A Statistical Assessment of a Process To Evaluate the Commercial Success Of Inventions

by

Glen Charles Sampson

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2001

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Acknowledgements

I would like to thank Dr. Thomas Åstebro, Associate Professor in the Department of Management Science at the University of Waterloo, for his advice and guidance on my research. In addition, I would like to thank Dr. Åstebro for his patience. I would also like to thank Dr. Paul Patterson, Chair in the Management of Technological Change at the University College of Cape Breton, for his generous support of my graduate work. Two other individuals at the University of Waterloo who I would like to extend special thanks to are Marion Reid and Lynne Wight for their patience and invaluable assistance enabling me to work on my thesis from Nova Scotia. There are also two individuals who deserve thanks for supporting my communication with Dr. Patterson at the University College of Cape Breton. These two individuals are Lee Singleton and Jennifer Pino. I would also like to thank my parents, Charles and Sally, as well as my brother, Scott, for their endless support and encouragement. Finally, I would like to thank my wife Catherine for her never-ending love and encouragement.

I also acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada's joint program in Management of Technological Change.

Abstract

In over twenty years of operations the Canadian Innovation Centre has evaluated, through its Inventor's Assistance Program (IAP), the commercial potential of over 12,000 early stage inventions. Prior to 1989, the Canadian Innovation Centre (CIC) used a version of the Preliminary Innovation Evaluation System (PIES), developed by Gerald Udell at the Oregon Innovation Center in 1974, to evaluate the commercial potential of early stage inventions. Since 1989, the CIC has used a modified version of PIES in their evaluation process. I first estimate the ability of this program's analysts to forecast the probability that an invention will become commercialized. I also estimate a model to predict the probability that an invention will become commercialized based on the IAP's evaluation of several underlying early stage characteristics of the invention. I find that such a statistical model is based on a limited set of variables and predicts future commercial success almost as well as the IAP's forecast of success. I then use factor analysis to determine if the ratings provided by the CIC evaluation service are representative of the underlying theoretical variable structure of PIES or their modified version. Factor analysis is applied to two distinct periods that are separated by a distinct alteration of the theoretical variable structure in 1989. While I find that the factor analysis provides evidence that the post 1989 theoretical structure does provide interpretation of some of the dimensions in the ranking variables, when a combination of the post 1989 and the pre 1989 structure are examined interpretability of the extracted factors is significantly improved. Finally, I compare the model estimated on the underlying early stage characteristics with a model estimated on the extracted factors. When the predictive accuracy of the two models is compared, I find that both procedures produce models that predict almost equally well. The models and the IAP perform better than R&D managers' predictions of their own R&D projects' successes. The thesis provides recommendations for the assessment and maintenance of evaluation models for inventions, innovations and R&D projects.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Introduction

A common myth is that in the modern economy of large scale R&D, independent inventors do not contribute significantly to economic development.[1] However, the percentage of U.S. patents granted to U.S. independent inventors has actually remained steady at about 13% between 1983 and 1996 (USPTO, 1997). Thus, although independent inventors might represent a small fraction of the general population, their contribution to the total amount of patents granted is certainly not ignorable. Similarly, the economic returns to independent inventors' development activities are not ignorable. Åstebro (1999) estimates the return on an investment in inventive activity by an independent inventor to be between 3.5% and 8.7% above a comparable stock market investment.

Given the contribution to the development of technology by independent inventors, and their inventions' economic significance, it is not surprising that programs designed to help inventors and entrepreneurs to commercialize their ideas are becoming increasingly popular. The number of non-profit inventor support organizations in the U.S. exceeded 150 in 1991

[1] Throughout this thesis the terms "I" and "my" are used. However, chapters 1, 2, 4, 5 and section 3.1 are co-authored by Dr. Thomas Åstebro and build upon Åstebro and Sampson (1999). Section 3.2 as well as chapters 6, 7, and 8 are contributed soley by the author. Furthermore, the author was not involved in the data collection process which was performed by Åstebro and others.

(Wall Street Journal, 1991), and at least some of these are partially or fully funded by federal and state government.

In general, government supported programs for small businesses and entrepreneurs have been found to have positive effects on clients' sales and employment (e.g., Chrisman et al. 1985). For example, focusing on "strategic planning assistance", Robinson (1982) reports that sales and employment increases enjoyed by clients of the U.S. Small Business Development Center program were significantly greater than what they would have been had no assistance been given. As well, Jarmin (1999) shows that clients that receive advice from the manufacturing extension program run by the National Institute of Standards and Technology have on average between 3.4% and 16% higher labor productivity growth than those that do not.

In this thesis I focus on the predictive accuracy of so-called Inventor's Assistance Programs (IAPs). IAP's help entrepreneurs evaluate a specific idea or invention *before* it has reached the marketplace and advise the potential entrepreneur on whether and how to continue efforts (Udell, 1989). Although IAP's have been in existence since 1974 they, to my knowledge, have not been subject to rigorous analysis regarding their ability to forecast the success of inventions. Udell and colleagues have written voluminously on the topic of evaluation of inventions and on supporting independent inventors. But none of his publications conclusively show that the evaluation process he originally devised in 1974, which spread to Universities and non-profit organizations in the USA and Canada in the 1970's, has any predictive accuracy. Recent publications, however, indicate that the recommendation delivered by the IAP at the Canadian Innovation Centre to an inventor regarding the commercial prospects of their proposed invention does correlate positively with the invention's subsequent probability of commercial success (Åstebro and Bernhardt, 2000; Åstebro and Gerchak, 2001). But a positive correlation between a recommendation and an invention's subsequent commercialization might simply be a function of self-selection. That is, those reviewed by the IAP as having low prospects and therefore are recommended to stop might be more likely to stop development efforts without necessarily having lower inherent technical or commercial qualities. Similarly, those reviewed by the IAP as having high prospects and therefore are recommended to continue development efforts might be more likely to continue development efforts without necessarily having higher inherent technical or commercial qualities. In this thesis I therefore take a closer look at what drives the probability of commercial success of inventions developed by independent inventors.

First, I estimate the ability of the CIC's IAP to forecast, at an early stage of an invention's development, the probability that an invention will subsequently become commercialized. I investigate whether this correlation is driven by some underlying characteristics of the invention or whether the correlation might be attributable to self-selection. Second, I examine alignment and appropriateness of the theoretical variable structure that the CIC indicates as underlying their ranked variables. I use factor analysis to determine the correspondence between the factors and the theoretical structure both before and after a substantial change was made to the structure in 1989, enabling the change to be assessed. Finally, I estimate the ability of the empirically determined factors to forecast the probability that an invention will become commercialized and compare the predictive capability to the original models based on the ranked criteria.

To perform these estimations a group of 1,095 inventors that received advice from the IAP at the Canadian Innovation Centre (CIC) in Waterloo, Ontario, Canada, were randomly sampled and compared to the *ex ante* evaluations and recommendations by the CIC with the *ex post* probability of success of these 1,095 inventions. Data on a broad range of characteristics of the inventions and their projects were collected and recorded by the CIC during the period between 1976 and 1993. Data on the outcomes of these projects were collected in 1996.

My analysis relates to a rather large body of literature on the subject of new product and R&D project evaluation (for reviews see Balachandra and Friar, 1997; Lilien and Yoon, 1989). This body of literature has not come to any conclusions regarding the predictors of success of R&D projects. Balachandra and Friar (1997) attributes this failure to several methodological problems associated with previous studies. They identified four major sources of weakness in previous studies, namely quality of data, the definition of a new product, factor selection and definition, and measurement of factors. My method resolves most of them. First, I use a large sample of observations in relation to the number of predictors investigated and therefore avoid problems associated with incorrectly accepting the null hypothesis (of no relationship) due to low power of the test (Rosenthal and Rosnow, 1991). Second, the data were collected using standard and accepted sampling and survey techniques rather than based on convenience. Third, my definition of project success is objective (not subjective) and is easily replicated. Fourth, I estimate models using both the raw variables and factors determined through factor analysis. When factor analysis is used it produces interpretable factors. Fifth, the measures of predictors of success were taken at an early stage of the development of the inventions whereas success was measured after the projects were completed and independent of the collection of data on predictors. I therefore avoid two methods bias: hindsight bias (Fichhoff, 1975) and common method variance bias (Campbell and Fiske 1959).

The results indicate several findings. First, the IAP assessments predict success better than R&D managers predict success of their own efforts. Second, the findings indicate that the post outcome results are not due to self-selection. Finally, the factor analysis demonstrates that the changes to the theoretical variable structure in 1989 were warranted.

These results add to the body of research in two ways. First, by providing empirical evidence to support the validity of IAP assessments. Second, by demonstrating the effective use of factor analysis for the assessment and maintenance of IAP evaluation models. Research and practice within this field could benefit through consistent use of both these empirical techniques.

Chapter 2
The IAP

2.1 The IAP at the Canadian Innovation Centre

IAP's were first launched in 1973 in the U.S.A. A result of these early efforts was a venture evaluation system (Udell, 1989) that was used at the Canadian Innovation Centre in Waterloo, Canada (CIC) from its inception. Since the IAP's inception in 1976, and through 1999, over 12,000 inventions have been evaluated. The IAP was launched at the University of Waterloo in 1976 and moved to the newly founded non-profit organization CIC in 1981. During 1976-1981 the IAP at the University of Waterloo normally used 2-3 evaluators, who were either professors at the University of Waterloo or other experts. Since 1982 the CIC has used full-time in-house analysts and continuously revised and improved their evaluation method. Up until 1998 the program's budget was supported 50% by the Canadian government and 50% by service fees. By the fall of 1998 government support had been discontinued and service fees to inventors were subsequently doubled.

The CIC evaluates the potential entrepreneurs and their inventions on 37 different criteria. These are in four groups: technical, production, market and risk factors. Thirty-three of these criteria were developed by Gerald Udell at the Oregon Innovation Center in 1974 as critical for venture success (Udell, 1989), and were used at Waterloo from the start in 1976. In 1989 the CIC introduced a revised list with four more criteria.

To have an idea/invention evaluated, the inventor/entrepreneur fills out a questionnaire and pays a fee. The fee for an independent inventor in 1995 was Cdn. $262 (about U.S.

$185). In addition to background information about the entrepreneur, the questionnaire asks for a brief description of the idea. It asks questions regarding the idea, and asks for supplementary documentation such as patent applications, sketches, and test reports. The questionnaire also asks about market information, manufacturing, product costs, and the entrepreneur's skills, plans and professional goals. In comparison with models of R&D project's success (Cooper, 1981), this review does not consider organizational factors, since all IAP reviews were for independent inventors where there is usually no organization to value. The in-house analyst compares the submission to other similar submissions and searches various on-line databases. The review is confidential.

The analyst then subjectively rates the idea on 37 criteria. A weight is assigned to each factor and an overall score is determined. The weight is not derived based on a statistical analysis but is perceptual. The perceived relative weights might therefore differ across evaluations and evaluators. Analysts have a group meeting where the evaluating analyst presents a summary and a final rating is agreed upon. The evaluation process typically takes five to seven hours. There are five possible ratings (with some minor variations): E - unacceptable, strongly advise project termination; D - doubtful, one or more factors strongly unfavourable, advise project termination; C - possible, may be modestly successful, invention has merit as a part-time endeavor; B - invention looks promising but more information is needed; A - invention is worth commercializing by full-time entrepreneur. Rating B serves to advise the inventor what information is missing and the inventor is asked to collect the additional information before determining whether to continue further work.

The amount of advice delivered by the CIC has grown over the years. In the beginning reports basically contained the overall score, ratings on each criteria (with explanations) and a few comments. In later years the report has come to be a 25-30 page document. It contains the overall score and the scores on each of the 37 criteria, summaries of information searches, specific recommendations about how, if at all, to commercialize the idea (five options), and recommendations on how to approach critical weaknesses.

Chapter 3
Literature Review

3.1 Before Market Product Prediction Literature Review

In chapter two I reported how inventions are assessed in the IAP at CIC. For more analysis of IAP's see Åstebro and Bernhardt (2000), Åstebro and Gerchak (2001), Udell (1989), and Udell et al. (1993). While Åstebro and Bernhardt (2000) analyze the social value of an IAP and Åstebro and Gerchak (2001) analyze the value of information provided by an IAP to inventors, there are no good statistics on the predictive accuracy of these processes. In particular, while Åstebro and Gerchak (2001) conclude that IAP's provide valuable information that inventors don't have, one cannot deduce from this study how well analysts at the IAP perform compared to other alternative evaluation models or procedures. It should be noted, however, that it is hard to find information on the predictive accuracy of any project evaluation process, be it of inventions, innovations, or R&D projects. However, this lack of analysis of model accuracy is not due to a shortage of studies on the subject.

There has been a great deal of research on the determinants of new product performance. The research falls into three perspectives: 1) research on factors leading to success (Cooper, 1984; Yoon and Lilien. 1985); 2) research on factors leading to failure (Constandse, 1971; Voss, 1985); and 3) research on factors that separate success from failure (Cooper, 1979; Cooper, 1985; Maidique and Zirger, 1984; Yoon and Lilien, 1985). In general, these studies suggest normative strategies to enhance success or avoid failure and have provided considerable evidence that a great number of factors can influence the outcomes of new product development. The factors studied describe various combinations of product characteristics, development processes, organizational characteristics, strategic factors and market environmental factors. Recent reviews can be found in Balachandra and Friar (1997) and Lilien and Yoon (1989). Rather than re-reviewing this extensive literature I provide a few examples and then proceed to draw conclusions for my research design based on the excellent meta analysis by Balachandra and Friar (1997).

Hopkins (1981) and Lazo (1965) found that the major reasons for the failure of new industrial products were inadequate market analysis, product deficiencies, higher than expected costs and poor timing in development and introduction. A more recent study (Abratt and Lombard 1993) indicates similar results. Cooper (1981) conducted a study on 195 new product projects to compare/contrast success (102) and failure (93). A factor analysis on forty-eight variables was conducted to generate a smaller and more manageable subset of predictors. The purpose for performing factor analysis was to solve the problem of inter-correlation among the original variables and to reduce the number of predictors to increase the power of tests of significance. Thirteen factors were identified and they explained 69.3 percent of the variance of the original forty-eight variables. A total of seven of the thirteen factors were significantly related to perceived project success at least at the 0.10 level. These were (in decreasing order of significance): product superiority and uniqueness, project/company resource compatibility, market need/growth/size, economic disadvantage to consumer, newness to firm, technological resource compatibility and finally market competitiveness. The model had an R^2 of 0.42 and an overall prediction accuracy of 84.1 percent, and performed well in a naïve split-sample test. As far as I know this is the only available study that has produced some data on the predictive accuracy of a project evaluation model.

These and other studies find little common agreement as to the relevant "success factors" (Balachandra and Friar 1997: BF97). Performing a meta-analysis of 60 papers BF97 found contradictory results and little stability of success factors across these studies: there seems to be no clear agreement on the direction of influence of the factors analyzed. BF97 went on to isolate only one study from an author, rather then assessing several papers from the same author(s) based on the same data, and further deleted those studies with little empirical content or results. This did little to clear up the confusion. BF97 found that among 72 compiled significant factors across nineteen studies (which typically used factor analysis), half of the significant factors were unique to specific studies and about 75% of the final factors were identified in just one or two studies. Even for similar factors their meaning and interpretation may not be the same because of the differences in context across the studies.

BF97 identified four major sources of weakness in previous studies, namely quality of data, the definition of a new product, factor selection and definition, and measurement of factors. I will discuss these issues in some detail as my method resolves most of them.

Most studies on the determinants or factors influencing the success of new product development and R&D projects have been conducted by simultaneously collecting information on independent and dependent variables *after* the projects have been completed (e.g. Cooper 1981, Lilien and Yoon 1989, Maidique and Zirger 1985, Yap and Souder 1994). These studies therefore suffer from both common method variance bias (Campbell and Fiske 1959) and hindsight bias (Fischhoff 1975). Since data on independent and dependent variables are collected at the same time by the same method from the same respondents the measured associations are larger than what would otherwise have been the case. In addition, recollection of conditions after the fact overstates causal relationships.[2]

BF97 lament the bias towards equal representation of successful projects in the studies they reviewed although in reality there are likely to be nine failures for each success (Griffin 1997). The approach typically used is to ask for one successful and one failed project from each firm. A matched sample is thus obtained with a mean probability of success around 0.5. This does not necessarily lead to bias in other parameter estimates, as implied by BF97. Indeed, Maddala (1983, pp. 90-91), shows that if one draws separate samples from two populations only the constant term changes, not the slopes. If one knows the true frequency of successes in a population it is merely a matter of rescaling the constant term to find the population-level intercept. However, I agree that researchers have rarely considered bias in parameters to be of critical concern. The generation of samples seem to have been driven more by convenience then by an attempt to provide generalizeable results (e.g. Maidique and Zirger, 1984).

A serious problem in previous work on predicting the outcome of R&D projects is the relatively sparse amount of observations that are used. Maidique and Zirger (1985), for example, estimate the relationship between 60 independent variables and the outcome of

---

[2] A typical example of such methods bias is the study by Maidique and Zirger (1984) where managers were asked to select a pair of innovations, one success and one a failure, and then asked to determine, for each innovation, whether a particular variable related to the outcome, (positively or negatively) or not. This procedure is likely to overstate relationships.

R&D projects using 59 respondents, where the highest number of observations reported for any variable was 52 and the lowest was 24. It is well known that a multivariate analysis requires about 20 observations *per variable* to be accurate. The above study fails dismally on this point, and the authors resort to bivariate associations, which are less useful as they do not inform of the relative importance of any one variable and an overall model prediction is not possible.

If a sample is of the character that the number of variables is high and the variables are likely to be collinear then one may want to use data reduction techniques to reduce model complexity and increase interpretability (Cliff, 1987). Researchers in this field have typically preferred factor analysis for data reduction (e.g. Cooper, 1981). Factor analysis requires large samples to be successful. Nunnally (1978) writes: "A goof rule is to have at least 10 times as many subjects as variables." (p. 421) and Cliff (1987) suggests: "With 40 or so variables, a group of 150 persons is about the minimum, although 500 is preferable." (p.339). Previous analyses of factors associated with the success of R&D projects all fail to reach the preferable sample size and only Cooper (1981) reaches the absolute minimum, as defined by Cliff (1987). This failure creates instability of results in test-retests and partly explains the lack of convergence on a stable set of “success factors”.

An additional issue identified by BF97 is the wide variety of R&D projects and new product development projects studied. This is not necessarily a methodological failure, as BF97 suggest, but related to the inherently large variance in the types of R&D projects conducted. The large variance reduces coefficient estimates unless one chooses to study a more well defined subset of all types of projects to generate more reliable but less general results.

There is also confusion in the use and meaning of derived success factors (BF97). This confusion is partly a result of using factor analysis with limited sample sizes, which produces unstable results, and possibly due to the lack of instrument validation, rather than confusion, per se, among the researchers.

