impact on their

usage pattern of web based sources. Once this information of various users has been compiled it would be interesting to see whether and how these four dimensions are related to the usage pattern of information seekers. For example, it may be revealed that new knowledge seeking behavior has higher impact on one's web based usage pattern or that a conventional mode of information search might have less impact. These types of findings can be opted from multiple regressions.

The regression coefficient R is observed in the analysis. In multiple regressions R is the measure of association between the observed value and the predicted value of the dependent variable. R Square or R^2 in multiple regressions indicates the variability in the dependent variable. Adjusted R^2 in multiple regressions takes the number of variables in the model and the number of data points in the model into consideration and examines how well the data fits the model. In other words, it says what percentage of information in the dependent variable is explained by several behavioral dimensions.

Effect of information seeking behavior dimensions on one's inclination to use web /e-mail as a source of information:

In order to assess the effect of four information seeking behavior dimensions i.e. specific document search behavior, mode of search behavior, related document search behavior and new knowledge seeking behavior on a dependent variable i.e. one's inclination to use web /e-mail as source of information , multiple regressions is carried out. The value of adjusted R^2 is 68.8%, i.e. 68.8% of the variation in dependent variable 'one's inclination to use web /e-mail as source of information' can be explained by the four information seeking dimensions. The ANOVA table shows that the regression model is significant in <1% level of significance. Higher value of F statistics (F = 169.368) and high significance level (<1%) suggests that there exists a linear relationship between the dependent variable (one's inclination to use web /e-mail as source of information) and independent variables (four information seeking dimensions). The impact of information seeking behavior dimensions on the dependent variable is given in table2:

Table 2: Coefficients of each information seeking behavior dimensions

Information seeking Behavior Dimensions	Coefficient value	Level of Significance
Specific document search behavior	.220	.000
Mode of search behavior	.449	.000
Related document search behavior	.062	.196
New knowledge seeking behavior	1.147	.000

It can be understood from the table of individual coefficients that new knowledge seeking behavior has maximum effect on one's inclination to use web /e-mail as a source of information. It implies that as a new knowledge seeker tries to gather new knowledge on a particular topic, they are more prone to use web based sources and e-mail interactions. The other two information seeking behavior dimension i.e. specific document search behavior and mode of search behavior also have significant positive impact on one's inclination to use web /e-mail as a source of information. But the related document seeking behavior dimension does not have a significant effect on one's inclination to use web /e-mail as source of information. In other words, it implies that information seekers use web/email based source when they look for some specific type of information; however they do not prefer to try web based source when they look for some related information. Instead they depend on other possible sources to look for related information.

Effect of information seeking behavior dimensions on usage of online database:

To see the effect of four information seeking behavior dimensions on dependent variable i.e. on how long one uses online database, multiple regressions are carried out. The value of adjusted R^2 is .578, i.e. 57.8% of the variation in dependent variable 'how long one uses online database' can be explained by the four information seeking dimensions. It can be gathered from the ANOVA table that the regression model is significant at <1% level of significance. A high value of F statistics ($F = 105.680$) and high significance level (<1%) suggests that there exists a linear relationship between the dependent variable and independent variables (four information seeking dimensions) and that these four dimensions have, to a large extent, explained the

change in how long one uses database. . The impact of each information seeking behavior dimension on the dependent variable is given in the following table 3.

Table 3: Coefficients of each information seeking behavior dimensions

Information seeking Behavior Dimensions	Coefficient value	Level of Significance
Specific document search behavior	-.081	0.139
Mode of search behavior	-.019	0.733
Related document search behavior	-.008	0.878
New knowledge seeking behavior	1.116	0.000

A study of the table of individual coefficients reveals that new knowledge seeking behavior has maximum effect on the time spent on online database. It suggests that when new knowledge seekers gather information about a particular topic, they are more prone to use online database. The remaining three information seeking behavior dimensions i.e. specific document search behavior, mode of search behavior and related document seeking behavior dimensions do not have a significant effect on how long one uses online databases. This proves that information seekers use online source when they look for a new type of information. In situations where they are aware about the source, they may seek an existing source. It is apparent that an information seeker would spend a significant time looking for new information in online data bases. However, the time he would spend on online data base would be insignificant when he looks for specific information or a particular mode since he would be familiar with the source and be able to retrieve the information quickly. Since related information seeking behavior does not have a significant effect on one's search for web based source, it also do not have a significant effect on the time spent using online database. Thus these findings reveal the specific information seeking characteristics of the information users.