There is no reason why there should be any confusion about the interpretation of the dependent variable. Nevertheless, there is. Cooper and Kleinschmidt (1987) identify three measures of success – financial performance, opportunity window and market share. Lilien and Yoon (1989) add another dimension, the length of the development process. A study by Griffin et al. (1993) illustrates the divergence of views about how to define successful projects. Summarizing their meta analysis approach BF97 states that "Since there is no common measure of success, and success is a composite of a number of subjective and objective measures, we have used success as defined by the individual authors of the studies" (p. 277).

Summarizing BF97 and drawing conclusions as to the design of a well executed study on project success I consider one of the major deficiencies in previous studies on predictors of success to be that relationships may be severely misrepresented by the method of collecting data on independent and dependent variables at the same time and after the fact (Campbell and Fiske, 1959; Fischhoff, 1975). This study therefore relies on data about independent variables collected before the projects started and on data about outcomes after the projects finished.

I am also concerned that previous results are biased by non-random sampling techniques and will rectify this by paying close attention to the sampling methodology. I, in addition, am concerned that previous analyses have been hampered by low sample sizes leading to unreliable results. Åstebro and others therefore select a "large enough" sample that satisfies statisticians (e.g. Cliff, 1987). Finally, I am concerned that previous research has not clearly defined the type of projects studied and the dependent variable measuring success, leading to confusion about how to interpret results. I therefore spend some time clearly defining the sample and provide an objective and easily interpretable measure of success.

Since the fairly comprehensive review by Balachandra and Friar in 1997, several authors have attempted to provide further insight into the assessment of inventions, innovations and projects. First, some research effort (Perlitz, Thorsten and Schrank 1999; McGrath and MacMillan 2000) has been applied to the area of migrating the options technique from finance to R&D project success. Second, some authors (Cooper 2000; Murphy and Kumar 1997) have criticized the research of the 1990's for focusing on the process as opposed to project selection. Finally, at least two articles (Murphy and Kumar 1997; Davidson, Clamen and Karol 1999) emphasized the importance of empirical testing and metrics for evaluation models.

According to Perlitz, Thorsten and Schrank (1999), the real options valuation technique has been a subject of growing interest for assessing R&D projects. In their (Perlitz, Thorsten and Schrank 1999) article discuss the pros and cons of utilizing real options for project evaluation. The process they discuss requires new decisions being made at various stages as new information is obtained through the ongoing process of product development and introduction. An organization is essentially investing enough at one stage to gain the option to invest at the next stage with an understanding of the overall value of the option (Perlitz, Thorsten and Schrank 1999). However, McGrath and MacMillan (2000) refer to their model as using real options reasoning, but their focus is on the overall option value and early stage assessment rather than the stage gate approach that seems to make the option principle more useful according to Perlitz, Thorsten and Schrank (1999). Both the theoretical structure and the process of the STAR system (McGrath and MacMillan 2000) possess similarities to previous models. The theoretical structure presented seems to draw upon the work of both Udell (1989) and Cooper (1981), but with useful additions and alterations pulled from the management of technology literature. The process for implementation of the STAR system (McGrath and MacMillan 2000) is virtually identical to the process recommended by Cooper (1981) for the implementation of his NewProd system.

The new product development literature of the 1990's focused on improving the process of new product development to the detriment of project selection research (Cooper 2000; Murphy and Kumar 1997). According to Murphy and Kumar (1997), the front end of new product development, from generation of an idea to its approval for development, remains a neglected topic in the development literature. Cooper (2000) succinctly stated that the focus in the 90's has been on doing the project right, but doing the right projects is critical. Murphy and Kumar (1997) reiterate the same thought when they point out that improving the development process downstream while neglecting upstream stages may be a fruitless exercise.

Finally, two articles provide support for the development of metrics for assessing the accuracy and the maintenance of new product evaluation models. Murphy and Kumar (1997) point out that very little attention has been paid to empirically testing the activities performed in the front end of new product development. Davidson, Clamen and Karol (1999) stress the importance of the maintenance of both performance and process metrics. They indicate that in order to maintain the effectiveness models and processes these metrics must be continually monitored. The result is empirical data that can be used to adjust, improve and provide feedback (Davidson, Clamen and Karol 1999).

Overall, since the fairly comprehensive review by Balachandra and Friar in 1997, research has investigated new alternatives, melded new alternatives with old and continued to criticize the lack of research effort at the front end of the new product development process. Therefore, I seek to add to this area of literature by providing empirical research at the front end of the new product evaluation process.

3.2 Factor Analysis Literature Review

While the Canadian Innovation Centre has evaluated over 12,000 inventions since its inception in 1974, it has modified the variables it uses in its innovation evaluation system several times. While the CIC made minor adjustments throughout the years, the most notable alteration was in 1989 when the original 33 variables that were developed by George Udell at the Oregon innovation center (Udell 1989) became 37 variables. For this thesis, it is important to note that the CIC not only increased the number of variables but also re-categorized a considerable number of the original variables. In other words, the underlying variable structure underwent extensive alterations in 1989. Although it is somewhat uncertain what precipitated this reorganization, there are two possible explanations that are not mutually exclusive. First, the CIC, after fifteen years of innovation evaluation experience, assumed that they possessed sufficient knowledge to alter the underlying variable structure that defined the Preliminary Innovation Evaluation System (PIES) as originally developed by Udell. A second, plausible explanation is that the CIC attempted to merge external research such as Rogers' Diffusion of Innovation (Rogers 1995) or Cooper's New Product Project Selection Model (Cooper 1981) with the PIES system to arrive at their new system.

Interestingly, Udell indicates knowledge of such alterations to the original PIES system (Udell, 1989). Although his PIES system has spread to a vast array of innovation centres across North America, many of the practitioners, whether intentional or not, are using, according to Udell, inappropriate systems for their evaluations. In 1989, Udell points out that some centres where using outdated versions of the PIES system due to the spread of unlicensed or illegal copies of the computerized PIES system. Others, as the CIC had done in 1989, had taken it upon themselves to make additions, deletions, and alterations to the original system. Unfortunately, many of these ad hoc adjustments neglected to utilize the considerable research that Udell, and others, had completed through the Oregon project. Indeed, 18 research reports and over 100 articles and papers, most of which dealt with the innovation process, independent inventors, or the PIES format, were published or presented by 1980 as a result of the Oregon project (Udell 1989). In fact, Udell claims that some centres made changes to the PIES system that were counter to the research (Udell 1989).

However, innovation centres may have had no other choice but to eventually make alterations to the PIES system. At the end of the Oregon project, further research and development of the PIES system was transferred to a small for-profit enterprise, which admittedly resulted in a reduced level of research and development into the PIES system (Udell 1989). Even Udell, himself, altered the criteria or variables used from the original 29 to 39 in 1989 (Udell 1989).

The intent of the factor analysis within this thesis is two-fold. First, to determine if the alterations made by the CIC to the PIES system were warranted. Second, to determine if the changes were in fact real. To arrive at a determination to satisfy this second goal I ask whether the changes in the theoretical latent variable structure beneath the CIC rated variables are reflected by the ranking procedures used by the analysts. This assessment is possible because, while the CIC re-categorized the variable structure, they left most of the ranked variables the same. Through factor analysis, it can be determined if the rankings provided by the CIC analysts do indeed reflect this restructured theoretical structure.

3.2.1 Two Existing Models

Although there is considerable research into the topic of innovation assessment and R&D project assessment [For reviews see Cooper (1983), Lilien and Yoon (1989), and Balachandra (1997)] I will examine the relevance of two key models. First, Udell (1989) claims that the PIES format was derived through factor analysis. He indicates that initially a considerably larger number of criteria or variables had been identified through the work at the Oregon Center, but an informal factor analysis had reduced the number to 29. While I am uncertain as to the exact meaning of "informal" I must assume that the process possessed some relevance to the practice of factor analysis to be referred to as such. Regardless of the details of this initial effort, Udell (1989) continues by stating that a later formal factor analysis indicated that the number of criteria could be reduced to about eight. In other words, the latent variable structure of the initial PIES structure consisted of eight factors when an actual factor analysis was performed. However, these eight factors seemed to have been dismissed since they failed to provide sufficient information to provide useful feedback to the submitter of the innovation. Although Udell (1989) mentions the considerable volume of research that was conducted on the PIES format, there is no mention of factor analysis to assess the ongoing validity of the underlying variable structure.

The work of Roger G. Cooper is relevant to this thesis for several reasons. First, within the R&D project assessment literature, especially when focusing on the assessment and prediction of success of early stage projects, Cooper provides one of the most comprehensive bodies of work. Second, Cooper developed a system, NewProd (Cooper 1985; Cooper 1992), that might be considered comparable to the PIES format (although it should be noted that much of Cooper's work is focused on R&D projects within large and medium sized firms while the PIES format was designed specifically for innovation centres that tend to assess a large portion of submissions from independent inventors). Finally, for a period of time Mr. Cooper was the director of research at the Canadian Industrial Innovation Centre (now the Canadian Innovation Center or CIC), which is the source of the data for this thesis. Given these facts, it is feasible to assume that some of Mr. Cooper's work could have influenced the modification of the PIES format at the CIC; either directly or indirectly. However, I will not make any specific assumptions regarding such influence and will simply proceed to provide a brief overview of his relevant work.

Cooper utilized factor analysis as a method in his work. As an example, in his 1981 study, Cooper started with 48 variables and employed exploratory factor analysis to reduce the number of predictors to thirteen and then regressed these constructs on a continuous measure of perceived success. Although Cooper was obviously interested in determining the underlying factors that influenced R&D project success, one of his main goals in the use of factor analysis was data reduction. He was faced with the obvious statistical problem of regressing a large set of variables (48) on one dependent variable (success of a project) with a dataset consisting of a comparatively small sample (195 cases). Indeed, a number of Cooper's studies (Cooper 1987; Cooper 1994; Cooper, Kleinschmidt and Elko 1993; Cooper and Kleinschmidt 2000) involve sample sizes of one to two hundred cases and usually these samples are multiple projects from a smaller number of firms. Therefore, Cooper's factor analysis work tends to focus on exploratory research with a key purpose being data reduction for further analysis.

Although there is some consistency in the factors extracted by Cooper's research, the authors are unaware of any factor analysis work to assess the continued validity of the factors determined by Cooper. It is also unknown if others have used altered versions of the Cooper (1981) structure to develop their own models and, if so, whether such activity relied on factor analysis to continually assess the validity of the variable structure.

3.2.2 Appropriateness of the Data Matrix for Factor Analysis

Prior to performing factor analysis I examined the data matrix to determine if it was appropriate for factoring. The literature indicates several methods for such tests. First, two of the simplest procedures for determining the appropriateness for factoring are the examination of the correlation matrix and the plotting of the latent roots obtained from matrix decomposition. If the correlation coefficients are small throughout the matrix, factoring may be inappropriate (Stewart 1981). In addition to the examination of the correlation matrix and plotting the latent roots, there is Bartlett's test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy (MSA). While Bartlett's test was found to be inappropriate for the large sample set of 581 due to it's virtual certainty to reject the independence hypothesis when the sample size is greater than 200 (Stewart 1981), the MSA (Kaiser 1970) was found to be suitable and provided a calibration scale (Kaiser and Rice 1974) that the data set could be assessed by. Furthermore, according to Stewart (1981) the MSA appears to have considerable utility and may be the best of the methods currently available for measuring dataset appropriateness for factor analysis.

3.2.3 Selecting an Appropriate Factor Extraction Method

While the various factor extraction methods are often debated, Stewart (1981) claims that when communalities are high, as they are in this data set, there are virtually no differences among the procedures. Furthermore, Stewart's (1981) research concludes that empirical evidence comparing the results obtained from principal components, principal factors, alpha analysis and maximum likelihood analysis supports this statement.

In addition, since previous factor analysis work has been undertaken with the same dataset, it is useful to examine these previous efforts. In this previous work (Åstebro, Michela and Zhang 2001) the principal components method was found to offer the best orthogonal decomposition of the correlation matrix in the most straightforward way.

3.2.4 Determining the Appropriate Number of Factors to Extract

Without question, determining the number of factors to retain is the most debated topic in factor analysis. As Stewart (1981) indicates, this is an important decision since too many factors will result in factor splitting and too few factors can seriously distort the solution. As a result, researchers have devised numerous methods and procedures. Each of these has their respective supporters and critics: First there is the roots criterion where factors are extracted until the eigenvalue of the next factor to extract is less than one. In other words, all factors with eigenvalues greater than or equal to one are extracted. To add to the confusion, this extraction rule is referred to by a number of names. Among them are the Kaiser rule, eigenvalues=1, Kaiser-Guttman test, and the K1 rule. A second test is Bartlett's which tests the hypothesis that the correlation matrix is not significantly different from the identity matrix (Stewart 1981). This test, which is applicable to the maximum likelihood extraction, is a chi-square test of significance that assesses the significance of the residual variation and provides a score and p-value for interpretation. Finally, there is the scree test that involves plotting the eigenvalues and determining where the curve bends or the slope experiences a considerable decrease. Due to the controversy that surrounds the number of factors to extract, I will address each of these choices in an attempt to justify my selections. The researcher is not limited to these techniques as numerous others are available, however, I will concentrate my discussion on these three commonly used methods.

First the roots criterion where the eigenvalues equal one is one of the most commonly used methods for determining the number of factors and is a built in feature in virtually all statistical software. The research community has a range of opinions on this rule. First, Gorsuch (1974) determined that it was accurate with a small to moderate number of variables but particularly inaccurate with a large number of variables, which he defined as greater than forty. Cattell and Vogelman (1977) concluded that it tends to extract too many factors when a large number of variables are used and too few factors when a small number of variables are used. Zwick and Velicer (1986) found that the K1 rule severely and consistently overestimated the number of factors to retain. Stewart (1981) claims that the test is only suitable as an indication of the minimum number of factors to extract and not the maximum. Finally, Cattell and Vogelman (1977) give the harshest criticism claiming that the Kaiser test is simply misleading as a single test.

While some of these opinions may seem contradictory, I concluded that, although the number of variables could be considered bordering on large at 37, this test is the most common and should be examined and reported. It should also be noted that previous research with the same data matrix (Åstebro, Michela and Zhang 2001) used this rule to determine the number of factors to extract.

The second test, Bartlett's, is also widely used to determine the number of factors to extract in a factor analysis procedure. However, Gorsuch (1974) determined that Bartlett's was only applicable for determination of the maximum number of factors and concluded that the test was not appropriate as a routine for selecting the number of factors. Furthermore, Zwick and Velicer (1986) found that Bartlett's chi-square test was less accurate and more variable than the scree test.

Given these comments, I concluded that Bartlett's chi-square was not the best method for determining the number of factors to extract. I was further influenced to reject this method by previous research (Åstebro, Michela and Zhang 2001) that attempted to use Bartlett's chi-square test with a maximum likelihood extraction. This effort found too many factors and many variables without salient loadings.

The scree method is one of the simplest to apply and, according to much of the literature, one of the most accurate. The process involves plotting the eigenvalues and assessing where the curve bends or the slope experiences a considerable decrease.

Many researchers provide empirical evidence attesting, in whole or in part, to the accuracy and usability of the scree test including: Cattell and Dickman 1962; Cattell and Sullivan 1962; Cattell and Gorsuch 1963; Cliff and Hamburger 1967; Linn 1968; Tucker, Koopman, and Linn 1969; Cliff 1970; Stewart 1981; Zwick and Velicer 1982; and Zoski and Jurs 1996. For example, Tucker, Koopman and Linn (1969) found the scree test to be accurate in 12 of 18 attempts. Cliff (1970) found the scree test to be accurate, particularly when questionable components were included. While Zwick and Veliver (1986) did not recommend the scree test as their number one choice, they clearly preferred it to both Kaiser's method and Bartlett's chi-square. Finally, Cattell and Vogelman (1977) found the scree test to be more accurate than the Kaiser rule.

Yet the scree is certainly not without criticism. First, Horn and Engstrom (1979) conclude that the scree recommends the extraction of too many factors in small samples. Second, Linn (1968) also indicates that the scree was more interpretable when using formal model matrices as opposed to simulation model matrices. Finally, while Cattell and Vogelman (1977) report that application of the test is quite straightforward and consistent, others (Zwick and Velicer 1986) imply that training in the technique is required to obtain consistent results. Furthermore, Zwick and Velicer (1986) indicate complications that can introduce problems when applying the scree test such as a continuous gradual slope or multiple breaks in the line.

However, when I examine the properties of my specific dataset the evidence is quite supportive of the scree method. Our dataset, by most research standards, is considered large having 581 cases and 37 variables. The research provides strong support for the use of the scree in datasets of this size. Linn (1968) found that changes in the slope of the curve were more readily identifiable in 40 variable matrices than in 20 variable matrices and also in sample sizes of 500 more so than in samples of 100. Cliff and Hamburger (1967) also noted the improvement in detectible slope reductions in samples of 400 as opposed to 100. Even Zwick and Velicer (1982) noted that the scree test was most accurate when larger samples were used.

After examining all of the evidence, both pro and con, I selected the scree test as a second test for selecting the number of factors to extract from the variable matrix. By examining both the eigenvalues and the scree plot a decision will be made as to the number of factors to extract. However, it should be recognized that I can compare findings from former analysis (Åstebro, Michela and Zhang 2001) using the Kaiser rule and Bartlett's chi-square performed on the same matrices.

3.2.5 Selecting an Orthoginal Rotation Method

In selecting an orthogonal rotation method I examined several options and settled on the varimax rotation. According to Stewart (1981), the majority of the standard orthogonal rotations result in the same factors. Nevertheless, varimax was selected for two reasons. First, it received strong support from the research (Dielman, Cattell, and Wagner 1972; Gorsuch 1974; Stewart 1981). Second, it is the most commonly used orthogonal rotation procedure.

3.2.6 Selecting an Oblique Rotation Method

Prior to discussing an oblique rotation method, I determined a need to communicate the reason for using an oblique rotation. Although there is controversy concerning the use of oblique rotations due to the allowance of correlation among the resulting factors, this approach is thought to be useful for my particular factor analysis. Not only did Rummel (1970) provide a logical explanation for the use of oblique rotations when he said that the real world should not be treated as though phenomena coagulate in unrelated clusters; but even the previous research (Åstebro, Michela and Zhang 2001) with the same dataset indicated some criteria in different categories seemed likely to covary substantially.

I decided to perform an oblique rotation (promax) for three reasons. First, my goal was to determine the validity of my hypothesis. Since an oblique rotation provides increased insight into the factors that load on each, an oblique method was considered important for providing the level of information required to assess my hypothesis. Second, I believe, like Rummel (1970), that real world factors are rarely unrelated and it is unrealistic for us, as researchers, to ignore that possibility. Finally, the oblique rotation provides us with the actual correlations that concern its opponents, enabling us to assess the level of correlation among the factors and make decisions.