Effect of information seeking behavior dimensions on frequency of usage of online database:

The effect of four information seeking behavior dimensions on the dependent variable 'frequency of usage of online database' is estimated by multiple regressions. It is ascertained that the value of adjusted R^2 is .670, i.e. 67% of the variation in

dependent variable ‘frequency of usage of online database’ can be explained by the four information seeking dimensions. It is determined from the ANOVA table that the regression model is significant in <1% level of significance. A high r value of F statistics ($F = 156.377$) and high significance level (<1%) suggests that there exists a linear relationship among the dependent variables and independent variables (four information seeking dimensions).The impact of each information seeking behavior dimension on the dependent variable is given in Table 4.

Table 4: Coefficients of each information seeking behavior dimensions

Information seeking Behavior Dimensions	Coefficient value	Level of Significance
Specific document search behavior	-.209	0.000
Mode of search behavior	-.351	0.000
Related document search behavior	.099	0.046
New knowledge seeking behavior	1.154	0.000

It is established from the table of individual coefficients that new knowledge seeking behavior has the maximum positive effect on the frequency of online database usage. It implies that as new knowledge seekers try to gather new information on a particular topic, they use online database more frequently. Two information seeking behavior dimensions i.e. specific document search behavior and mode of search behavior dimension have a significant negative effect on the frequency of usage of online database. It means that the frequency of search decreases when the document is more specific or mode of search behavior improves (this may be a hard copy document as compared to telephonic search or an electronic search.). This result depicts an actual pattern of the effect of these two dimensions on usage frequency of online database. It implies that an information seeker increases usage frequency of online source when he or she looks for new type of information. However, for other type of information, which are more specific in nature, the usage frequency decreases.

Segment of people based on same information seeking behavior variables:

After identifying information seeking behavior, cluster analysis is carried out to classify the users into relatively homogeneous groups called clusters. The

respondents in each cluster tend to be similar to each other and differ from individuals in other groups. Concerned heterogeneity among the information seekers would come out in this analysis as it is based on the algebraic distances between the coordinates of the respondents.

Clustering based on information seeking behavior variables:

We selected three clustering variables based on the characteristics of users' information seeking behavior. These clustering variables are:

1. Frequency of usage of web based / e-mail sources for acquiring information.
2. Length of time spent on online database by an individual user. .
3. Length of time individual user continues searching information when she does not get the required information.

Then cluster analysis technique is carried out in two steps using SPSS software package. In the first step we used hierarchical clustering procedure to identify how many homogeneous groups are present in the data.

In an attempt to generate clusters with minimum variance, we selected Wards method. K-mean clustering technique was used in the second step to identify the location of the segments on clustering variables. K-mean clustering also helped us estimate the approximate size of user segments in the population. We identified three distinct clusters through k-mean procedure and their behavioral criteria. We also calculated the size of each segment and their characteristics using the k-mean procedure. . Following are the profiles of these clusters along with their size and characteristics.

Profile of Respondent Segments

Segment 1: This segment consists of 12.3% of the information users. It is ascertained from the final cluster centers of cluster 1 that the respondents in this segment use website / e-mail as a source of information very frequently. The average usage is approximately 2.66 years, which suggest that users in this segment use online database for quite some time. Also, that they are very patient while searching some information in a printed document.