3.2.7 Regressing Success on Factors

Although I recognized the statistical merit in performing factor analysis on the dataset to reduce the correlations, factor analysis also formed a key statistical tool within the relevant research. Many models estimated for the evaluation of R&D projects (Cooper 1981) or independent inventions (Udell 1989) are developed through statistical procedures that include a first stage where factors are extracted from a relatively large set of variables and a second stage where the dependent variable is regressed on the resulting factors.

Although factor analysis has been used to estimate parameters for many models, it is important to note that factor analysis was utilized to determine some of the specific models, such as Udell (1989) and Cooper (1981), that have a direct connection to the evaluation process at the CIC. First, in developing the original PIES format, Udell (1989) utilized factor analysis. This process included both an informal and formal factor analysis extracting twenty-nine and eight factors respectively (Udell 1989). Second, considerable R&D project evaluation work performed by Cooper (1985; 1992) utilized factor analysis to reduce the number of independent variables for further regression analysis. Finally, in a survey of the literature, Balachandra and Friar (1997) indicated that factor selection and definition as well as measurement of the factors are two of the key problems within this area of research. Clearly, the factor problem stated by Balachandra and Friar (1997) is an indication that factor analysis is a common procedure within the field.

Chapter 4
Data

4.1 Data

I selected to study projects undertaken by independent inventors. Their projects are many times less likely to reach the market compared to R&D projects in established organizations (Mansfield et al. 1971, Griffin 1997, Åstebro 1998) and thus represent one of the highest levels of uncertainty. If I find robust predictors for this sample it is likely that other researchers will find the same predictors to appear even stronger for projects with less uncertainty, as uncertainty is likely to reduce precision in estimates. Indeed, I argue that patterns and relationships that appear in this study are robust to high levels of uncertainty.

It is important to recognize that inventions developed by independent inventors may have characteristics that distinguish them from R&D projects undertaken in large established organizations. Except for their lower probability of success, inventions developed by independent inventors are, however, not as different to those developed in large established organizations as one is often led to believe in the popular press. Indeed, inventions that are patented by independent inventors are technically no different in terms of their degree of novelty and their degree of detail in their specification and are as likely to have their patent fees maintained as inventions patented by established firms (Dahlin et al. 1999). Inventions patented by independent inventors do appear, however, to have a narrower scope of application then inventions patented by established firms (Dahlin et al. 1999). It has also been observed that inventions developed by independent inventors have out-of-pocket costs (excluding the inventor's own labor costs) that are about one eighth of project costs in established firms while those projects that succeed are as profitable as R&D projects in established firms (Åstebro, 1998).

Åstebro and others used the Inventors' Assessment Program (IAP) at the Canadian Innovation Centre (CIC) as the source for project data because the IAP advises primarily independent inventors. Indeed, only four per cent of the submissions to the IAP in 1993 were derived from corporations with more than 200 employees (CIC, 1996).

In order to avoid sampling bias, Åstebro and others targeted a random sample of 20% of submission of projects submitted to the IAP at the CIC from each year between 1976 and 1993. The sample frame consisted of all 8,797 valid records of IAP submissions during 1976 to 1993. Initially 3,282 records were sampled. Using a CD-ROM of Canadian residential addresses 1,826 records were updated with current addresses. This number represents 21% of the sample frame. Åstebro and others could not reject the hypothesis that the updated records were a randomly selected subset of the sample frame across the years of submission (χ^2=0.19, d.f.=16, n.s.).

Åstebro and others followed the total survey design method outlined by Dillman (1978). This involved pre-tests of the survey instrument on a sample of inventors and detailed reviews of the instrument by analysts at the IAP. The telephone survey method was chosen for its ability to generate high response rates and for greater control of the data collection process (Lavrakas, 1993). The inventor was first mailed a letter informing that someone would be calling. Telephone calls were made primarily in evenings to residential telephone numbers during an eight-week period in the spring of 1996. Among the 1,826 records, the research team obtained 1,095 responses representing an adjusted response rate of 75%.

An overwhelming majority of inventors who responded are male (89 per cent) and a plurality of their inventions are consumer oriented (47 per cent). However, there is a significant fraction of "high-tech" inventions (6 per cent), and industrial equipment inventions (6 per cent) (CIC, 1996). The majority of inventors (72 per cent) are from the Province of Ontario. A number of tests were conducted to establish that the variation in sampling and response proportions across the years of submissions, provinces in Canada, gender, and rating were random: that is, no selection bias were detected (for details Dr. Thomas Åstebro at the University of Waterloo). The background demographic characteristics of this sample correspond well to other samples of independent inventors (see Albaum, 1975; Parker et al., 1996). In addition, the sample contains very few multiple submissions from the same inventor: Over a period of sixteen years, 1,044 inventors made one submission, 21 made two submissions, and three provided three submissions. There are thus few "professional" inventors in this sample. It is also safe to assume that projects are independent, which simplifies the statistical analysis.

For each record, a file number was assigned by the IAP to indicate the submission year and month. Åstebro and others went back to the IAP to find the physical record of the evaluation information for the 1,095 responses. They obtained evaluation information for 1,093 observations. Evaluation information obtained in the CIC record included ratings for each of the 37 early stage characteristics as well as the invention's overall rating. Data on the independent variables were consequently collected before outcomes were observed and independently of this study. I therefore avoid any potential methods bias (Rosenthal and Rosnow, 1991).

Because the inventions are developed, to a large extent, by independent inventors, and the inventions are assessed by the IAP at an early stage of development, there are few organizational characteristics that are measurable at the time of the evaluation. This should be kept in mind when interpreting results as organizational factors are nevertheless likely to affect the success of projects (BF97). Such an early judgment of a project may, superficially, seem like a waste of time given the potential lack of relevant data and the great uncertainty associated with available data. But this argument is refuted by the study by Mansfield et al. (1977a) who found clear evidence that the earlier assessment of a project the greater the future probability of technical, commercial as well as financial success (pp. 25-32). Therefore, any relationship I find is of interest and is likely to have an even stronger effect when there is less uncertainty.

Åstebro and others were concerned that they would not be able to observe enough successes to estimate meaningful models. The research team therefore included a subset of inventions where analysts at the CIC had information indicating that the invention might have reached the market. Analysts had obtained this information through various sources such as newspaper clippings. There were 75 additional observations included this way. These additional observations increase the analysis sample to 1,170 observations. Forty-eight of the 75 non-randomly sampled inventions were subsequently reported directly by the inventors to have reached the market. In comparison, the random sample of 1,095 inventions contained 75 inventions that had reached the market. Finally, the research team removed four inventions that were reported as having reached the prototype stage of development where I could not yet determine whether the inventions were successful or not. The sample therefore consists of either 1,091 random observations or 1,166 observations where some are choice-based.

Table 4-1, columns (2) and (3), reports the frequency distribution of the responses over the IAP's ratings for the random sample of responses (i.e., excluding the choice-based observations). A majority of submissions (rating D and E) are advised to terminate efforts; fifteen percent receive rating E and 60% receive the rating D totaling 75% of all submissions. Four percent receive the most favorable rating (A), four percent are advised to collect additional market or technical analysis (B), and 18% are advised the innovation is suitable to launch as a limited (i.e. part-time) effort (C).

Table 4-1 also reports the probability of commercial success for each of the different ratings and for the (random) sample as a whole (column 4). "Success" is defined in this case as successfully reaching the market place and selling at least one unit. There can be many definitions of success (Cooper and Kleinschmidt, 1987). Balachandra and Friar (1997) lament the lack of uniformity in the use of measures of success in the literature. The one I employ, however, is easily operationalized, easy to replicate across studies, does not depend on a subjective evaluation, and is certainly a necessary but not a sufficient condition for financial success.[3] [See Åstebro (1999) for a detailed analysis of the financial success of

[3] We asked the inventor "Did you ever start to sell <NAME> or a later revised or improved version of this invention?"

these inventions.] As seen, the probability of success is clearly increasing with the overall rating by the IAP.

Data spanned two submission periods with somewhat different evaluation procedures, with the first period from 1976 to 1989 (early July) including 564 random observations and the second from July 21, 1989 to 1993 including 527 random observations (four observations being deleted). During the first period, 1976-1989, each of the 33 underlying criteria was rated on the scale of 1 to 5 by analysts at the IAP, whereas in the second period, 1989-1993, the 37 criteria were graded on a three-point scale, i.e. A (Acceptable, which means that the criteria appears to be favorable or satisfactory), B (Borderline. The criteria rated as B needs to be improved or strengthened.), and C (Critical Weakness. It usually means it may be necessary to discontinue the effort to commercialize the invention.) Sometimes a B^+ was assigned, as well. For the purposes of multivariate analysis the research team decided to convert the scores on the underlying criteria for the second period into numerical data according to the following: $A = 5$; $B^+ = 4.5$; $B = 4$; and $C = 3$.[4]

For the regression analysis I then included the choice-based observations. The resulting analysis data set for the period 1989-93 consisted of 581 inventions containing 517 failures and 64 successes. While the resulting sample is considerably smaller then originally intended, it is still "large enough" to comply with standard requirements for multivariate analysis (Cliff, 1978).

---

[4] If these scores were used the mean of the scores on each variable for the two periods would come close to equal, which motivates this otherwise rather arbitrary choice of scaling.

Table 4-1 Inventions Developed by Independent Inventors

	1976-1993 sample**			1989-1993 sample***		
Rating	**Sample Total**	**Percent of all**	**Percent Commercial**	**Sample Total**	**Percent of all**	**Percent Commercial**
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A - recommended for development.	24	2%	50%	30	5%	67%
B - may go forward, but need to collect more data.	45	4%	16%	40	7%	25%
C - recommended to go forward, returns likely modest.	204	19%	16%	88	15%	22%
D - doubtful, further development not recommended	655	60%	4%	344	59%	4%
E - strongly recommend to stop further development	163	15%	0%	79	14%	1%
Weighted Average			7%			11%
Total*	1091	100%		581	100%	

* Two inventions miss data on their rating and are not included. Four inventions are deleted as I could not determine whether they were successful or not.

** Data represents a random selection of inventions.

*** Data represents a random selection of 527 inventions augmented by a choice-based sample of 54 inventions.

Columns 5, 6 and 7 of Table 4-1 report information for inventions assessed after 1989 where choice-based observations have been added. The addition of the choice-based observations did not change the distribution across ratings for the full 1976 – 93 sample ($\chi2$=6.39, d.f.=4, n.s.). Neither did the addition of the choice-based observations change the distribution across ratings for the 1989 – 93 sub-sample ($\chi2$=6.34, d.f.=4, n.s.). Maddala

(1983, pp. 90-91), shows that if one draws separate samples from two populations only the constant term change, not the slopes. Given that I show no changes in the underlying distribution with the addition of the choice-based observations, there will be no bias in regression parameter estimates, only a change in the constant. However, the reduction in the sample to cover only 1989 – 93 resulted in a marginally significant change in the distribution across ratings ($\chi 2$=8.94, d.f.=4, p<0.10). The main change is a reduction in the number of inventions rated "C", which is attributable to a policy change at the IAP. After a review, analysts found that those rated "C" were earning very little returns and decided to either be more encouraging or more discouraging to the marginal inventor.

With the inclusion of the choice-based sample there is an increase in the average probability of success from 0.07 to 0.11 and, in particular, an increase in the probability of success for those rated A by the IAP. When I perform chi-squares using the probability of success I find no significant differences. The probability of success distribution across ratings did not change when the full 1976 – 93 sample is compared to the sub-sample 1989 – 93 ($\chi 2$=4.13, d.f.=4, n.s.). Neither did the addition of the choice-based observations change the distribution across ratings for the 1989 – 93 sub-sample ($\chi 2$=4.07, d.f.=4, n.s.). In addition, the increase in the average probability of success can be attributed, in part, to the improvement over time in the services provided by the IAP. The IAP started to use in-house specialists rather then outside experts in 1981 and that likely contributed to a faster rate of learning within the IAP and, possibly, a more profitable interaction between the IAP and the inventor. In later years, and particularly as part of the changeover in 1989, the IAP started to focus on serving inventors with more targeted advice on how to proceed, given that an invention was rated favorably. This, apparently, had quite a large impact on the success rate for those rated "A". Finally, the impact of raised prices for the service over time cannot be ignored, as this is likely to lead to higher quality submissions through self-selection.

4.2 Theoretical Variable Structures

Table 4-2 Post 1989 Theoretical Structure (CIC Modified Variable Structure)

Latent Variable Structure: Category	Sub Category	Ranked Variables
Technical		Technical Feasibility
		Functional Performance
		Research & Development
		Technology Significance
		Safety
		Environmental Impact
Production		Technology of Production
		Tooling Cost
		Cost of Production
Market	Demand	Need
		Potential Market
		Trend of Demand
		Duration of Demand
		Demand Predictability
		Product Line Potential
	Acceptability	Societal Benefits
		Compatibility
		Learning
		Visibility
		Appearance
		Function
		Durability
		Service
	Competition	Price
		Existing Competition
		New Competition
	Effort	Marketing Research
		Promotion Cost
		Distribution
Risk		Legality
		Development Risks
		Dependence
		Protection
		Size of Investment
		Potential Sales
		Payback Period
		Profitability

The structures, contained in Tables 4-2 and 4-3, form the basis for the assessment of a factor analysis procedure. It is important to note that a substantive change occurred in this underlying theoretical structure in 1989. This change is represented by the discrepancies between Table 4-2 and 4-3. These theoretical structures provide the framework to compare my empirically derived factors against. For comparison purposes and to assess the structural change in 1989, I use data from both pre and post 1989.

Table 4-3 Pre 1989 Theoretical Structure (PIES Variable Structure)

Latent Variable Structure Category	Ranked Variables
Societal	Legality
	Safety
	Societal Benefits
	Environmental Impact
Business Risk	Functional Feasibility
	Production Feasibility
	Stage of Development
	Investment Costs
	Payback Period
	Profitability
	Marketing Research
	Research and Development
Demand Analysis	Potential Market
	Potential Sales
	Trend of Demand
	Demand Predictability
	Demand Life Cycle
	Product Line Potential
Market Acceptance	Compatibility
	Learning
	Need
	Dependence
	Visibility
	Promotion
	Distribution
	Appearance
Competitive	Service
	Durability
	Function
	Price
	Existing Competition
	New Competition
	Protection

Chapter 5
Evaluating, Ex Ante, the Ex Post Commercial Success of Inventions

5.1 Method

The purpose of my analysis is to compare the ability of two models to predict at an early stage the probability that an invention will later become commercialized. First, a model was estimated that used the IAP's overall rating as the independent variable (i.e. the single item A through E rating) and whether or not the invention later became successful as the dependent variable. Second, I estimated a model that used the underlying early stage characteristics as independent variables and whether or not the invention later became successful as the dependent variable. Third, I estimated a model that regressed the overall rating on the underlying early stage characteristics of the invention. The third model was estimated to examine whether there was any truth to the claim that the correlation between the overall rating and the probability of success is due solely to self-selection or whether the correlation could be traced to the underlying characteristics of the innovations. In summary, I investigate the extent to which the underlying characteristics are correlated with the overall rating, whether the underlying characteristics determine project outcomes and whether the overall rating determines project outcomes. Note that while my regression analysis is associative between the underlying characteristics and the overall rating, I can infer causality from the impact of the two on project outcomes. Causality can be inferred because the outcome occurs after characteristics are observed.

The comparison of the first two models was based on three criteria. First, overall predictive accuracy was examined. This number was calculated by taking the number of correct predictions of the model and dividing it by the total number of observations examined. Although the overall predictive accuracy of the models is important, it is also critical to calculate the ability of each model to separately predict success and failure. These criteria are commonly called Type I and Type II errors. In this application, I term the Type I error the probability that the model predicts a failure given that the invention is successful. I term the Type II error the probability that the model predicts a success given that the invention fails to reach the market.

Even though I decided to use the logit model, a number of link functions are available for binary logistic regression (see Agresti, 1990). Rather than arbitrarily selecting one function, three link functions were explored: logit, normit (also called probit), and gompit (also called complementary log-log). The general class of models that comprises these particular functions is defined by:

$$g(\pi_j) = \beta_0 + \beta x'_j \quad , \qquad (1)$$

where

π_j = the probability of a response for the j'th factor/covariate pattern;

$g(\pi_j)$ = the link function;

β_0 = the intercept;

x'_j = a column vector of predictor variables associated with the j'th factor/covariate pattern;

β = a row vector of unknown coefficients associated with the predictors.

5.2 Results

After running numerous regressions with the three link functions, I determined that all three generated qualitatively similar results. Knowing that the results were robust to model specification, I selected the logit for further analysis as it is most widely used.

The first model estimated the correlation between the IAP's overall rating and the probability of success. It does not take into account the 37 underlying characteristics and their ratings. This model predicted 457 out of the 581 possible inventions correctly (78.7 percent). Table 1-2 indicates that the model predicted 49 of the 64 possible successes correctly, (76.6 percent), with a probability of a Type I error of 0.234. The model predicted 408 of a possible 517 failures correctly (78.9 percent), with a probability of a Type II error of 0.211.

The IAP's overall rating's ability to predict success can also be calculated by means of a simpler method but which does not use all available information. By taking those inventions that receive a rating of A, B, or C and assuming that they are predictions of success and by taking those inventions that receive a rating of D or E and assuming that they are predictions of failure, the predictive accuracy of the IAP can be assessed. These calculations reinforce the predictions determined through the logistic regression model. However, since the five ratings are collapsed into two categories, valuable information is lost. I therefore refrain from reporting the details of this analysis.

For the second model I used hierarchical regression with a forward variable selection process among the set of 37 criteria (the underlying characteristics of the invention) with probability of success as the dependent variable. I used a p-value of 0.05 to determine the inclusion of predictors. That is, the final model only contains predictors that were significant at the 5% level. Following this forward selection process, the final model consisted of only four independent variables. These are: V23 (Function), V33 (Development risk), V35 (Protection) and V40 (Profitability). This model correctly predicted 441 out of the 581 possible inventions (75.9 percent). The definitions of these significant variables are described in Appendix A. I also investigated results using a backward variable selection process. The variables selected for inclusion using this procedure were identical to the ones selected for inclusion using the forward selection model with the exception of V23 (Function), which was replaced by V2 (Functional Performance). However, the forward selection model using V23 (and not V2) was superior with a predictive accuracy 2.6 percentage points higher. Therefore this model was selected. The results of the forward variable selection was thus:

$$Y = -14.71 + \underset{(0.02)}{0.67 * V23} + \underset{(0.000)}{0.87 * V33} + \underset{(0.01)}{0.57 * V35} + \underset{(0.001)}{1.12 * V40} \tag{2}$$

where Y = the log of the odds of success and the p-values are reported in the parentheses. The model obtained a psuedo-R^2 = 0.20.

Columns 4 and 5 of Table 5-1 indicate that the model correctly predicted 49 of the 64 possible successes (76.6 percent) with a Type I error of 0.234. Table 1-2 also indicates that the model predicted 392 of a possible 517 failures correctly (75.8 percent) with a Type II error of 0.242.

Table 5-1 Predictive Ability of the Two Models.

	IAP Overall Rating		Underlying Characteristics	
	Number	Percent	Number	Percent
(1)	(2)	(3)	(4)	(5)
Correctly Predicts Success	49	0.77	49	0.77
Type I Error Actual success but model predicts failure	15	0.23	15	0.23
Correctly Predicts Failure	408	0.79	392	0.76
Type II Error Actual failure but model predicts success	109	0.21	125	0.24
Overall Predictive Ability		0.79		0.76

I continued by examining whether a statistical model could capture the decisions made by analysts at the IAP. As described in Section 2 the process of determining the overall score is built on analysts' perceptions of the relative importance of the various factors that the invention is judged against, which might change from innovation to innovation. The process also involves a group decision-making process. It is therefore not clear, a priori, whether there is any statistical association between the underlying characteristics and the overall rating. Due to the categorical rankings of the IAP's overall rating having values of A, B, C, D or E, an ordinal logistic regression model was fitted (Maddala, 1983, p. 46). Using a p-value of 0.05 to determine inclusion of predictors in the final model, the resulting model contained 11 of the possible 37 explanatory variables. The variables included in this model were V1 (Technical feasibility), V14 (Duration of demand), V22 (Appearance), V23 (Function), V31 (Distribution), V32 (Legality), V33 (Development risk), V35 (Protection), V37 (Size of investment), V38 (Potential sales) and V40 (Profitability). The pseudo-R^2 of the model was 0.60. I can therefore conclude that there is indeed a strong statistical association between the overall rating and several of the underlying characteristics of the innovations.

At this point, the models using either the IAP's overall rating or the probability of success as the response variables can be compared. A quick examination indicates that all of the predictors that are in the fitted model determining the probability of success are also in the fitted model determining the overall rating. This common set of explanatory variables consists of V23 (Function), V33 (Development risk), V35 (Protection) and V40 (Profitability). Although the model with the overall rating as the dependent variable has six additional variables, the four common predictors provide evidence that the correlation between the overall rating and probability of success depends on the underlying qualities of the inventions rather than self-selection.

To compare the predictive accuracy of the two main models, I examined the overall predictive accuracy of each, the ability of each to predict success, and the ability of each to predict failure. The first model, which uses the overall rating as a predictor of success, is superior to the second model in two out of three instances (see Table 1-2). The IAP's overall rating has a predictive accuracy of 78.7 percent while the underlying innovation characteristics correctly predict 75.9 percent of the observations. In addition, the IAP's overall rating outperforms the underlying characteristics when predicting failures. The overall rating predicts failure correctly 78.9 percent of the time, while the underlying characteristics predict failure correctly only 75.8 percent of the time. However, the underlying characteristics predict success equally well as the overall rating with both models correctly predicting successes 76.6 percent of the time. In other words, the model based on the IAP's overall rating gained its entire victory by being superior at predicting failures.

5.3 Discussion

I found that in making an overall assessment of an innovation at an early stage of development analysts at the IAP in Canada outperformed a statistical model based on the underlying characteristics of the invention (similarly rated by the analysts) in two out of three comparisons while performing equally well on the third. Overall this means that analysts at the IAP's correctly predict 78.7% of the observations, while a model based on the underlying characteristics correctly predicts 75.9% of all outcomes. These prediction accuracies are both higher than R&D department managers' ability to predict the technical success of their own R&D projects, which was estimated by Mansfield (1968) to be 66%. Industrial R&D projects might be perceived as more complex and thus more difficult to predict than inventions developed by independent inventors. However, the R&D projects in Mansfield's study had quite low uncertainty. In about three-fourths of the cases, the estimated probability of technical success exceeded 0.80 (Mansfield, op. cit., Table 3.4). In comparison, the average probability of commercial success for inventions developed by independent inventors varied between 0.07 and 0.11 in my study, indicating a much higher degree of uncertainty. Analysts in the IAP at CIC estimate the likelihood of market success at a very early stage of the inventions' development (on average two years before reaching the market), which undoubtedly is quite difficult (Balachandra and Friar, 1997). It is therefore fair to say that the IAP's ability to predict success compares very favorably to R&D managers' predictive accuracy. The estimated models, however, perform less well than the new product project selection model derived by Cooper (1981), which obtained a within-sample prediction accuracy of 84.1%. But since Cooper's study suffers from methods bias where data on independent and dependent variables were collected at the same time and after the fact it is not clear how well his results stand up against ours, where data on predictors are obtained prior to observing outcomes. It is also not clear how well Cooper's model performs when used in actual decision situations. Results in this article illustrates that the IAP process performs well in actual decision situations. Nevertheless, the problem of subjective ratings of the underlying characteristics remains. While the CIC has built up a strong knowledge base in their IAP, this knowledge is only partially captured by the estimated weights associated with the predictive characteristics. Another important facet of the IAP is the determination of the scores of an invention on each of the underlying characteristics. Our model does not inform the non-CIC analyst how to perform this subjective assessment.

The results may illustrate that the overall rating provided by the IAP influences the inventor to follow the recommendation of the IAP while not necessarily reflecting the actual chances of success. In other words, regardless of the true underlying merits of the invention, if the inventor receives a rating of A, B, or C she will be more likely to continue to pursue the commercialization of her invention. Similarly, if the inventor receives a rating of D or E she will be more likely to stop the pursuit of commercialization. In response to this criticism, I found that the same underlying characteristics that predict project success also predict the overall rating by the IAP. That is, the merits of the invention affect both the IAP's overall assessment and the invention's subsequent probability of commercial success. Therefore, the hypothesis that self-selection is solely responsible for the correlation between the overall rating and ensuing success is rejected.

It is, however, still unknown to what degree the rating by the CIC predicts success as opposed to predicting the decision by the inventor to continue with development efforts. Indeed, the models in question do not take into account whether the inventor curtailed his/her pursuits conditional on a specific rating. The binary variable which represents success is only able to inform us that those who were successful continued to pursue their efforts. Those who continued to pursue their efforts and subsequently failed and those who, at some point, and possibly following a D or E rating from the IAP, decided to discontinue all inventive efforts are both classified as failures. Similarly, those inventors who were given an overall rating of A, B or C and eventually failed, may or may not have proceeded on the recommendation of the IAP. Future research might model this more complex structure of only partially observable decisions using nested logit models.

Chapter 6

Testing a Priori Hypothesis about the Underlying Dimensions

6.1 Hypotheses

The organization of each evaluation report provided by the CIC reflects a theoretical variable structure used during the evaluation process. Prior to 1989, the CIC used a theoretical structure (Table 4-3) based on Udell's (1989) PIES structure. From 1989 onward the CIC used a modified version (Table 4-2) of Udell's PIES structure as the theoretical underpinning of their evaluation process. These two structures and the validity of the alteration in the structure can be tested using factor analysis. The goal is to compare the empirical results of the factor analysis procedure with the theoretical structures that the CIC claims underlie their evaluation process. In addition, the adjustment in 1989 can be further assessed with regression analysis by comparing the predictive ability pre and post alteration. Therefore, I formulate the following hypotheses to achieve these goals:

> ***H_1****: The post 1989 empirical factor structure is aligned with the post 1989 theoretical factor structure presented in Table 4-2. In other words, the empirical factor loadings on the 37 variables indicate alignment with seven distinct factors which are: Technical, Production, Market Demand, Market Acceptability, Market Competition, Market Effort and Risk.*

***H_2**: The pre 1989 empirical factor structure is aligned with the pre 1989 theoretical factor structure presented in Table 4-3. In other words, the empirical factor loadings on the 33 variables indicate alignment with five distinct factors, which are: Societal, Business Risk, Demand Analysis, Market Acceptance and Competitive.*

***H_3**: The post 1989 empirical factor structure is aligned with the pre 1989 theoretical factor structure presented in Table 4-3. In other words, the empirical factor loadings on the 37 variables indicate alignment with five distinct factors, which are: Societal, Business Risk, Demand Analysis, Market Acceptance and Competitive.*

***H_4**: The post 1989 empirical factor structure is aligned with a combination of the pre 1989 and post 1989 theoretical factor structures presented in Table 4-3 and Table 4-2. In other words, the empirical factor loadings on the 37 variables indicate alignment with a mixture of the twelve possible factors which are: Technical, Production, Market Demand, Market Acceptability, Market Competition, Market Effort and Risk from Table 4-2 as well as Societal, Business Risk, Demand Analysis, Market Acceptance and Competitive from Table 4-3.*

***H_5**: The IAP improved its ability to predict success following the alteration of the theoretical variable structure in 1989. Therefore, the empirical regression results using the post 1989 data should demonstrate an improved capability to predict success when compared to the empirical regression results using the pre 1989 data.*

6.2 Methodology

As Rummel indicates, the vast majority of published factor analyses are exploratory (Rummel 1970). Rather than hypothesize the structure of the latent variables prior to analysis, most researchers perform exploratory research and report the factors discovered. Depending on the factor loadings of the various variables involved, the researcher makes an effort to interpret and usually name the discovered factors. Indeed, within the field of innovation assessment, both Udell (1989) and Cooper (1981) utilized this exploratory technique to construct their models.

While I recognize the benefit of the exploratory technique in the initial creation of such models, many variables affect the use of models – especially when the use is continued over extended periods of time. For example, the models used by innovation evaluation centres, such as the CIC, have been used over a twenty five year period with limited, if any, validity tests. Indeed, when alterations are made to such evaluation models the research is rarely empirical and, as Udell claims, sometimes runs contradictory to the existing research (Udell 1989). During these twenty years not only have model variables and structures been altered, but evaluators have come and gone, business climates have changed multiple times and economic factors for success and their relationships may have experienced some change.

Therefore, in order to ensure the validity of innovation models initially developed through factor analysis, ongoing assessments must be carried out. (Although this applies to almost any model developed through exploratory factor analysis, I will focus on the example of innovation assessment.) In the case of underlying variable structure, which presumably serves to guide the evaluator as well as the entrepreneur, inventor or company receiving an evaluation, regular factor analysis could be conducted to assess whether the evaluation and the underlying variable structure are properly aligned. If such regular assessments are not carried out, the evaluation process can become inconsistent and ultimately incorrect. Furthermore, if an alteration is being considered to the variables or the underlying variable structure, it is critical that such a change is assessed through empirical methods such as factor analysis and is not allowed to occur based solely on anecdotal evidence. Beyond factor analysis, several other regular assessments should occur. First, the predictive ability of the assessments should be scored through follow up research on the outcomes of the innovations and statistical assessments of the predictive abilities. Second, alignment between the ranked variables and the overall rating should be regularly assessed through statistical procedures (Åstebro and Sampson 1999). Essentially, if models are going to be used to provide critical assessments of early stage projects some level of quality control must be employed rather than either assuming a models eternal viability or making ad hoc changes.

Although exploratory factor analysis is commonly used for the initial construction of models, it is, unfortunately, not often utilized for the ongoing assessment of the initially determined underlying variable structures. However, this restraint is not due to methodology but to research tradition (Rummel 1970). I believe that much of the debate surrounding the use and validity of factor analysis procedures could be avoided if researchers and the practitioners who use the resulting models, embraced factor analysis as a required follow up to models based on exploratory research. Indeed, to test whether these patterns actually exist (or continue to exist) is the factor analysis task (Rummel 1970).

For hypotheses one through four this thesis postulates, in advance, the underlying variable structure or dimensions that are resident in the empirical factor analysis. As mentioned above, these advance postulations are based on the theoretical variable structures utilized in the invention evaluation process by the CIC and presented in Table 4-2 and 4-3. First, to assess whether the post 1989 rankings of the CIC analysts support the theoretical factor structure in Table 4-2 (Hypothesis 1), I perform factor analysis on the 37 ranked variables for the 581 cases in the post 1989 period. Second, to assess whether the pre 1989 rankings of the CIC analysts support the theoretical factor structure in Table 4-3 (Hypothesis 2), I perform factor analysis on the 33 ranked variables for the 583 cases in the pre 1989 period. Next, to assess whether the post 1989 rankings of the CIC analysts actually support the theoretical factor structure in Table 4-3 (Hypothesis 3), I utilize the empirical results from the factor analysis performed for Hypothesis 1, but determine if the factor loadings map to Table 4-3 instead of Table 4-2. Finally, to assess whether the post 1989 rankings of the CIC analysts support a mixture of the theoretical factor structures presented in Tables 4-2 and 4-3 (Hypothesis 3), I perform factor analysis on the 37 ranked variables for the 581 cases in the post 1989 period. The first two factor analysis procedures test whether the empirical results are aligned with the theoretical structure used by the CIC in their respective periods. The third procedure tests whether the results from the post 1989 period remain aligned with the theoretical structure from the pre 1989 period. The final analysis is designed to test whether the 1989 alteration in the theoretical structure is reflected in the analysts' rankings or whether the theoretical structure from the pre 1989 period remains partially resident in the minds of evaluators during the post 1989 period.

The results section will include my selection process for each of the steps listed within the methodology as well as the outcomes of each step. Following the results, I will provide a discussion and my conclusions concerning my findings. In an effort to assess the first four hypotheses I use the following methodology (See Section 3.2 for a detailed literature review of these steps):

1) Determine the appropriateness of the data matrix for factor analysis.

2) Select an appropriate method for the factor extraction.

3) Determine an appropriate number of factors to extract (in this case I am testing the hypothesis of seven factors for the post 1989 period and five factors for the pre 1989 period).

4) Select an orthogonal rotation method.

5) Select an oblique rotation method.

6) Interpret the results with respect to interpretability and simple structure.

In addition, to assess the fifth hypothesis, regression analysis is performed on the pre 1989 period and the post 1989 period separately. In both instances I regress the success outcome on the ranking variables to determine whether the predictive capabilities of the CIC show improvement following the alteration of the theoretical variable structure.

6.3 Results

6.3.1 Appropriateness of the data matrix for factor analysis

I chose to examine the correlation matrix and assess the MSA on the calibration scale provided by Kaiser and Rice (1974). Examination of the correlation matrix for the post 1989 period data found that the 37 variables were highly correlated with many of the correlations exceeding .70 and most being over .20. The overall MSA score was determined to be 0.87. Examination of the correlation matrix for the pre 1989 period data found that the 33 variables were also highly correlated with many of the correlations exceeding .70 and most being over .20. The overall MSA score was determined to be 0.97.

Given the scale developed by Kaiser and Rice (1974) (.90+ Marvelous, .80+ meritorious, .70+ middling, .60+ mediocre, .50+ miserable, below .50 unacceptable) the post 1989 data set would fall in the category of meritorious and is approaching marvelous while the pre 1989 data set would definitely be classified as marvelous. Therefore, both the correlations and the MSA provide strong indication that both data sets are, indeed, appropriate for factor analysis.

6.3.2 Selecting an appropriate factor extraction method

While the various factor extraction methods are often debated, Stewart (1981) claims that when communalities are high, as they are in these data sets, there are virtually no differences among the procedures. The empirical evidence examined by Stewart (1981) comparing the results obtained from principal components, principal factors, alpha analysis and maximum likelihood analysis supports this conclusion. Although some indicate that the choice of method is not likely to affect results, I consider two factors prior to selecting a factor extraction method.

First, I consider the fact that prior work was done with this dataset and principal components was chosen as the method of factor extraction. In this previous work, the principal components method was found to offer the best orthogonal decomposition of the correlation matrix in the most straightforward way (Åstebro, Michela and Zhang 2001). Second, I consider the fact that the maximum likelihood method is the method of choice when the factor analysis is testing a priori hypothesis (Stewart 1981). Given my focus on testing a priori hypothesis coupled with an inability of the previous work to discover interpretable factors through the use of principal component analysis, I proceed with maximum likelihood extraction.

6.3.3 Determining the appropriate number of factors to extract

I am forming prior hypotheses concerning the number of factors that should be extracted and then testing those hypotheses against empirical results. In hypothesis one, the post 1989 period is hypothesized to have seven factors. In hypothesis two, the pre 1989 period is hypothesized to contain five factors. In hypothesis three, the post 1989 period is hypothesized to have five factors. However, in hypothesis four, although the post 1989 data is assessed using a total of twelve factors they are not designed to be twelve mutually exclusive factors. Therefore, I make the assumption that there are still seven factors[5] but the alignment with underlying factors can be drawn from any of the twelve theoretical factors. Determining the number of factors to retain is likely the most controversial topic in factor analysis. After examining three methods for factor extraction, including Eigenvalues, Bartlett's and the scree, I concluded the most appropriate method for this dataset was to examine both the Scree and the Eigenvalues.

To apply the scree test to the post 1989 dataset I first plotted the eigenvalues using the maximum likelihood method to obtain Figure 6-1. According to Linn (1968), when there is a clear break in the eigenvalue curve there is little question about the correct number of factors. However, Linn (1968) also concludes that when there is no clear break in the curve, there probably is no good answer. As Figure 6-1 indicates, there is a definite change in the slope of the curve. However, the exact point of that change in slope may be open to debate.

Examination of the curve indicates that 7, 10, and 14 factors could all be plausible ends of the scree when the inclusion rule of the last factor on the scree is used. This phenomenon of a split scree line is a regular occurrence. Indeed, the single scree line may well be the exception (Catell and Vogelman 1977). However, when I combine information from the Eigenvalue test, as is recommended in the literature (Zwick and Velicer 1986; Stewart 1981) the picture becomes clearer. I can consider that the eigenvalue of 7 is almost precisely equal to 1 (0.96) and the remainder of the eigenvalues, to follow, obviously drop below 1.

---

[5] Both the Egenvalues and the Scree test indicate that seven is an appropriate number of factors for the post 1989 dataset.