Segment 2: This segment consists of 53% of the information users. It is observed from the final cluster centers of cluster 2 that respondents in this segment do not use website / e-mail as a source of information very frequently. They rank it quite low

(close to 6th in their ranking) in their list of sources of information. However, the average usage of online databases, which is 6.23 years, proves that they use online database for a long time. The finding also suggests that the users in this segment are not very patient while seeking some information in a printed document.

Segment 3: This segment is the largest and consists of 34.7% of information users. It is discovered from final cluster centers of cluster 3 that respondents in this segment also do not frequently use website / e-mail as a source of information. They also rank it quite low (above 6th in their ranking) in their list of sources of information. However, the average usage of online databases, which is approximately 5 years, proves that they use online databases for a significant amount of time. . However, they display a degree of patience while searching for information in a printed document.

Individual level Bayesian analysis and comparison with aggregate analysis:

In this research, we estimated separate regression estimates for each individual respondent. In the past, researchers have tried to handle this problem with an adjustment by way of ignoring heterogeneity among individuals. They pool all the data together, and estimate a single set of regression coefficients that describes the “average” of all individuals. However, an alternative solution has recently become available with the introduction of “hierarchical Bayes” (HB) methods. In this method, the analysis is carried out by considering each individual as a sample from a population of similar individuals, and “borrowing” information from other individuals in the estimation for each one.

Individual level effects are estimated for four segments. Respondents of each segment are taken out for segment level analysis. To estimate the individual level coefficients, Hierarchical Bayes regression (HB Reg) module of Sawtooth software is used for data analysis. This software estimates individual level coefficients considering individual level data. While doing simulation, the ordinary least square (OLS) estimates are taken as the starting values of the parameters. Since the target distribution of the simulation is not dependent on the starting value of parameters, starting with OLS estimate would not affect the posterior parameter value. It will only lead to relatively quick convergence. As the extreme data points may hinder rather than help the convergence, trial runs are carried out with both OLS starting value and zero starting value. It is found that the convergence is reached faster when simulation is started with OLS starting value.

1,40,000 iterations in the simulation process were used to reach convergence. These draws were burnt off (not saved or used for estimating parameter). A study of the visual graph reveals that the parameters have converged in their posterior distribution. After the chain converged, 40,000 draws were picked up from subsequent iteration and every 100th draw was saved for each coefficient. These saved draws were used to estimate effects of each dimension.

The following table provides the segment wise individual level Bayesian estimates of the effects of underlying dimensions of information seeking behavior on several dependent variables. . Table 5 shows the effects of four segments on one's inclination to use web /e-mail as a source of information.

Table 5: Segment wise effects of Information seeking behavior dimensions on one's inclination to use web /e-mail as source of information: Individual level Bayesian estimation (IBE)

	Segment 1 (IBE)	Segment 2 (IBE)	Segment 3 (IBE)	Segment 4 (IBE)	Aggregate Level
R²	53.8%	72.6%	42.9%	48.9%	68.8%
Std. Error of estimate	1.057	0.613	0.697	0.494	0.842
Coefficient of new Knowledge seeking behavior	1.204**	1.61**	1.164**	1.014**	1.147**
Coefficient of specific document search behavior	0.313	0.197**	0.116	.233**	0.220**
Coefficient of mode of search behavior	.762**	0.725**	0.339*	0.486**	0.449**
Coefficient of related document search behavior	0.122	0.037	0.022	0.02	0.062

***Denotes that the coefficient is significant at <1% level of significance*

**Denotes that the coefficient is significant at <5% level of significance*

Table 5 proves that segment 2, with individual level model, has a better explanatory power than the aggregate level model (72.6% as opposed to 68.8%). In the case of the other three segments, however, the explanatory power is less than that of an aggregate level model. The standard error of the estimate is found to be significantly lower (better) in individual level model than that of aggregate level model in three segments. It is also evident from this that the coefficient of new

knowledge seeking behavior has significantly higher impact on segment 2 respondents than overall (aggregate level) respondents. It is interesting to note that while aggregate level coefficient of ‘specific document search behavior’ is significant, the same is insignificant for segment 1 and segment 3. Similarly, coefficient of ‘mode of search behavior’ is significant at <1% level for aggregate level estimate, however, for segment 3, it is significant at <5% level but not at <1% level of significance. It is also found that the effect of ‘mode of search behavior’ is significantly higher in the case of both segment1 and segment 2 respondents than that of on aggregate level respondents.

Table 6: Segment wise effects of Information seeking behavior dimensions on how long one uses online database: Individual level Bayesian estimation (IBE)

	Segment 1 (IBE)	Segment 2 (IBE)	Segment 3 (IBE)	Segment 4 (IBE)	Aggregate Level
R²	37.4%	42.4%	25.6%	48.5%	57.8%
Std. Error of estimate	1.123	0.521	0.849	0.741	0.952
Coefficient of new knowledge seeking behavior	0.514*	0.745**	1.024**	1.573**	1.116**
Coefficient of specific document search behavior	-0.208	-0.045	-0.098	-.224**	-0.081
Coefficient of mode of search behavior	0.613**	0.306**	0.20*	0.628**	0.019
Coefficient of related document search behavior	0.064	0.023	0.043	0.143	0.062

***Denotes that the coefficient is significant at <1% level of significance*

**Denotes that the coefficient is significant at <5% level of significance*

Table 6 displays the effects of four segments on one’s usage of online databases. It is found from Table 6 that the aggregate level model has better explanatory power than individual level model for all the segments. It is also revealed that the standard error of the estimate is found to be significantly lower in an individual level model than that the aggregate level model in three out of four segments. Hence the efficiency of the estimation is better in case of Bayesian analysis when considering individual

level data. It is also evident from the finding that the coefficient of new knowledge seeking behavior has significantly greater impact on segment 4 than other segments as well as overall (aggregate level) respondents. It not only means that the effect of the underlying dimension increases for a particular segment of people, but also that it can have a better impact in terms of new knowledge seeking behavior on the people of segment 4. It is interesting to note that while aggregate level coefficient of ‘specific document search behavior’ is not significant, the same is highly significant (significance level <1%) for segment 4. Another interesting find is that the coefficient of ‘mode of search behavior’ is insignificant for aggregate level estimate, but it is significant in all the segments and highly significant (<1%) in segment1, 2 and 4, while in segment 3, it is significant at <5% level but not at <1% level. It is also found that the effect of ‘mode of search behavior’ is significantly high in both segment1 and segment 4 and has a strong positive impact.

Table 7: Segment wise effects of Information seeking behavior dimensions on how often one uses online database: Individual level Bayesian estimation (IBE)

	Segment 1 (IBE)	Segment 2 (IBE)	Segment 3 (IBE)	Segment 4 (IBE)	Aggregate Level
R²	78.3%	43.1%	63%	37.2%	67%
Std. Error of estimate	0.99	0.793	0.672	0.894	0.859
Coefficient of new knowledge seeking behavior	1.611**	1.333**	1.439**	1.057**	1.154**
Coefficient of specific document search behavior	-0.407*	-0.298**	-0.336**	-0.051**	-0.209**
Coefficient of mode of search behavior	-0.064	0.274*	-0.522**	-0.392**	-0.351**
Coefficient of related document search behavior	0.355*	0.144	0.128	-0.049	0.099*

***Denotes that the coefficient is significant at <1% level of significance*

**Denotes that the coefficient is significant at <5% level of significance*

Table 7 shows the effects of four segments on one’s inclination to use web /e-mail as source of information. It is established from Table 7 that segment 1 with

individual level model has a better explanatory power than the aggregate level model (78.3% as opposed to 67%) while the explanatory power of other three segments is less than that of aggregate level model. The standard error of the estimate is found to be lower (better) in individual level model than that of aggregate level model in three segments out of four segments. Hence the efficiency of the estimates is better in case of Bayesian analysis considering individual level data. It is also evident from the finding that the coefficient of new knowledge seeking behavior has significantly greater impact in segment 1, 2 and 3 respondents than overall (aggregate level) respondents. It means that the effect of the underlying dimension increases for three segments of people and can have better impact in terms of new knowledge seeking behavior on the people of these segments. An interesting find here is that coefficient of 'mode of search behavior' is highly significant for aggregate level estimate as well as for segment 3 and segment 4; however, it is insignificant in segment 1. The finding also shows that while coefficient of 'Related search behavior' is significant in aggregate level, it is insignificant in three segments out of four segments though it has a much stronger effect in segment 1 which is more than three times that of aggregate level estimate.