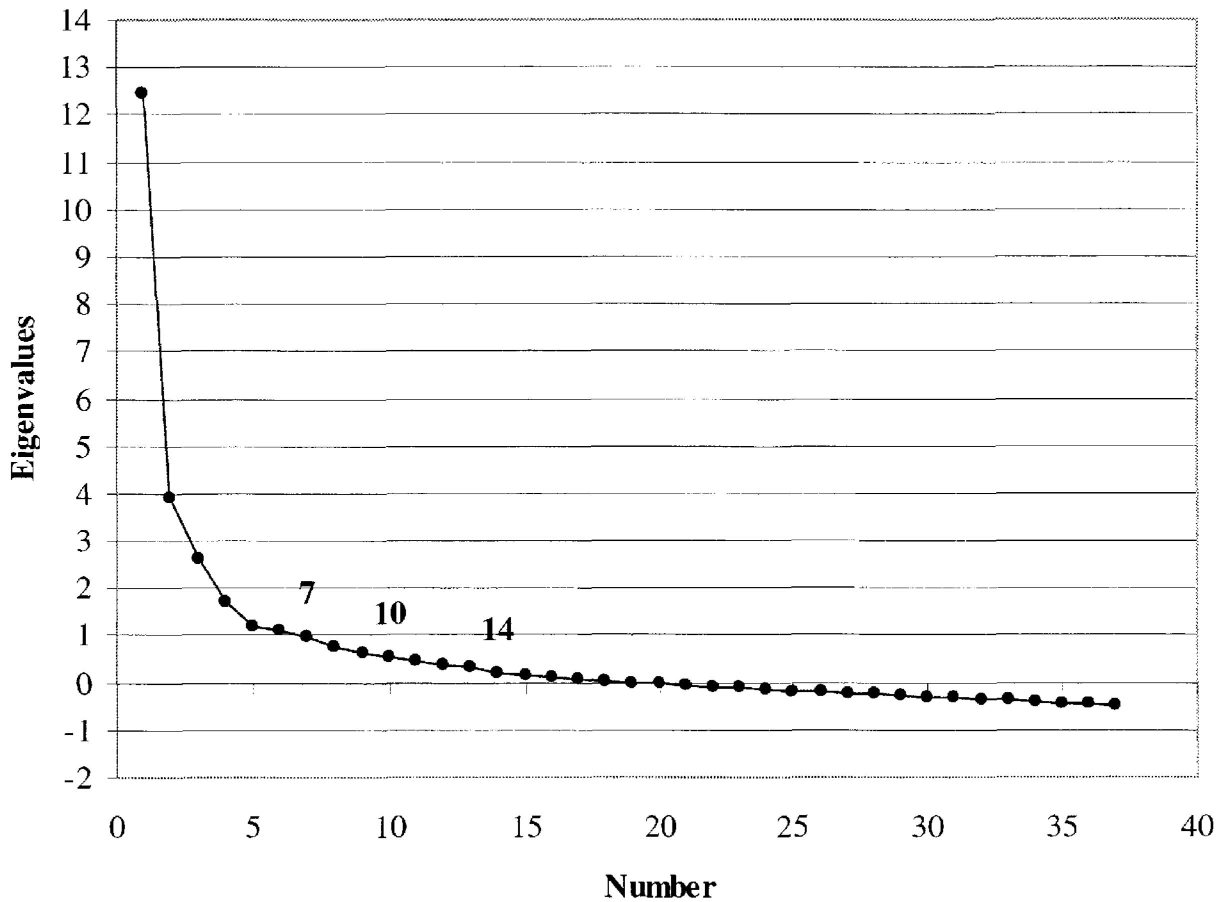

Figure 6-1 Scree Plot of Eigenvalues for Post 1989 Data

Given that both the scree and the eigenvalues indicate that 7 factors is a suitable number for the post 1989 data, I continue towards my hypothesis that the underlying variable structure of the CIC rankings consist of 7 factors. However, I also continue to analyze the 10 factor solution because it serves as a comparison for testing my 7 factor hypothesis.

However, only the 7 factor results are presented in the body of the thesis, while the results of the 10 factor solution are reported in Appendix C.

To apply the scree test to the pre 1989 dataset I plotted the eigenvalues using the maximum likelihood method to obtain Figure 6-2. As Figure 6-2 indicates, there is a definite change in the slope of the curve. However, the exact point of that change in slope may again be considered open to debate.

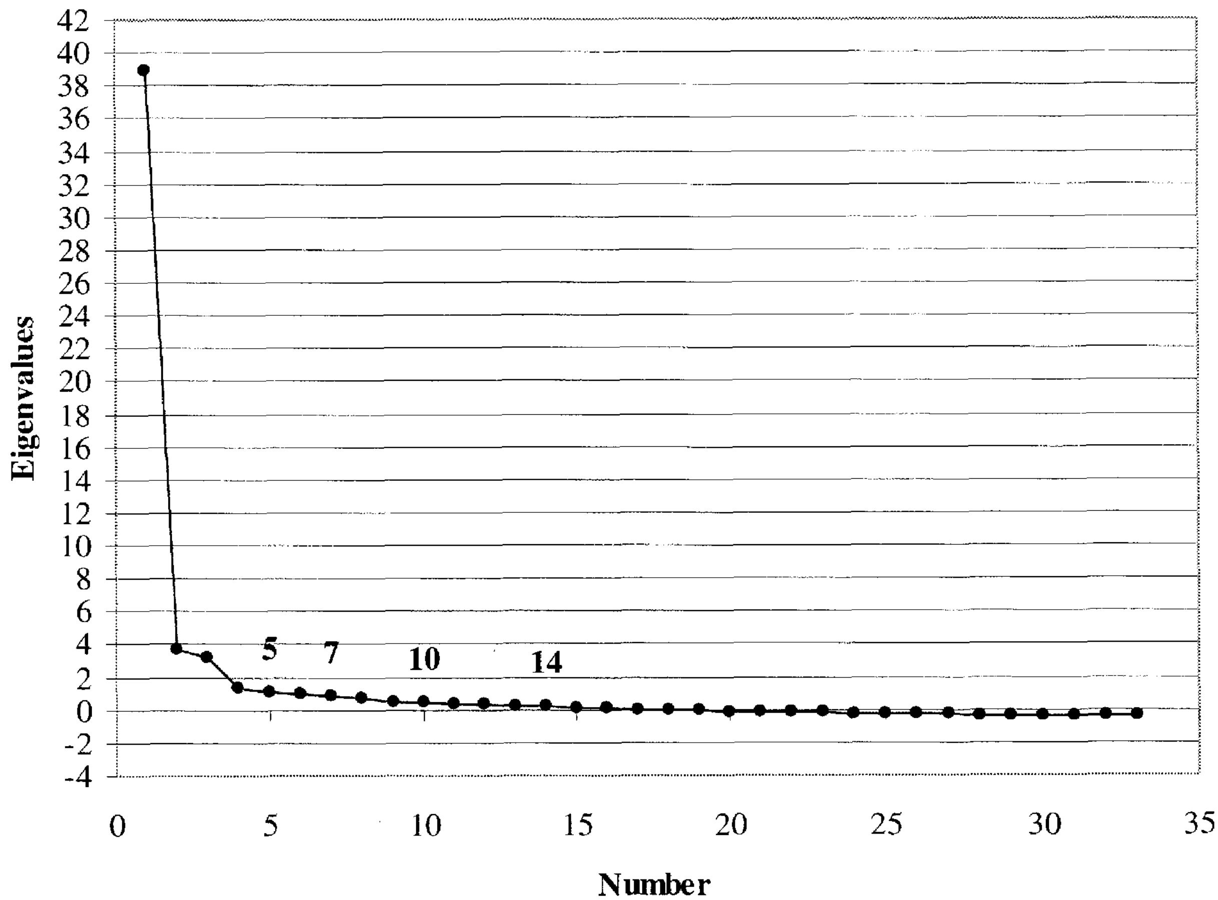

Figure 6-2 Scree Plot of Eigenvalues for Pre 1989 Data

Examination of the curve indicates that 2, 4, and 5 factors could all be plausible ends of the scree. However, when I combine information from the eigenvalue test, as is recommended in the literature (Zwick and Velicer 1986; Stewart 1981) the result becomes clearer. If I consider that the eigenvalue of 5 is 1.15 while the eigenvalue of 6 is 0.85 five factors would be the appropriate number of factors to extract.

Given that both the scree and the eigenvalues indicate that 5 factors is a suitable number for the pre 1989 data, I continued towards my hypothesis that the underlying variable structure of the pre 1989 CIC rankings did indeed consist of 5 factors. However, I also continued to analyze the 7 factor solution because it served as a comparison for testing my 5 factor hypothesis. Only the 5 factor results are presented in the body of the thesis, while the results of the 7 factor solution are reported in Appendix C.

6.3.4 Selecting an orthogonal rotation method

To select an orthogonal rotation method I examined several options and settled on the varimax rotation. According to Stewart (1981), the majority of the standard orthogonal rotations result in the same factors. Nevertheless, varimax was selected for two reasons. First, it received strong support from the research (Dielman, Cattell, and Wagner 1972; Gorsuch 1974; Stewart 1981). Second, it is the most commonly used orthogonal rotation procedure. While the orthogonal method is used with for each analysis, only the results of oblique rotations are presented in the body of the thesis. All orthogonal results are provided in Appendix C.

6.3.5 Selecting an oblique rotation method

Prior to discussing an oblique rotation method, I determined a need to communicate the reason for using an oblique rotation. Although there is controversy concerning the use of oblique rotations due to the allowance of correlation among the resulting factors, this approach is thought to be useful for my particular factor analysis. Not only did Rummel (1970) provide a logical explanation for the use of oblique rotations when he said that the real world should not be treated as though phenomena coagulate in unrelated clusters; but even the previous research (Åstebro, Michela and Zhang 2001) with the same dataset indicated some criteria in different categories seemed likely to covary substantially.

I decided to perform an oblique rotation (promax) for three reasons. First, my goal was to determine the validity of my hypothesis. Since an oblique rotation provides increased insight into the factors that load on each, an oblique method was considered important for providing the level of information required to assess my hypothesis. Second, I believe, like Rummel (1970), that real world factors are rarely unrelated and it is unrealistic for us, as researchers, to ignore that possibility. Finally, the oblique rotation provides us with the actual correlations that concern its opponents, enabling us to assess the level of correlation among the extracted factors and make decisions based on this information.

6.3.6 Interpretability and simple structure

In an effort to determine the validity of each of the hypotheses, one through four, I examined the factor loadings of two types of rotations for two different numbers of factors on each of the two data sets. Since hypotheses one, three and four can utilize the same factor tables, the result is a total of eight factor tables. However, the findings indicate two facts. First, there are minimal discrepancies between the orthogonal rotation and the oblique rotation with respect to their factor loadings. Second, there are minimal differences between the two distinct number of factors tested on each of the two datasets. As previously stated, my goal is to perform factor analysis to test prior hypotheses about the underlying variable structure. Thus, I focus on the rotation method that lends itself to improving pattern identification within the loadings, as I believe that correlation among the resulting factors is a reasonable assumption. For these reasons I will only present the oblique results in detail. Furthermore, I will only present the empirical results of the 7 factor extraction for the post 1989 data and the 5 factor extraction for the pre 1989 data. The remaining factor tables, results and discussions can be found in the appendices.

I examined the loadings for interpretability and assessed their attainment of simple structure. Interpretability, for my purposes, is defined by salient loadings equal to or greater than .40. In other words, if a variable has a loading of .40 or greater on a particular factor, I considered the variable to represent a strong component of that factor. My definition of simple structure, while being aware of Thurstone's (1947) five criteria for simple structure, is based on Catell's (1952) hyperplane count. However, instead of using plus or minus .10 as the range, I selected plus or minus .15 which was found to be common in application.

6.3.7 Interpreted results with respect to interpretability and simple structure

***H**$_1$: The post 1989 empirical factor structure is aligned with the post 1989 theoretical factor structure presented in Table 4-2. In other words, the empirical factor loadings on the 37 variables indicate alignment with seven distinct factors which are: Technical, Production, Market Demand, Market Acceptability, Market Competition, Market Effort and Risk.*

To test the first hypothesis I examine the factor loadings of the oblique ration for seven factors (Table 6-1). In this case the empirical factor loadings for the first empirical factor (F1) are divided across three hypothesized factors in the theoretical structure. First, Technology Production (0.62), Tooling Cost (0.59) and Cost of Production (0.57) all have salient loadings on Factor 1 and are aligned with the hypothesized factor Production. Second, Technical Feasibility (0.65) and Research & Development (0.62) both have salient loadings on Factor 1 as well, but are aligned with the hypothesized factor Technical. Finally, Development Risks (0.49) and Investment Costs (0.55) both have salient loadings on Factor 1, but are aligned with the hypothesized factor Risk.

Table 6-1 Factor Loadings for 7 Factors (Post 1989 Theoretical Structure / Post 1989 Data Empirical Results)

Post 1989 Theorectical Structure			Oblique Rotated Factor Pattern						
Factor	**Variable**		**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
	Technical Feasibility	V1	**0.65**	-0.14	0.12	0.13	0.00	-0.31	0.05
	Functional Performance	V2	0.38	0.05	-0.03	0.36	0.14	-0.22	0.10
Technical	Research & Development	V3	**0.62**	0.02	-0.03	0.09	0.12	-0.12	0.00
	Technology Significance	V4	-0.06	0.04	**0.49**	-0.10	0.11	0.08	0.23
	Safety	V5	0.06	-0.06	-0.05	0.26	0.23	**0.43**	0.07
	Environmental Impact	V6	0.03	-0.01	0.10	0.10	0.14	**0.52**	-0.02
	Technology Production	V8	**0.62**	-0.06	-0.09	0.10	-0.11	0.18	0.00
Production	Tooling Cost	V9	**0.59**	0.15	-0.07	-0.16	-0.01	0.19	-0.01
	Cost of Production	V10	**0.57**	0.02	0.12	0.05	-0.18	0.28	-0.05
	Need	V11	-0.09	0.19	**0.44**	0.16	0.05	0.12	-0.03
	Potential Market	V12	0.04	-0.08	0.38	0.20	-0.10	0.16	-0.27
Market	Trend of Demand	V13	0.09	0.07	**0.56**	-0.03	0.03	0.00	-0.20
Demand	Durations of Demand	V14	-0.06	-0.06	**0.51**	0.15	0.05	-0.06	-0.08
	Demand Predictability	V15	-0.01	0.03	0.21	0.07	0.29	0.27	-0.02
	Product Line Potential	V16	-0.06	-0.02	0.29	0.03	0.16	0.28	0.12
	Societal Benefits	V18	-0.04	-0.05	0.21	0.01	**0.61**	0.11	0.01
	Compatibility	V19	-0.13	0.26	0.05	**0.43**	0.23	0.08	-0.08
	Learning	V20	0.17	-0.04	-0.10	**0.48**	0.13	0.13	-0.20
Market	Visibility	V21	-0.06	0.04	0.11	**0.56**	0.08	0.10	-0.05
Acceptability	Appearance	V22	-0.04	0.17	0.04	**0.50**	-0.09	0.10	0.08
	Function	V23	0.08	0.05	0.10	**0.43**	-0.10	-0.03	**0.45**
	Durability	V24	0.10	-0.08	0.08	**0.47**	-0.11	0.31	0.18
	Service	V25	0.17	0.04	0.00	0.28	0.05	0.29	-0.03
Market	Price	V26	0.31	0.21	0.17	-0.13	0.02	0.00	0.00
Competition	Existing Competition	V27	-0.03	0.13	-0.08	0.02	0.02	0.13	**0.52**
	New Competition	V28	-0.02	-0.06	-0.12	-0.03	0.11	-0.06	**0.45**
Market	Marketing Research	V29	0.28	0.09	0.06	0.10	0.15	0.09	-0.09
Effort	Promotion Cost	V30	0.05	0.29	0.04	0.03	0.26	-0.07	0.05
	Distribution	V31	0.14	0.07	0.09	-0.06	0.30	0.11	0.06
	Legality	V32	0.14	-0.16	-0.19	0.10	**0.44**	0.28	0.13
	Development Risks	V33	**0.49**	0.21	-0.06	-0.01	0.20	-0.03	0.00
	Dependence	V34	0.10	-0.06	0.29	0.14	0.00	0.01	-0.11
Risk	Protection	V35	0.07	0.00	**0.47**	-0.18	0.01	0.01	0.34
	Investment Costs	V37	**0.55**	0.33	0.03	-0.12	0.12	0.11	-0.02
	Potential Sales	V38	0.01	**0.59**	0.17	0.20	-0.08	-0.02	-0.01
	Payback Period	V39	0.18	**0.78**	-0.08	-0.02	0.00	-0.03	-0.01
	Profitability	V40	0.03	**0.83**	-0.01	0.11	-0.08	-0.05	0.09

Next I examine the second empirical factor (F2). In this case all three of the salient factor loadings are aligned with one of the hypothesized factors. Potential Sales (0.59), Payback Period (0.78) and Profitability (0.83) are all aligned with the hypothesized factor Risk.

Moving to the third empirical factor (F3) of the oblique rotated factors indicates more division of the salient loadings across the hypothesized factors. First, Need (0.44), Trend of

Demand (0.56) and Durations of Demand (0.51) all have salient loadings on Factor 3 and are aligned with the hypothesized factor Market Demand. Second, Technology Significance (0.49) also has a salient loading on Factor 3, but is aligned with the hypothesized factor Technical. Finally, Protection (0.47) has a salient loading on Factor 3 as well, but is aligned with the hypothesized factor Risk.

The fourth empirical factor (F4), similar to Factor 2, has all of its salient loadings on one of the hypothesized factors. Compatibility (0.43), Learning (0.48), Visibility (0.56), Appearance (0.50), Function (0.43) and Durability (0.47) all have salient loadings on Factor 4 and are aligned with the hypothesized factor Market Acceptability.

The fifth empirical factor (F5) has its two salient loadings divided across two of the hypothesized factors. First, Societal Benefits (0.61) has a salient loading on Factor 5 and is aligned with the hypothesized factor Market Acceptability. However, Legality (0.44) has a salient loading on Factor 5, but is aligned with the hypothesized factor Risk.

Next, the sixth empirical factor (F6) has both of its salient loadings on the same hypothesized factor. Both Safety (0.43) and Environmental Impact (0.52) have salient loadings on Factor 6 and are aligned with the hypothesized factor Technical.

Finally, the seventh empirical factor (F7) has two of its salient loadings on one hypothesized factor and a third salient loading on a different hypothesized factor. First, Existing Competition (0.52) and New Competition (0.45) have salient loadings on Factor 7 and are both aligned with the hypothesized factor Market Competition. Second, Function (0.45) also has a salient loading on Factor 7, but is aligned with Market Acceptability. It should be noted that Function has a split loading, as it has a salient loading on both Factor 4 (0.43) and Factor 7 (0.45).

Overall, the oblique rotation with seven factors extracted, provides only three factors with salient loadings that are not divided across the hypothesized factors. These factors are Factors 2, 4 and 6. However, the fourth empirical factor contains a split loading with the seventh empirical factor leaving only two factors with clear interpretability aligned with the post 1989 theoretical structure presented in Table 4-2. This assessment indicates that the factor loadings of the oblique rotation for seven factors provide sufficient evidence to reject hypothesis one. The factor analysis indicates that the seven factors indicated as the underlying theoretical structure and the 37 ranked variables used by the CIC in post 1989 rankings are insufficiently aligned to warrant acceptance of the hypothesis. With regard to simple structure, the oblique seven factor loadings have a hyperplane count of 176 out of 259 or 68.0%.