Discussions and Conclusion:

The KMO test and the Bartlett test of sphericity indicate strong association or correlation between the variables. Thus it can be concluded that individual users' information seeking behavioral characteristics are not discrete or independent but have adequate association within each other. The four information seeking behavior dimensions that emerged in the factor analysis suggest that there exist four underlying dimensions of information seekers. These information seeking behavioral dimensions are based on the type of information they want and the manner in which it will be used. . Thus it can be concluded that an individual's information seeking behavior would show different effects on the type of documents being searched or the mode of search availed during the search process.

In this research, first information seeking behavior of the information seekers is calculated at an aggregate level. Then by using Hierarchical Bayesian Statistics the same characteristics are calculated. It shows different results in every step. The results obtained from the study of every behavioral dimension at an individual level deviate from the aggregate level study of the respective behavioral dimensions. In a few cases

it shows reverse results, i.e., reverse effect when the calculation is done in individual level.

In previous researches on information seeking behavior the population was always grouped according to their demographic variables, i.e. age, sex, educational level, subject specialization, working place, occupation etc. These people were grouped according to any one or two of the above characteristics and then the information seeking behavior was observed. But people having the same demographic characteristics may not have the same choice or may not act similarly at the time of searching information. For example, behavioral characteristics of information seeker having same educational level or same age might vary due to the differences in their culture, their previous knowledge, and their background.

In this study, parameters are estimated for each individual and then they are grouped into segments according to their observed behavioral characteristics. Then the regression coefficients are compared. Thus, this grouping may give more accurate results at the time of estimation of behavioral pattern. This finding is very useful for information providers in the context of designing information center. If the regular information seekers can be grouped according to their behavioral characteristics, they can be served in a better way. This will save the time of both the information user and the information provider.

We selected MBA students as a sample. Though the sample has similar demographic profile in terms of age and education level, they behave differently and this is due to individual behavioral characteristics. At the same time it is found that students having different age group, different subject of specialization, different occupation, different experience, and having local variation also are on the same segment, i.e. behave similarly. The behavioral characteristic of each segment is similar and different from the other segment. By considering the previous behavior of the information seeker the information providers can serve the sources of information in such a way that they can be satisfied by getting their required information easily.

While analyzing the effect of individual information seeking behavior dimensions on one's inclination to use web/e-mail as a source of information, it is found that new knowledge seeking behavior has a very significant effect on one's usage of web/e-mail. When a person tries to gather new knowledge, he/she would not have prior knowledge about the exact source of that information. At the same time, the user would like to know if the information is easily available or not. Then

searching the web would be the first choice for the information seekers. Web would provide them some source of the new knowledge that the information seeker is looking for. The other two information seeking behavior dimensions i.e., specific document search behavior and mode of search behavior also have significant positive impact on one's inclination to use web /e-mail as source of information. However, the intensity of the effect of these two dimensions against the first dimension is considerably less. It can be concluded that web/e-mail usage may not have significant effect on the information users who are looking for specific information. . In practice it is seen that the information seekers would search their required material directly when the source of information is known to them or the source is related to the known source. Again, at the time of searching known documents users prefer printed documents and the search is direct, which would also save the time of the user. However, in case of searching new documents the web /e-mail search is the best way chosen by the users. This may be due to easy accessibility of the document containing information.