***H_2*: The pre 1989 empirical factor structure is aligned with the pre 1989 theoretical factor structure presented in Table 4-3. In other words, the empirical factor loadings on the 33 variables indicate alignment with five distinct factors, which are: Societal, Business Risk, Demand Analysis, Market Acceptance and Competitive.*

Table 6-2 Factor Loadings for 5 Factors (Pre 1989 Theoretical Structure / Pre 1989 Data Empirical Results)

Pre 1989 Theoretical Structure			Oblique Rotated Factor Pattern				
Factor	**Variable**		**F1**	**F2**	**F3**	**F4**	**F5**
	Safety	V5	<u>0.59</u>	0.39	-0.05	-0.10	-0.08
Societal	Environmental Impact	V6	<u>0.56</u>	0.32	-0.21	0.06	0.09
	Societal Benefits	V18	<u>0.58</u>	0.21	0.02	-0.11	0.08
	Legality	V32	0.21	<u>0.48</u>	-0.06	-0.05	0.01
	Functional Feasibility	V2	0.12	<u>0.68</u>	0.20	-0.14	-0.05
	Research & Development	V3	-0.05	<u>0.84</u>	0.06	-0.03	0.10
	Production Feasibility	V7	0.07	<u>0.65</u>	0.04	-0.01	0.14
Business	Stage of Development	V17	0.28	<u>0.47</u>	0.14	0.15	-0.09
Risk	Marketing Research	V29	-0.04	0.23	0.10	0.01	<u>0.53</u>
	Investment Costs	V36	-0.19	0.37	<u>0.59</u>	0.05	0.11
	Payback Period	V39	0.00	0.26	<u>0.60</u>	0.01	0.08
	Profitability	V40	-0.01	0.09	<u>0.73</u>	0.10	0.00
	Potential Market	V12	<u>0.74</u>	0.06	-0.06	0.05	0.08
	Trend of Demand	V13	<u>0.65</u>	-0.14	0.24	-0.10	0.16
Demand	Demand Life Cycle	V14	0.33	-0.20	0.33	-0.11	0.26
Analysis	Demand Predictability	V15	<u>0.43</u>	0.00	-0.05	0.16	0.27
	Product Line Potential	V16	<u>0.75</u>	-0.04	0.06	0.07	-0.08
	Potential Sales	V38	<u>0.48</u>	-0.06	<u>0.52</u>	0.14	-0.10
	Need	V11	<u>0.66</u>	-0.08	0.16	-0.02	0.15
	Compatibility	V19	<u>0.61</u>	0.24	0.10	-0.07	0.00
	Learning	V20	<u>0.46</u>	<u>0.46</u>	-0.04	-0.08	0.07
Market	Visibility	V21	<u>0.43</u>	0.16	0.23	0.10	0.09
Acceptance	Appearnance	V22	<u>0.56</u>	0.24	-0.04	0.10	0.00
	Promotion	V30	0.34	0.03	0.24	0.12	0.37
	Distribution	V31	0.29	0.14	0.04	0.11	<u>0.48</u>
	Dependence	V34	<u>0.56</u>	0.15	-0.02	0.02	-0.09
	Function	V23	0.38	<u>0.40</u>	0.18	0.12	-0.11
	Durability	V24	<u>0.46</u>	<u>0.41</u>	-0.10	0.18	-0.06
	Service	V25	-0.05	<u>0.56</u>	0.07	0.10	0.09
Competitive	Price	V26	0.35	0.24	0.15	0.08	0.01
	Existing Competition	V27	-0.12	0.02	0.26	<u>0.63</u>	0.08
	New Competition	V28	0.20	0.03	-0.04	<u>0.51</u>	0.09
	Protection	V35	0.09	-0.13	0.29	0.37	-0.07

To test the second hypothesis I examine the factor loadings of the oblique ration for five factors (Table 6-2). In this case the empirical factor loadings for the first empirical factor (F1) are divided across four of the five hypothesized factors in the theoretical structure. First, Safety (0.59), Environmental Impact (0.56) and Societal Benefits (0.58) all have salient loadings on Factor 1 and are aligned with the hypothesized factor Societal. Second, Potential Market (0.74), Trend of Demand (0.65), Demand Predictability (0.43), Product Line Potential (0.75) and Potential Sales (0.48) all have salient loadings on Factor 1, but are aligned with the hypothesized factor Demand Analysis. Third, Need (0.66), Compatibility (0.61), Learning (0.46), Visibility (0.43), Appearance (0.56) and Dependence (0.56) all have salient loadings on Factor 1, but are aligned with the hypothesized factor Market Acceptance. Finally, Durability (0.46) has a salient loading on Factor 1, but is aligned with the hypothesized factor Competitive.

Although the remainder of the factor loadings are presented in Table 6-2, the loadings on the first empirical factor provide considerable evidence to reject the second hypothesis. This first empirical factor possesses fifteen of a possible thirty-three salient loadings and they are distributed across four of the five factors from the pre 1989 theoretical structure. In fact, Table 6-2 indicates that only one of the empirical factors (F4) is aligned with a single theoretical factor. This assessment indicates that the factor loadings of the oblique rotation for five factors provide sufficient evidence to reject hypothesis two. The factor analysis indicates that the five factors indicated as the underlying theoretical structure and the 33 ranked variables used by the CIC in pre 1989 rankings are insufficiently aligned to warrant acceptance of the hypothesis. With regard to simple structure, the oblique five factor loadings have a hyperplane count of 96 out of 165 or 58.2%.

***H_3*: The post 1989 empirical factor structure is aligned with the pre 1989 theoretical factor structure presented in Table 4-3. In other words, the empirical factor loadings on the 37 variables indicate alignment with five distinct factors, which are: Societal, Business Risk, Demand Analysis, Market Acceptance and Competitive.*

Table 6-3 Factor Loadings for 7 Factors (Pre 1989 Theoretical Structure / Post 1989 Data Empirical Results)

Pre 1989 Theoretical Structure			Oblique Rotated Factor Pattern						
Factor	**Variable**		**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
	Safety	V5	0.06	-0.06	-0.05	0.26	0.23	**0.43**	0.07
Societal	Environmental Impact	V6	0.03	-0.01	0.10	0.10	0.14	**0.52**	-0.02
	Societal Benefits	V18	-0.04	-0.05	0.21	0.01	**0.61**	0.11	0.01
	Legality	V32	0.14	-0.16	-0.19	0.10	**0.44**	0.28	0.13
	Functional Performance	V2	0.38	0.05	-0.03	0.36	0.14	-0.22	0.10
	Research & Development	V3	**0.62**	0.02	-0.03	0.09	0.12	-0.12	0.00
	Technology Production	V8	**0.62**	-0.06	-0.09	0.10	-0.11	0.18	0.00
Business	Tooling Cost	V9	**0.59**	0.15	-0.07	-0.16	-0.01	0.19	-0.01
Risk	Cost of Production	V10	**0.57**	0.02	0.12	0.05	-0.18	0.28	-0.05
	Marketing Research	V29	0.28	0.09	0.06	0.10	0.15	0.09	-0.09
	Payback Period	V39	0.18	**0.78**	-0.08	-0.02	0.00	-0.03	-0.01
	Profitability	V40	0.03	**0.83**	-0.01	0.11	-0.08	-0.05	0.09
	Potential Market	V12	0.04	-0.08	0.38	0.20	-0.10	0.16	-0.27
	Trend of Demand	V13	0.09	0.07	**0.56**	-0.03	0.03	0.00	-0.20
Demand	Durations of Demand	V14	-0.06	-0.06	**0.51**	0.15	0.05	-0.06	-0.08
Analysis	Demand Predictability	V15	-0.01	0.03	0.21	0.07	0.29	0.27	-0.02
	Product Line Potential	V16	-0.06	-0.02	0.29	0.03	0.16	0.28	0.12
	Potential Sales	V38	0.01	**0.59**	0.17	0.20	-0.08	-0.02	-0.01
	Need	V11	-0.09	0.19	**0.44**	0.16	0.05	0.12	-0.03
	Compatibility	V19	-0.13	0.26	0.05	**0.43**	0.23	0.08	-0.08
	Learning	V20	0.17	-0.04	-0.10	**0.48**	0.13	0.13	-0.20
Market	Visibility	V21	-0.06	0.04	0.11	**0.56**	0.08	0.10	-0.05
Acceptance	Appearance	V22	-0.04	0.17	0.04	**0.50**	-0.09	0.10	0.08
	Promotion Cost	V30	0.05	0.29	0.04	0.03	0.26	-0.07	0.05
	Distribution	V31	0.14	0.07	0.09	-0.06	0.30	0.11	0.06
	Dependence	V34	0.10	-0.06	0.29	0.14	0.00	0.01	-0.11
	Function	V23	0.08	0.05	0.10	**0.43**	-0.10	-0.03	**0.45**
	Durability	V24	0.10	-0.08	0.08	**0.47**	-0.11	0.31	0.18
	Service	V25	0.17	0.04	0.00	0.28	0.05	0.29	-0.03
Competitive	Price	V26	0.31	0.21	0.17	-0.13	0.02	0.00	0.00
	Existing Competition	V27	-0.03	0.13	-0.08	0.02	0.02	0.13	**0.52**
	New Competition	V28	-0.02	-0.06	-0.12	-0.03	0.11	-0.06	**0.45**
	Protection	V35	0.07	0.00	**0.47**	-0.18	0.01	0.01	0.34

To test the third hypothesis I examine the factor loadings of the post 1989 data using an oblique ration for seven factors (Table 6-3). These factor loadings are identical to those found in Table 6-1; however, in Table 6-3 the empirical results are reorganized to determine their alignment with the pre 1989 theoretical structure. There are two obvious concerns with this process. First, I chose to use the factor loadings from the seven factor solution rather than use a five factor solution. The main reason for this decision was that the seven factor solution was found to be superior to the five factor solution by both the scree and eigenvalue tests.

Therefore, the extraction of seven factors was the most appropriate number of factors to extract form the post 1989 dataset. The second obvious concern involves the differences among variables used in the pre and post alteration periods. Thankfully, the CIC maintained detailed records of their variable alterations over the years (Appendix D) and this documentation was used to map the variables from the post 1989 period onto the pre 1989 theoretical structure.

Examining the post 1989 empirical factor loadings for the first empirical factor (F1) I find the loadings focused on a single hypothesized factor from the pre 1989 theoretical structure. Research and Development (0.62), Technology Production (0.62), Tooling Cost (0.59) and Cost of Production (0.57) all have salient loadings on Factor 1 and are aligned with the hypothesized factor Business Risk.

Next I examine the second empirical factor (F2). In this case three factor loadings are divided across two hypothesized factors. First, Payback Period (0.78) and Profitability (0.83) are both aligned with the hypothesized factor Business Risk. However, Potential Sales (0.59) also has a salient loading on Factor 2, but is aligned with the hypothesized factor Demand Analysis.

The third empirical factor (F3) of the oblique rotated factors is divided across three of the hypothesized factors. First, Trend of Demand (0.56) and Durations of Demand (0.51) both have salient loadings on Factor 3 and are aligned with the hypothesized factor Demand Analysis. Second, Need (0.44) also has a salient loading on Factor 3, but is aligned with the hypothesized factor Market Acceptance. Finally, Protection (0.47) has a salient loading on Factor 3 as well, but is aligned with the hypothesized factor Competitive.

The fourth empirical factor (F4) has six salient loadings divided across two of the hypothesized factors. First, Compatibility (0.43), Learning (0.48), Visibility (0.56) and Appearance (0.50) all have salient loadings on Factor 4 and are aligned with the hypothesized factor Market Acceptance. Second, Function (0.43) and Durability (0.47) both have salient loadings on Factor 4 but are aligned with the hypothesized factor Competitive.

The fifth empirical factor (F5) has both of its salient loadings aligned with one of the hypothesized factors. Both Societal Benefits (0.61) and Legality (0.44) have salient loadings on Factor 5 and are aligned with the hypothesized factor Societal.

Next, the sixth empirical factor (F6) also has both of its salient loadings on the same hypothesized factor. Both Safety (0.43) and Environmental Impact (0.52) have salient loadings on Factor 6 and are aligned with the hypothesized factor Societal.

Finally, the seventh empirical factor (F7) has all three of its salient loadings on one hypothesized factor. Function (0.45), Existing Competition (0.52) and New Competition (0.45) all have salient loadings on Factor 7 and all aligned with the hypothesized factor Competitive. Since these loadings are the same as those found in the analysis of hypothesis one, Function still has a split loading with a salient loading on both Factor 4 (0.43) and Factor 7 (0.45).

When the loadings from the oblique rotation with seven factors extracted are assessed using the pre 1989 theoretical structure only four of the seven empirical factors are not divided across the hypothesized factors. These factors are Factors 1, 5, 6 and 7. However, as was the case in hypothesis one, the fourth empirical factor contains a split loading with the seventh empirical factor leaving only three factors with clear interpretability aligned with the pre 1989 theoretical structure presented in Table 4-3. Although the results indicate a slight improvement over the assessment of hypothesis one, this assessment indicates that the factor loadings of the oblique rotation for seven factors provide sufficient evidence to reject hypothesis three. The factor analysis indicates that the five factors indicated as the underlying pre 1989 theoretical structure and the 37 ranked variables used by the CIC in post 1989 rankings are insufficiently aligned to warrant acceptance of the third hypothesis.

***H_4*: *The post 1989 empirical factor structure is aligned with a combination of the pre 1989 and post 1989 theoretical factor structures presented in Table 4-3 and Table 4-2. In other words, the empirical factor loadings on the 37 variables indicate alignment with a mixture of the twelve possible factors which are: Technical, Production, Market Demand, Market Acceptability, Market Competition, Market Effort and Risk from Table 4-2 as well as Societal, Business Risk, Demand Analysis, Market Acceptance and Competitive from Table 4-3.*

To test the fourth hypothesis I examine Table 6-1 and Table 6-3 simultaneously (Table 6-1 is placed on top of Table 6-3 in Appendix E to improve the visual interpretation of the results). Both these tables contain the factor loadings for the post 1989 data using an oblique ration for seven factors. These two sets of factor loadings are identical with two exceptions. First, Table 6-1 contains 37 variables while Table 6-3 contains 33 variables. Second, some equivalent variables are named according to the names given them by the CIC during each of the two respective periods (A graphical representation of the mapping of variables between the post 1989 theoretical structure and the pre 1989 theoretical structure is contained in Appendix D).

Since the seven factor post 1989 factor loading details have already been provided in the discussion of hypotheses one and three, I refrain from restating these results. However, an examination of Tables 6-1 and 6-3 demonstrates that six of the seven empirical factors align with one of the twelve theoretical factors from the two periods. The first empirical factor (F1) from the seven factor extraction of post 1989 data contains four salient loadings on the hypothesized factor Business Risk from the pre 1989 theoretical structure (Table 6-3). The second empirical factor (F2) contains three salient loadings aligned with the hypothesized factor Market Demand from the post 1989 theoretical structure (Table 6-1). However, examination of either table indicates that the salient loadings of the third empirical factor (F3) align with three hypothesized factors. Next, the fourth empirical factor (F4) from the seven factor extraction of post 1989 data contains six salient loadings on the hypothesized factor Market Acceptability from the post 1989 theoretical structure (Table 6-1). The fifth empirical factor (F5) contains two salient loadings on the hypothesized factor Societal from the pre 1989 theoretical structure (Table 6-3). Interestingly, the sixth empirical factor (F6) contains two salient loadings that align with hypothesized factors from each of the theoretical structures: Societal from the pre 1989 theoretical structure (Table 6-3) and Technical from the post 1989 theoretical structure (Table 6-1). Finally, the seventh empirical factor (F7) contains three salient loadings on the hypothesized factor Competitive from the pre 1989 theoretical structure (Table 6-3).

Overall, when using the post 1989 dataset, the alignment present between the seven empirically extracted factors and the twelve hypothesized factors provided by the combination of both pre 1989 and post 1989 theoretical structures is extremely promising. In fact, six of the seven empirically derived factors are aligned with one, and only one, of the factors from the theoretical structures. Therefore, the empirical results provide sufficient evidence to accept the fourth hypothesis, which indicates that the pre 1989 theoretical structure remained partially resident in the minds of the CIC analysts during the post 1989 period.

***H_5**: The IAP improved its ability to predict success following the alteration of the theoretical variable structure in 1989. Therefore, the empirical regression results using the post 1989 data should demonstrate an improved capability to predict success when compared to the empirical regression results using the pre 1989 data.*

To test the fifth hypothesis I regress the success variable on the ranked variables for each data set. First, I perform the logistic regression for the pre 1989 data set and then for the post 1989 data set. The goal is to determine if the CIC demonstrates an empirical improvement in their ability to predict success following the alteration of the theoretical structure in 1989. For each period, backward, forward and stepwise variable elimination techniques were attempted and the method providing the best result was selected.

The model for the pre 1989 period predicted 369 out of the 583 possible inventions correctly (63.3 percent). Columns (2) and (3) of Table 6-4 indicate that the model predicted

38 of the 59 possible successes correctly, (64.4 percent), with a probability of a Type I error of 0.356. The model predicted 331 of a possible 524 failures correctly (63.2 percent), with a probability of a Type II error of 0.368. The pre 1989 model has a Pseudo R^2 of 0.10.

Table 6-4 CIC Predictive Ability Pre 1989 vs. Post 1989

	Pre 1989 Period		Post 1989 Period	
	Number	Percent	Number	Percent
(1)	(2)	(3)	(4)	(5)
Correctly Predicts Success	38	0.64	49	0.77
Type I Error Actual success but model predicts failure	21	0.36	15	0.23
Correctly Predicts Failure	331	0.63	392	0.76
Type II Error Actual failure but model predicts success	193	0.37	125	0.24
Overall Predictive Ability		0.63		0.76

The model for the post 1989 period predicted 441 out of the 581 possible inventions correctly (75.9 percent). Columns (4) and (5) of Table 6-4 indicate that the model predicted 49 of the 64 possible successes correctly, (76.6 percent), with a probability of a Type I error of 0.234. The model predicted 392 of a possible 517 failures correctly (75.8 percent), with a probability of a Type II error of 0.242. The post 1989 model has a Pseudo R^2 of 0.20.

Overall, the post 1989 period demonstrates superiority to the pre 1989 period regardless of the measure used. The capability of the CIC to predict success improves 12.2 percent in the post 1989 period. Similarly, the capability of the CIC to predict failure improves 12.6 percent in the post 1989 period. Overall, the post 1989 period improves the predictive ability of the CIC 12.6 percent over the pre 1989 period. As a result, I accept hypothesis five and conclude that the IAP improved its ability to predict success following the alteration of the theoretical variable structure in 1989.