Consequently, our research reveals different aspects of information seeking behavior, which the earlier researches have not done. Considering individual level heterogeneity this research has managed to accomplish an important and pioneering work in the area of information seeking behavior research. Behavioral researchers in information science have investigated information seeking behavior under several contexts and user profiles. Most of the research on information seeking behavior consists of queries by type of subject or by age or by the gender of information seeker. Here user is not considered as an individual, but understood in the context of his cultural system where the effect is determined as average effect of all the information seekers. This study emphasizes the importance of information seeking behavior in the context of personal behavior at an individual level and finds that there exists considerable difference in the behavior when it is considered at an individual level as opposed to the entire population at a time. This conceptualization can also be applied with several characteristics of information needs and information seeking behavior.

This research reveals several interesting and thought provoking findings when dealt with individual level study. We find that the explanatory power of the individual level study does not differ much from the aggregate level study. However, the segment level analysis shows how the actual segment level results differ from the

aggregate level analysis results. For example, while estimating segment wise effects of information seeking behavior dimensions on one's inclination to use web /e-mail as source of information through individual level Bayesian estimation (IBE), we find that aggregate level result explain the model up to 68.8% whereas the results of segment 3 of the individual level analysis is 42.9%. The individual level result for the segment 3 is significantly less than the aggregate level model. However, this result gives the actual behavioral responses of respondents of segment 3. Thus as a decision maker, one would like to work with this actual result while making a decision for the segment 3 rather than consider the results obtained from the aggregate level analysis which might lead to a wrong decision. The same can be observed in all the segment level analysis results. Thus it can be concluded that the individual level analysis shows the actual effects of the behavioral responses which would give more accurate predictive power of decision making for the information providers.

We find that the standard errors of estimates are significantly lower in individual level analysis than that of the aggregate level analysis. The standard error of the estimates in almost all the segments is significantly less, showing the higher efficiency of the estimates. Hence the efficiency of the estimates is better in case of Bayesian analysis considering individual level data.

Considering the estimated coefficients of mode of search behavior (Table 6), it is found that coefficient of 'mode of search behavior' is insignificant for aggregate level estimate. However, it is significant in all the segments and highly significant (<1%) in segment1, 2 and 4, and in segment 3, it is significant at <5% level. It is also found that the effect of 'mode of search behavior' is significantly high in both segment1 and segment 4 and has strong positive impact on the web search behavior of the information seeker. From the segment profile it is found that both these segments are constant users of online sources in their search for information. Hence, the information providers can devise their plan of the mode of search for each segment since it has significant effects on the different segment. However, if they focus on aggregate level effect, they would simply remove this dimension from their consideration since this dimension does not have any significant effect at an aggregate level.

In this present study, information needs identifies the user's subjective opinion and action of several characteristics of individual of a specific population. This study involves a deliberate choice decision of the individual that includes different types of

information search which may be a specific type of information, related type of information or new information. The individual level heterogeneity study suggests that different seekers behave differently while seeking information for their need. Aggregate level behavioral study does not consider this individual variation and it may lead to wrong decision making by the information providers. Behavioral assumptions underlying aggregate level information seeking process do not provide a behavioral explanation of different characteristics of the information seekers at individual level which may be captured very efficiently through this individual level heterogeneity study.

Future research

This study provides an extended view in information seeking behavioral research. It also gives several avenues for future research. . Carrying out the research in a complete unit level approach may give additional findings in the area of information seeking behavior. Replicating this approach for paid and non-paid information could be an interesting area to study. Future research would, perhaps, concentrate on children's information seeking behavior as the world of electronic information increases rapidly. Soon, different children will have different areas of interest to be catered to. Research in areas that can be addressed using individual level analysis proposed in this study is looking into design systems that adapt as users develop competence and provide areas for exploration and learning. Future research may also try to find out what aspects of searching behavior is universal (can be done through aggregate level analysis) and identify situation-/ respondents specific (individual level analysis) cases, if a system is available that can serve an increasingly diverse end-user population such as students or different levels, children, adolescent as well as mature information seekers.

REFERENCES

1. Ainslie, A. & Rossi, Peter E. (1998). Similarities in choice behavior across product categories. *Marketing Science*, 17 (2), 91-106.
2. Allenby, Greg M. & Ginter, J.L. (1995). Using extremes to design products and segment markets. *Journal of marketing research*, 32 (4), 392 – 403.
3. Belkin, N. (1980). Anomalous State of Knowledge as a Basis for Information Retrieval. *The Canadian Journal of Information Science*, 5, 133-143.
4. Bettman, J.A. (1978). Consumer information acquisition and search strategies. In A.A. Mitchell (Ed.), *The effects of information on consumer and market*

behavior. Chicago: American Marketing Association.

5. Brown, C.M. (2007). The role of web-based information in the scholarly communication of chemists: Citation and content analysis of American chemical society journal. *Journal of the American Society for Information Science and Technology*, 58 (3), 2055 – 2065.
6. Buente, W., & Robbin, A. (2008). Trends in internet information behavior, 2000 – 2004. *Journal of the American Society for Information Science and Technology*, 59 (11), 1743 – 1760.
7. Cronback, L. J. (1951). Coefficient Alpha and internal structure of tests. *Psychometrika*, 16, 297-334.
8. Dervin, Brenda. & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology*, 21, 3-33.
9. DeVellis, R. F. (2003). *Scale development: Theory and applications*, Second Edition. Thousand Oaks, CA: Sage Publications.
10. Eisenberg, M.B. & Berkowitz, R.E. (1992). Information problem solving: The Big Six Skills approach to library & information skills instruction. *School Library Media Activities Monthly*, 8 (5), 27.
11. Ellis, D. A (1989). Behavioral approach to information retrieval system design. *Journal of Documentation*, 45 (3), 171-212.
12. Foster, Allen (2004). A non-linear model of information-seeking behavior. *Journal of the American Society for Information Science and Technology*, 55 (3), 228-237.
13. Haines, L.L., Light, J., O'Malley, D., & Delwiche, F.A. (2010). Information-seeking behavior of basic science researchers: implications for library services. *Journal of the Medical Library Association*, 98 (1), 73-81.
14. Hjørland, Birger (1997). *Information seeking and subject representation: an activity theoretical approach to information science* (213p). Westport, CT: Greenwood Press.
15. Howard, Gardner (1993). *Multiple intelligences: The theory in practice*. New York: Basic Books.
16. Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction. *Journal of Documentation*, 52 (1), 3-50.
17. Ippolite, P. & Mathios, A.D. (1990). Information, advertising and health choices: a study of the cereal market. *RAND Journal of Economics*, 21(3), 459-480.
18. Johnson, J.D. & Meischke, H. A (1993). Comprehensive model of cancer-related information seeking applied to magazines. *Human Communication Research*, 19 (3), 343-367.
19. Kaiser, H. F. & Cerny, B. A. (1979). Factor analysis of the image correlation matrix. *Educational and Psychological Measurement*, 39, 711-714.
20. Krikelas, James. (1983). Information seeking behavior: Patterns and Concepts. *Drexel Library Quarterly*, 19(2), 5-20.