6.4 Concluding Remarks

Although three of my five hypotheses were rejected, the resulting findings provide useful insight into the innovation evaluation process. I discovered that factor analysis of the 37 variables ranked by the IAP analysts during the post 1989 period provides empirical results that can be interpreted by the theoretical structures. The key is the plurality of the word 'structures'. Neither the empirical results from the pre 1989 period nor the post 1989 period where found to be highly interpretable using the theoretical structures of their respective periods. Rather, it appears, the pre 1989 theoretical structure remained partially resident in the minds of the analysts even though the structure had been formerly altered. Intuitively, this empirical finding offers a logical and realistic result. The pre 1989 theoretical structure would have been entrenched in the minds of the CIC analysts after years of continued use with minimal alteration. Alteration of the theoretical structure in 1989 is simply that – theory. Therefore an evolutionary process must occur whereby the altered, post 1989, theoretical structure adds to the analyst process and begins to alter the evaluation process. However, it does not fully replace the former theoretical structure, but rather cohabitates with the former structure in the minds of analysts. Although more research may be required, these results indicate that the analysts may be mentally calling upon portions of both theoretical structures in their ranking procedures.

Another interesting finding was the regression analysis that compared the predictive ability in the pre 1989 period to the predictive ability in the post 1989 period. I found a noticeable improvement in the predictive capabilities of the CIC following the 1989 alteration of the theoretical structure. While there could be many reasons contributing to these improved capabilities, the results indicate that the alterations performed on the theoretical structure were beneficial to the CIC evaluation process. Since the factor analysis indicates that the analysts are mentally drawing upon theoretical structures from both pre and post alteration periods, it may offer insight into further improvement in predictive ability of the CIC evaluation process.

Beyond these two key results, the findings indicate the benefits of continued model assessment through factor analysis and predictive capability assessment. In order to maintain credible models IAP's and others utilizing predictive models should make a concerted effort to collect pre and post evaluation data to continuously assess their models. Without such efforts evaluation models can become inaccurate and invalid. In other words, their predictive ability and their alignment with their underlying structure could begin to diverge from what could be optimally attained through continued statistical analysis and adjustment. Model maintenance must become an ongoing process to maintain the validity of the evaluation model and process.

Further factor analysis could be performed with this data. First, differing numbers of factors could be extracted. For example, although it was obviously a different dataset, Udell (1989) mentioned that the first formal factor analysis performed on the data found eight factors to be the correct number to extract. It would thus be logical to assess the extraction of eight factors with the same dataset to determine if the factors have remained somewhat constant. Another avenue is to examine the same number of factors (7 and 10) with different factor extraction methods and rotations. While I anticipate the results of such analysis to be marginal, I believe that incremental improvements may be discovered. Beyond different numbers of factors and alterations in methods and extractions, it is recommended that formal confirmatory factor analysis be performed on the datasets within this thesis. Such formal confirmatory factor analysis should include overall fit statistics to provide additional insight beyond the examination of the factor loadings presented here.

Another avenue for further research, which would require minimal effort, would be to return to the previous factor analysis work (Åstebro, Michela and Zhang 2001) performed on the same dataset. These authors extracted factors from the dataset only to dismiss them due to their lack of interpretability. By returning to the factor loadings of these results and considering both the pre and post 1989 theoretical structures, these results may be considered more promising as a result of improved interpretability.

Chapter 7
Commercial Success: Factor Analysis vs. Raw Variables

7.1 Method

In Åstebro and Sampson (1999) [Chapter 5] we performed logistic regression to determine which of the CIC's 37 ranked criteria were significant in the determination of success and how well the resulting regression model could predict success. We also examined a second regression model based on these underlying ranked criteria and its ability to predict the overall rating provided by the CIC. In this chapter, I chose to apply the factor analysis method to the underlying ranked criteria data, prior to performing logistic regression, to develop additional models. I then compare the predictive ability of the new models with the predictive ability and interpretability of the logistic regression models developed in Åstebro and Sampson (1999). The intent is to determine the model with the best predictive ability and the most interpretable results irrespective of statistical method.

Although Åstebro and Sampson (1999) chose not to use factor analysis due to the problems associated with difficult to interpret factors in previous studies (Balachandra and Friar 1997), it was obvious to the authors that the high correlations among the independent variables provided evidence to suggest that logistic regression and variable elimination could be affected. Therefore, it was decided to perform factor analysis to eliminate these correlations and perform logistic regression using the same dependent variables but the extracted factors as the independent variables.

To perform regression analysis I use the factors extracted from the post 1989 period dataset in Chapter 2. These datasets represent four different sets of independent variables. First, seven factors were extracted using both orthogonal and oblique rotation methods. Second ten factors were extracted using both orthogonal and oblique rotations. In this chapter, I focus on the seven factor extraction as the scree and eigenvalues tests indicate seven to be the most appropriate number of factors to extract from the post 1989 dataset. In addition, as indicated in Chapter 5, the empirical results of the seven factor solution are interpretable when compared against the theoretical structure. Details of the regression analysis performed with ten factors are included for comparison purposes in Appendix F.

While the use of factors extracted through orthogonal rotation for further data analysis such as regression analysis is commonplace and indeed standard practice, the use of factors extracted through an oblique rotation method can be questioned. The main reason for not using factors from an oblique rotation for regression analysis is the allowance, by such rotations, of correlations across factors. However, I choose to proceed with the oblique factors due to the relatively low correlations found within the correlation matrix of the oblique factors (Table 7-1). In addition, the original variables were not constructed as completely independent. Indeed, previous research (Åstebro, Michela and Zhang 2001) using the same dataset indicated that some criteria in different categories seemed likely to covary substantially. I agree with this statement as the original variables and the resulting factors represent real world phenomenon, which Rummel (1970) indicates should not be treated as unrelated clusters. Overall, it was considered unrealistic to assume that all of the factors should be uncorrelated within this particular dataset.

Table 7-1 Oblique (Promax Rotation) Factor Correlations 7 Factors

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Factor1	1.00	0.35	0.04	<u>**0.47**</u>	0.23	-0.06	0.23
Factor2	0.35	1.00	<u>**0.45**</u>	0.25	0.13	0.21	0.21
Factor3	0.04	<u>**0.45**</u>	1.00	0.29	0.03	0.21	0.10
Factor4	<u>**0.47**</u>	0.25	0.29	1.00	0.34	-0.06	0.11
Factor5	0.23	0.13	0.03	0.34	1.00	-0.21	0.06
Factor6	-0.06	0.21	0.21	-0.06	-0.21	1.00	-0.17
Factor7	0.23	0.21	0.10	0.11	0.06	-0.17	1.00

The purpose of this chapter is to compare the predictive accuracy of a model estimated by performing factor analysis on the independent variables prior to using logistic regression with the predictive accuracy of a model estimated using logistic regression with the 37 early stage characteristics. In Åstebro and Sampson (1999) [Chapter 1] one of the estimated models had project success as the dependent variable and the 37 early stage characteristics as the independent variables. In this chapter, I use seven factors extracted from the 37 early stage characteristics and regress the success variable on them. (Details of the factor extraction method and results are contained in Chapter 6.) The resulting predictive capabilities of the two models are compared twice. First, the two models are assessed using the full post 1989 dataset to estimate and test the models. Then the same models are assessed with an out of sample prediction test.

For comparison of predictive accuracy of the respective models three criteria were used for these models with success as the dependent variable. First, overall predictive accuracy was examined. This number was calculated by taking the number of correct predictions of the model and dividing it by the total number of observations examined. Second, the Type I error is calculated as the probability that the model predicts a failure given that the invention is a success. Finally, I compute the Type II error as the probability that the model predicts a success given that the invention fails to reach the market.

All the models use binary logistic regression because of the format of the dependent variable. When we previously regressed the dependent variable, project success, on the original early stage characteristics we attempted analysis with three different link functions: logit, normit (also called probit), and gompit (also called complementary log-log). The results of these attempts indicated that all three functions generated qualitatively similar results. As a result, I again select the logit model for this analysis.

Since I use the factors from the oblique rotation for reasons previously discussed, both the models based on the extracted factors as well as those based on the early stage characteristics possess collinearity among their independent variables. As a result, I utilize variable elimination techniques in the estimation of all models. For each model, only factors or variables that were found to be significant at the 0.05 level were accepted into the final prediction models. In assessing each model I attempted three variable elimination methods. I used backward, forward and stepwise elimination. However, I found there to be no significant differences between these methods. In fact, in almost all cases the results are identical. Therefore, I only report the forward elimination results.

Initially, to assess and compare the models based on the initial characteristics ranked by the IAP with the models based on the extracted factors, four additional models where determined. The first two models used the factors from the factor analysis where seven factors were extracted as the independent variables using two rotation methods. While the last two models used the factors determined through factor analysis where ten factors were extracted using two rotation methods. Furthermore, each of the models were estimated and tested using both the full sample and a split sample test. For the split samples, data from the years 1989 through 1992 are used to estimate the parameters and then the models are tested on data from the year 1993. However, the seven factor model estimated using orthogonal factors as well as the two ten factor extractions are not reported in the body of the thesis. These additional models were estimated for comparison purposes and their results offer similar results (For details see Appendix F). Therefore, only the results of the model based on the factors from the oblique seven factor extraction are reported in this chapter. The predictive ability of this model is compared to the model estimated by regressing project success on the early stage characteristics.

7.2 Results

One of the models determined in Åstebro and Sampson (1999) used the 37 early stage characteristics ranked by the IAP as the independent variables and project success as the dependent variable. I now compare the predictive ability of this model with a model estimated using the factors extracted from these early stage characteristics as the independent variables.

Table 7-2 Ability to Predict Success: Early Stage Characteristics vs. 7 Oblique Factors

	Early Stage Characteristics		7 Oblique Factors	
	Number	Percent	Number	Percent
(1)	(2)	(3)	(4)	(5)
Correctly Predicts Success	49	0.77	48	0.75
Type I Error Actual success but model predicts failure	15	0.23	16	0.25
Correctly Predicts Failure	392	0.76	383	0.74
Type II Error Actual failure but model predicts success	125	0.24	134	0.26
Overall Predictive Ability		0.76		0.74

Following the variable elimination process, the model that uses seven oblique (promax) factors as the independent variables and project success as the dependent variable consists of four independent variables. The four of the seven factors that are retained are: Factor 1, Factor 2, Factor 3, and Factor 7. This model correctly predicted 431 out of the 581 possible inventions (74.2 percent) and had a Pseudo R^2 of 0.19. Columns 4 and 5 of Table 7-2 indicate that the model predicted 48 of a possible 64 successes correctly (75.0 percent) with a Type I error of 0.250. Table 7-2 also indicates that the model predicted 383 of the 517 possible failures (74.1 percent) with a Type II error of 0.259.

Columns 4 and 5 of Table 7-3 indicate that in a time split sample test this model predicted 131 out of the 172 possible inventions in the test sample correctly (76.2 percent). The model predicted 13 of a possible 18 successes correctly (72.2 percent), with a probability of a Type I error of 0.278. The model predicted 118 of the 154 possible failures correctly, (76.6 percent), with a probability of a Type II error of 0.234.

Table 7-3 Ability to Predict Success: Early Stage Characteristics vs. 7 Oblique Factors (Split Sample Test)

	Early Stage Characteristics		7 Oblique Factors	
	Number	Percent	Number	Percent
(1)	(2)	(3)	(4)	(5)
Correctly Predicts Success	13	0.72	13	0.72
Type I Error Actual success but model predicts failure	5	0.28	5	0.28
Correctly Predicts Failure	118	0.77	118	0.77
Type II Error Actual failure but model predicts success	36	0.23	36	0.23
Overall Predictive Ability		0.76		0.76

At this point, the model using the IAP's 37 early stage characteristics as the independent variables can be compared against the model using 7 oblique factors as the independent variables. To compare the models I examined both the interpretability of the underlying factors and the predictive accuracy. Therefore, I will first examine the interpretability of the underlying factors. Second, I will compare the predictive accuracy when tested on the sample data and when tested in a time split test.

To compare the interpretability of the factors I return to the results of Chapter 6, which indicated that the seven oblique factors extracted from the post 1989 dataset had clear alignment with the combined post 1989 and pre 1989 theoretical factor structures. With the exception of one factor the mapping to the predefined variable structures is aligned. As a result, the interpretability of the factors when seven factors are used is considered high. Through the variable elimination process the final model estimated using the seven oblique factors is reduced to four factors. The remaining factors are Factor 1 (F1) which is aligned with Business Risk from the pre 1989 theoretical structure, Factor 2 (F2) which is aligned with Risk from the post 1989 theoretical structure, Factor 3 (F3) which is the only difficult to interpret factor but is most closely associated with Market Demand from the post 1989 theoretical structure, and, finally, Factor 7 (F7) which is aligned with Competitive from the pre 1989 theoretical structure. It should be noted that the second significant factor (F2), Risk from the post 1989 theoretical structure, has salient loadings on the three variables Potential Sales, Payback Period and Profitability. Therefore, I am more inclined to interpret this Factor as Return on Investment (ROI) rather than the more general Risk. With respect to overall interpretability of the model it is interesting, and somewhat unfortunate, that one of the four significant Factors at the 0.05 level is the only difficult to interpret extracted factor.

To compare the predictive accuracy of the model I now compare the model that uses the original early stage characteristics as the independent variables with the model that uses the extracted factors as the independent variables. The model that uses the 37 early stage characteristics as the independent variables is superior, in almost every instance, to the 7 oblique factor model when the model is estimated and tested using the full sample. However, the margin of victory is quite small in some cases. As indicated in Table 7-2, the model that uses the underlying characteristics predicts 441 out of 581 cases producing an overall percentage correct of 75.9 percent. This is superior to the factor model by 10 predictions or 1.7 percent overall. When success and failure are considered separately, this model predicted 49 out of 64 successes correctly (76.6 percent) and 392 out of 517 failures correctly (75.8 percent). On both counts this is superior to the factor model. With respect to predicting success correctly, the early characteristics model is superior by 1 prediction or 1.6 percent. When predicting failure, the early stage characteristics model is superior by 9 predictions or 1.7 percent. In addition, the Pseudo R^2 of the early stage characteristics model is 0.20, which is again slightly higher than the 0.19 Pseudo R^2 of the factor model.

While victory for the early stage characteristics model is, although marginal, clear and consistent with the full sample results, the findings are much different when the time split sample is examined. As Table 7-3 clearly indicates, the early stage characteristics model and the factor model are equal on all measures when the time split assessment is performed.

7.3 Concluding Remarks

I found that in making an overall assessment of an innovation at an early stage of development a model estimated using extracted factors as the independent variables was comparable in predictive ability with a model estimated using the original early stage characteristics. In fact, when a time split sample is performed, the predictive ability of the two models are identical.

This result indicates that either of the two models is suitable for assessing the success of inventions. Therefore, both models could be successfully used by IAP's. Model selection depends more on the goals of a particular IAP. For example, if the goal of a particular IAP were to expedite the evaluation process, due to increasingly tight budgets, they would likely select the model based on the early stage characteristics as opposed to the model using the extracted factors. Although both estimated models contain four independent variables following variable elimination, the early stage characteristics model only requires the ranking of four characteristics while the factor model requires the ranking of all thirty-seven characteristics. In other words, since the four factors that remain in the factor model depend on loadings from all early stage characteristics, it is necessary to rank every single one. It could be argued that some of the variables with low loadings could be eliminated. For example, one could argue that only fifteen variables have salient loadings on the four factors that remain in the model so ranking those would be sufficient. However, while this method would maintain the fifteen most important loadings for our four factors, it would alter our factors and thus its ability to predict success. It could further be argued that lower cut of points for loadings could be used. For example, all loadings less than 0.10 or 0.20 could be dropped leaving our factor model depending on more than 15 but less than 37 early stage characteristics. However, the key point remains that the factor model can never be effective ranking only four early stage characteristics. As a result, the model estimated using the early stage characteristics, as the independent variables, will remain more efficient from the perspective of required ranking effort while providing an equivalent ability to predict project success.

Given the probability of dependences between the early stage ranked characteristics, it is unknown whether ranking only four characteristics is sufficient. If an IAP analyst is asked to streamline the assessment process by focusing on the four variables that remain in the early stage characteristics model will the rankings of these variables change? In other words, does the ranking process of these four particular variables depend on the analyst thinking about and actively assigning rankings to some or all of the other variables? If so, then streamlining the process may have negative effects on the predictive ability of the model. However, the effects of such streamlining are unknown and further research would be required to assess the full effects of streamlining the process at IAP's.

Overall, regardless of the model chosen, the predictive accuracy of the resulting models remains higher than the ability of R&D managers to predict technical success of projects (Mansfield 1968), but lower that the project selection model developed by Cooper (1981). However, the predictive capability comparison to Cooper can be questioned, as our model was tested on screening criteria as they were used, whereas Cooper's analysis was on variables ex post associated with success.

Although models have been developed and assessed using logistic regression with the IAP ranked characteristics and now extracted factors, other suitable methods remain for the development of a predictive model. Indeed, this data set lends itself to model development using at least two other methods and possibly more. While other methods may be suitable, two that I feel offer particular merit are neural network analysis and rough sets analysis. During the course of this research, both of these methods were explored to varying degrees. First, neural network analysis was performed using the backpropagation algorithm with numerous parameters. While these initial neural network efforts resulted in comparable predictive capabilities to those found with the logistic regression and factor analysis methods described in this chapter, it might be useful to attempt more in-depth analysis. Second, a small amount of effort was expended using rough sets analysis with this dataset. Examination of the variables indicated the existence of patterns, which suggested that rough sets analysis might be appropriate for the dataset. However, analysis led to limited success and the rough sets analysis was subsequently dropped. Therefore, it is still possible that a more comprehensive effort with rough sets could lead to more promising results. Models developed using these additional methods could then be compared against the ones presented in this thesis for their predictive capabilities as well as their interpretability.

Chapter 8
Summary and Conclusions

The intent of this thesis was to provide insight into the evaluation of before market inventions, innovations and research and development projects. This insight was derived from the assessment of a common invention evaluation model used in many IAP's throughout North America for the past twenty years. Through the research process each step was carefully considered with respect to the value it would add to the body of work within this research area. First, the data collection was designed to counter many of the potential biases in the research. Second, the methods used were selected based on the previous work of others, the task at hand and the overall goal of the research. Finally, an effort was made to perform the analysis and report the results with a focus on adding both value and insight.

Chapters five through seven provide the core of the results and findings within this thesis. However, the fourth chapter is important because of the data collection process utilized to form the foundation of the work in the chapters that it precedes. Indeed, the validity of the findings presented within this thesis is based on a foundation of unbiased data collection. As a result, it is critical to briefly reexamine this process and why it is important to the evaluation of before market inventions, innovations and research and development projects. Unlike much of the research performed within this particular field of model development and assessment, this research used data collected on independent variables and dependent variables that were not collected simultaneously. Instead the data on the dependent variables was collected years after the data on the independent variables was scored. This single fact provides considerable credibility to the findings by reducing the obvious biases found in simultaneous data collection.