21. Kuhlthau, C. C. (1991). Inside the information search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42 (5), 361-371.
22. Lenk, P., W. DeSarbo, P. E. Green & M. Young (1996). Hierarchical Bayes conjoint analysis: recovery of part-worth heterogeneity from reduced experimental design. *Marketing Science*, 15 (2), 173 – 191.
23. Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review*, 50 (4), 370-396.
24. McKenzie Pamela J. & Carey Robert F. (2000). What's wrong with that woman? Working Paper, University of Western Ontario.
25. Natter, Martin & Markus Feurstein (2002). Real world performance of choice based conjoint models. *European Journal of Operation Research*, 137 (2), 448-458.
26. Nicholas, D., Huntington, P., Jamali, H.R., & Watkinson, A. (2006). The information seeking behavior of the users of digital scholarly journals. *Information Processing and Management*, 42, 1345-1365.
27. Robertson, S. E. (1977). Theories and models in information retrieval. *Journal of Documentation*, 33 (2), 126-148.
28. Rossi, Peter E., Greg M. Allenby & R. McCulloch (2005). *Bayesian Statistics and Marketing*, 1st ed. USA: John Wiley.
29. Solomon, P. (1997). Discovering information behavior in sense making: I Time and timing. *Journal of the American Society for Information Science*, 48 (12), 1097-1108.
30. Sonnenwald, D. Evolving perspectives of human behavior: contexts, situations, social works and information horizons. In T.D. Wilson and D. Allen (Eds.) *Exploring the context of information behavior. Proceedings of the 2nd International Conference on Research in Information Needs, Seeking and use in Different Contexts*, 13–15 August 1998, Sheffield, UK (pp. 176–190). London, England: Taylor Graham
31. Taylor, Robert (1968). Question-Negotiation and Information Seeking in Libraries. *College and Research Libraries*, 29 (3), 178-194
32. Taylor, Robert. S. (1985). Information values in decision contexts. *Information Management Review*, 1(1), 47-55.
33. Wessel, C.B., Tannery, N.H., & Epstein, B.A. (2006). Information-seeking behavior and use of information resources by clinical research coordinators. *Journal of the Medical Library Association*, 94 (1), 48-54.
34. Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, 37 (1), 3-15.
35. Wilson, T. D. (1994) Information needs and uses: fifty years of progress? (Ed.), in B.C. Vickery, *Fifty years of information progress: a Journal of Documentation review*, London: Aslib. pp. 15-51.
36. Wilson, T. D. (1999). Models in information-behavior research. *Journal of Documentation*, 55 (3), 249-270.

INDIAN INSTITUTE OF MANAGEMENT KOZHICODE

<i>Type of Document: (Working Paper/Case/ Teaching Note, etc.)</i>	<i>Ref. No.:</i>
WORKING PAPER	IIMK/WPS/109/ MKTG /2012/12
<i>Title:</i> MICRO MODELING OF INDIVIDUAL LEVEL INFORMATION SEEKING BEHAVIOR: A HETEROGENEITY SPECIFIC STUDY	
<i>Author(s):</i>	<i>Institution(s)</i>
Mahuya Adhikary	
Atanu Adhikari	Professor Indian Institute of Management Kozhikode IIMK Campus PO-673 570, Kozhikode,Kerala,India, Phone: 91-495- 2809241), email: atanu.adhikari@iimk.ac.in
Chaitali Dutta	
<i>Subject Areas :</i> Marketing Management	<i>Subject Classification Codes, if any:</i>
<i>Supporting Agencies, if any:</i> N.A.	<i>Research Grant/Project No.(s):</i> N.A.
<i>Supplementary Information, if any:</i>	<i>Date of Issue:</i> May 2012
	<i>Number of Pages:</i> 26
<i>Abstract:</i> The responses of individual information seekers are best studied in isolation. In the past researchers did not consider information users as individuals; viewing them instead in the context of their cultural system as a whole. This study emphasizes the importance of heterogeneity in information seekers, arguing that the behavior and responses of different seekers would vary based on their demographic and socio-graphic profile, the extent of their background knowledge, and the variations in their risk adversity and cost consciousness. In order to understand this aspect of the information seeker’s behavior, information seeking behavior dimension was derived. The effect of this information seeking behavior dimension on the users was investigated. An in-depth questionnaire was administered to 307 respondents. Factor Analysis, Cluster Analysis, Multiple regression and Bayesian Regression were used on the responses in order to identify (i) information seeking behavior dimensions, (ii) different segments of information seekers, (iii) the effect of behavioral dimensions on the behavior of information seekers and (iv) the differences between the effect of these dimensions on users, both at an individual level and at the aggregate level. It was revealed that there is a significant difference in the behavior when the aggregate level estimates are considered and when individual level Bayesian estimates are made. These findings were discussed to understand the implications and to provide avenues for future research.	
<i>Key Words/Phrases:</i>	
<i>Referencing Style Followed:</i> American Psychological Association	