The goal of Chapter Five was to assess the predictive accuracy of the model used by the CIC and whether self-selection was occurring. This chapter provided evidence to suggest that the CIC model, which is similar to the invention evaluation model commonly used by IAP's across North America, was indeed accurate in predicting success and self-selection was unlikely to be responsible for this predictive capability.

The goal of Chapter Six was to assess the IAP model with respect to its underlying theoretical variable structure. In addition, this chapter examined the IAP's alteration of the theoretical structure in 1989 through empirical investigation of datasets from the pre and post alteration periods. The conclusion in Chapter Six was that the CIC analysts were mentally drawing upon a mixture of the pre and post theoretical structures during the post alteration period. Furthermore, it was found that the post alteration period model was superior in predicting project success. Combined these facts indicated that the alteration might have been successful in improving the accuracy of the evaluations at the CIC, but the factor analysis indicates that there could still be improvement with respect to the alignment of the underlying variable structure, which might in turn improve accuracy even more.

The goal of Chapter Seven was to compare the predictive accuracy of a model estimated using the early stage characteristics as the independent variables with a model using extracted factors as the independent variables. The conclusion of Chapter Seven was that both models performed almost equally well when predicting project success. Due to this finding, it was recommended that IAP's, and others interested in using such a model, would need to look at their specific requirements in determining which model was more appropriate.

Beyond the specific findings there are a number of issues that deserve discussion. First there is the reduction of governments funding for IAP programs and the resulting need for IAP's to alter their operations to reflect this new reality. The second issue is how IAP invention evaluation models can be expanded to new verticals and applications. Finally, there are a number of key elements that remain outside the scope of this thesis that need to be identified.

Ironically, although IAP's provide advice and recommendations to inventors and innovators, the evaluation process used at IAP's has undergone little if any true innovation over the past twenty-five years. Reliance on government funding over this period of time may be partially responsible for this lack of progress. Recently, much of this government funding, at least in Canada, has been substantially reduced. As a result, IAP's need to reexamine their internal processes and operations to develop increased efficiencies. The processes presented in this thesis provide a starting point for increased efficiencies within IAP's. Indeed, the statistical processes of logistic regression and factor analysis are obvious tools for building and assessing the evaluation models. However, the frequency of the assessments and the application of the empirical results have been less than optimal within IAP's.

One possible solution to the reduced funding and increased efficiency would be to utilize the Internet to aggregate both the ranking data and the outcome data across numerous, or all, IAP's. In other words, the IAP's could utilize the power of the Internet to collect and analyze data from multiple IAP's and then create a single continuously updated model based on empirically driven results. For even further efficiencies the actual inventor could rank inventions through an Internet interface. Alternatively, a franchise type system could be implemented where the IAP trains individuals or small businesses in dispersed geographical locations to offer the invention evaluation service and all interaction with the IAP could be done through the Internet. However, enabling either or both of these options would require considerable knowledge acquisition from the IAP analysts to derive the decision heuristics they utilize. These heuristics, if successfully extracted, could be used to develop a series of questions that underlie the ranking of each variable. The purpose of such questions would be to reduce the subjectivity of the ranked variables by focusing on structuring relatively objective underlying questions. Since one of the models developed in this thesis required the use of only four ranked variables and possessed a predictive accuracy of over seventy percent, questions could be developed that underlie only these four variables resulting in a model that could be both efficient and effective.

Although the current model utilized by the CIC and many IAP's is applied across all types of inventions and innovations and this horizontal application has its merits with regard to efficiency, different models for different verticals are becoming necessary. Both the CIC and McGrath and MacMillan (2000) demonstrate a demand for such vertical models. In the case of the CIC, they have adapted the PIES (Udell 1989) model to apply specifically to the software industry. In the case of McGrath and MacMillan (2000) they have developed a model that focuses on the specific nuances of technology businesses. In both situations, the parties have targeted specific verticals by moving beyond a more generic model. McGrath and MacMillan (2000) state that a number of venture capital organizations have used their model with success. Another possibility, with a certain degree of similarity to creating derivative models for application with specific industry verticals, would be to develop models that apply to features or incremental change in products.

While there are many possible applications of evaluation models, it is important to note some of the limitations of this thesis. First, the dependent variable termed "success" within this thesis, while easily defined and replicated, is more representative of commercialization than actual success. Reaching market with one unit is, indeed, a form of success, but can still lead to dismal failure in the marketplace. Realizing this, Åstebro and others are currently carrying out additional research to explore longer-term success by examining variables such as return on investment. A second important issue involves the people factor. In practice the CIC collects information on the actual inventor and utilizes that information in its decision heuristics. However, no information on the individual is considered within this thesis. Arguably, at the early stage of development, the individual or team is one of the more critical elements to assess and is paramount with most venture capitalists when considering early stage investments. Therefore, more research is necessary to determine the impact of the people factor in IAP assessments. Finally, the market assessment is another critical factor when examining early stage opportunities and, although this is considered in the models, it likely requires an extensive effort to provide a full assessment especially when analysts cannot be experts in all markets. My concern is that, in a percentage of cases, the market assessment is scored with insufficient knowledge. However, the CIC likely recognizes the importance of this component since it does offer in-depth market research as an additional service to clients.

Overall, this thesis provides methods that can be implemented by IAPs to continually assess and maintain their predictive models. These findings and methods could be applied to models for before market inventions, innovations and research and development projects. In order to maintain the quality of predictive models, IAP's, R&D departments and others need to assess the predictive ability and the theoretical structures of their models on a regular basis. Therefore, post evaluation data collection, regression analysis and factor analysis need to become part of the ongoing processes of those who use such models.

Appendix A
Variable Definitions

Technical Feasibility	Is the technical solution sound and complete?
Functional Performance	Does this innovation work better than the alternatives?
Research and Development	How great a burden is the remaining research and development required to bring the innovation to a marketable stage?
Technology Significance	How significant a contribution to technology or to its application is proposed?
Safety	Are potential dangers or undesirable side effects expected?
Environmental Impact	Will the innovation lead to pollution, litter, misuse of natural resources or the like?
Technology of Production	Are the technology and skills required to produce the invention available?
Tooling Cost	How great a burden is the cost of production tooling required to meet the expected demand?
Cost of Production	Does production at a reasonable cost level appear possible?
Need	Does the innovation solve a problem, fill a need or satisfy a want for the customer?
Potential Market	How large and how enduring is the total market for all products serving this function?
Trend of Demand	Will the demand for such an innovation be expected to rise, remain steady, or fall in the lifetime of this idea?
Duration of Demand	Is the demand for the innovation expected to be "long term"?
Demand Predictability	How closely will it be possible to predict sales?
Product Line Potential	Can the innovation lead to other profitable products or services?
Societal Benefits	Will the innovation be of general benefit to society?
Compatibility	Is the innovation compatible with current attitudes and ways of doing things?
Learning	How easily can the customer learn the correct use of the innovation?
Visibility	How evident are the advantages of the innovation to the prospective customer?
Appearance	Does the appearance of the innovation convey a message of desirable qualities?
Function	Does this innovation work better than the alternatives? – or fulfill a function not now provided?
Durability	Will this innovation endure "long usage"?

Service	Will this innovation require less servicing or less costly servicing than alternatives?
Price	Does this innovation have a price advantage over its competitors?
Existing Competition	Does this innovation already face competition in the marketplace that will make its entry difficult and costly?
New Competition	Is this innovation likely to face new competition in the marketplace from other innovations that must be expected to threaten its market share?
Marketing Research	How great an effort will be required to define the product and price that the final market will find acceptable?
Promotion Cost	Is the cost and effort of promotion to achieve market acceptance of the innovation in line with expected earnings?
Distribution	How difficult will it be to develop or access distribution channels for the innovation?
Legality	Does the invention meet the requirements of applicable laws, regulations and product standards and avoid exposure to product liability?
Development Risk	What degree of uncertainty is associated with complete successful development from the present condition of the innovation to the market ready state?
Dependence	To what degree does this innovation lose control of its market and sales due to its dependence on other products, processes, systems or services?
Protection	Is it likely that worthwhile commercial protection will be obtainable for this innovation through patents, trade secrets or other means?
Size of Investment	Is the total investment required for the project likely to be obtainable?
Potential Sales	Is the sales volume for this particular innovation likely to be sufficient to justify initiating the project?
Payback Period	Will the initial investment be recovered in the early life of the innovation?
Profitability	Will the expected revenue from the innovation provide more profits than other investment opportunities?

Appendix B
SAS Code

Chapter 5 Code:

```
* Set up data *;

data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=1;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;
  if proto=1 then delete;
run;


* Logistic Regression Analysis *;

proc logistic data=tob0 descend;
   model Q1 = V1 V2 V3 V4 V5 V6 V8 V9 V10 V11 V12 V13 V14
   V15 V16 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33
   V34 V35 V37 V38 V39 V40 / selection=backward
                             lackfit
                             rsq
                             ctable
                             pprob=(0 to 1 by 0.01);

run;


proc logistic data=tob0 descend;
   model Q1 = V1 V2 V3 V4 V5 V6 V8 V9 V10 V11 V12 V13 V14
   V15 V16 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33
   V34 V35 V37 V38 V39 V40 / selection=forward
                             lackfit
                             rsq
                             ctable
                             pprob=(0 to 1 by 0.01);
run;


proc logistic data=tob0 descend;
   model Q1 = V1 V2 V3 V4 V5 V6 V8 V9 V10 V11 V12 V13 V14
   V15 V16 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33
   V34 V35 V37 V38 V39 V40 / selection=stepwise
                             slentry=.2
                             lackfit
                             rsq
                             ctable
                             pprob=(0 to 1 by 0.01);
```

```
  run;



* Set up data *;

data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=1;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;
  if proto=1 then delete;
run;


%inc 'C:\Documents and Settings\gsampson\My Documents\Glen\Thesis
Stuff\GLEN\Thesis3\Step_6\Sas Programs\classify.sas';


* Logistic Regression Analysis *;

proc logistic data=tob0;
   model rating2 = V1 V2 V3 V4 V5 V6 V8 V9 V10 V11 V12 V13 V14
   V15 V16 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33
   V34 V35 V37 V38 V39 V40 / selection=backward
                             lackfit
                             rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )

run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD selection model';
   table rating2*_predlvl;

run;
```

```
proc logistic data=tob0;
   model rating2 = V1 V2 V3 V4 V5 V6 V8 V9 V10 V11 V12 V13 V14
   V15 V16 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33
   V34 V35 V37 V38 V39 V40 / selection=forward
                             lackfit
                             rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )

run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of FORWARD selection model';
   table rating2*_predlvl;

run;


proc logistic data=tob0;
   model rating2 = V1 V2 V3 V4 V5 V6 V8 V9 V10 V11 V12 V13 V14
   V15 V16 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33
   V34 V35 V37 V38 V39 V40 / selection=stepwise
                             slentry=.2
                             lackfit
                             rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )

run;
```

```
proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of STEPWISE selection model';
   table rating2*_predlvl;

  run;


* Set up data *;

data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=1;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;

  if rating2='A' then rating3=10;
  if rating2='B' then rating3=8;
  if rating2='C' then rating3=6;
  if rating2='D' then rating3=4;
  if rating2='E' then rating3=2;

  if proto=1 then delete;
  run;


* Logistic Regression Analysis *;


proc logistic data=tob0 descend;
   model Q1 = rating3 / lackfit
                        rsq
                        ctable
                        pprob=(0 to 1 by 0.01) ;

run;



* Set up data *;

data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=1;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;

  if rating2='A' then rating3=10;
```

```
  if rating2='B' then rating3=8;
  if rating2='C' then rating3=6;
  if rating2='D' then rating3=4;
  if rating2='E' then rating3=2;

  if proto=1 then delete;
run;


* Create the training dataset *;

data tobEstimate;
  set tob0;
  if year >= 89;
  if year <= 92;
run;


* Create the test dataset *;

data tobTest;
  set tob0;
  if year=93;
run;


* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> Q1 model *;
* Use the Variables determined in Step 1 *;

proc logistic data=tobEstimate outest=parms descend;
   model Q1 = V23 V33 V35 V40 / selection=none
                                lackfit
                                rsq
                                ctable
                                pprob=(0 to 1 by 0.01);

run;


* Use parameters on test data *;

proc logistic data=tobTest inest=parms descend;
   model Q1 = V23 V33 V35 V40 / selection=none
                                maxiter=0
                                lackfit
                                rsq
                                ctable
                                pprob=(0 to 1 by 0.01);

run;
```

```
* Estimate the Parameters for Logistic Regression Analysis *;
* rating2 --> Q1 model *;

proc logistic data=tobEstimate outest=parms2 descend;
   model Q1 = rating3 / selection=none
                        lackfit
                        rsq
                        ctable
                        pprob=(0 to 1 by 0.01);

run;


* Use parameters on test data *;

proc logistic data=tobTest inest=parms2 descend;
   model Q1 = rating3 / selection=none
                        maxiter=0
                        lackfit
                        rsq
                        ctable
                        pprob=(0 to 1 by 0.01);

run;


* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> R2 model *;
* Use the Variables determined in Step 2 Comparison -
Backward/Forward/Stepwise Elimination*;

proc logistic data=tobEstimate outest=parms descend;
   model rating2 = V1 V14 V22-V23 V31-V33 V35 V37-V38 V40 / selection=none
                                                            lackfit
                                                            rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- V1-V40 - Estimate';
```

```
   table rating2*_predlvl;

run;



* Use parameters on test data *;

proc logistic data=tobTest inest=parms descend;
   model rating2 = V1 V14 V22-V23 V31-V33 V35 V37-V38 V40 / selection=none
                                                             maxiter=0
                                                             lackfit
                                                             rsq;

   output out=thesis.predictR2 predicted=_pred;
run;

* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- V1-V40 - Test';
   table rating2*_predlvl;

  run;
```

Chapter 6 and Chapter 7 Code:

```
* Run initial factor analysis on Period 1 data *;


data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=1;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;

  if rating2='A' then rating3=10;
  if rating2='B' then rating3=8;
  if rating2='C' then rating3=6;
  if rating2='D' then rating3=4;
  if rating2='E' then rating3=2;

  if proto=1 then delete;
  run;


* Perform Orthoginal Rotations using Varimax on nfactors for 7, 10 and 14
Factors*;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=7
                  out=thesis.factors_7_VM
                  scree
                  msa
                  preplot
                  plot
                  score
                  round
                  flag=.5;
   title " Factor Analysis for Period II Varimax with 7 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=10
                  out=thesis.factors_10_VM
                  scree
                  msa
                  preplot
                  plot
```

```
                  score
                  heywood;
   title " Factor Analysis for Period II Varimax with 10 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=14
                  out=thesis.factors_14_VM
                  scree
                  msa
                  preplot
                  plot
                  score
                  heywood;
   title " Factor Analysis for Period II Varimax with 14 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;


* Perform Oblique Rotations using Promax on nfactors for 7, 10, and 14
Factors*;


proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=7
                  out=thesis.factors_7_PM
                  scree
                  msa
                  preplot
                  plot
                  score;
   title " Factor Analysis for Period II Promax with 7 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=10
                  out=thesis.factors_10_PM
                  scree
                  msa
                  preplot
                  plot
                  score
```

```
                  heywood;
   title " Factor Analysis for Period II Promax with 10 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=14
                  out=thesis.factors_14_PM
                  scree
                  msa
                  preplot
                  plot
                  score
                  heywood;
   title " Factor Analysis for Period II Promax with 14 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;
```

```
* Run initial factor analysis on Period 1 data *;

data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=1;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;

  if rating2='A' then rating3=10;
  if rating2='B' then rating3=8;
  if rating2='C' then rating3=6;
  if rating2='D' then rating3=4;
  if rating2='E' then rating3=2;

  if proto=1 then delete;
  run;


* Perform Orthoginal Rotations using Varimax on nfactors for 7 and 10
Factors*;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=7
                  out=thesis.factors_7_VM
                  scree
                  msa
                  preplot
                  plot
                  score;
   title " Factor Analysis for Period II Varimax with 7 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=10
                  out=thesis.factors_10_VM
                  scree
                  msa
                  preplot
                  plot
                  score
                  heywood;
   title " Factor Analysis for Period II Varimax with 10 Factors";
```

```
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

* Perform Oblique Rotations using Promax on nfactors for 7 and 10
Factors*;


proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=7
                  out=thesis.factors_7_PM
                  scree
                  msa
                  preplot
                  plot
                  score;
   title " Factor Analysis for Period II Promax with 7 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=10
                  out=thesis.factors_10_PM
                  scree
                  msa
                  preplot
                  plot
                  score
                  heywood;
   title " Factor Analysis for Period II Promax with 10 Factors";
   var v1-v6 v8-v16 v18-v35 v37-v40;
run;

* Logistic Regression Analysis Comparison to Step1 V1-V40 -> Q1 *;
* Backward *;

proc logistic data=thesis.factors_7_VM descend;
   model Q1 = Factor1-Factor7  / selection=backward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_10_VM descend;
   model Q1 = Factor1-Factor10 / selection=backward
                                 lackfit
```

```
                                    rsq
                                    ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_7_PM descend;
   model Q1 = Factor1-Factor7  / selection=backward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_10_PM descend;
   model Q1 = Factor1-Factor10 / selection=backward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;

* Forward *;

proc logistic data=thesis.factors_7_VM descend;
   model Q1 = Factor1-Factor7  / selection=forward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_10_VM descend;
   model Q1 = Factor1-Factor10 / selection=forward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_7_PM descend;
   model Q1 = Factor1-Factor7  / selection=forward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_10_PM descend;
```

```
   model Q1 = Factor1-Factor10 / selection=forward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;

* Stepwise *;

proc logistic data=thesis.factors_7_VM descend;
   model Q1 = Factor1-Factor7  / selection=stepwise
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_10_VM descend;
   model Q1 = Factor1-Factor10 / selection=stepwise
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_7_PM descend;
   model Q1 = Factor1-Factor7  / selection=stepwise
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_10_PM descend;
   model Q1 = Factor1-Factor10 / selection=stepwise
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;




* Logistic Regression Analysis Comparison to Step2 V1-V40 -> Rating2 *;
* 7 Factors - Orthoginal *;

%inc 'C:\Documents and Settings\gsampson\My Documents\Glen\Thesis
Stuff\GLEN\Thesis3\Step_6\Sas Programs\classify.sas';
```

```
proc logistic data=thesis.factors_7_VM;
   model rating2 = Factor1-Factor7  / selection=backward
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD selection model - 7 Factors -
Orthoginal';
   table rating2*_predlvl;

run;


proc logistic data=thesis.factors_7_VM;
   model rating2 = Factor1-Factor7  / selection=forward
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of FORWARD selection model - 7 Factors -
Orthoginal';
   table rating2*_predlvl;

run;
```

```
proc logistic data=thesis.factors_7_VM;
   model rating2 = Factor1-Factor7  / selection=stepwise
                                      slentry=.2
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of STEPWISE selection model - 7 Factors -
Orthoginal';
   table rating2*_predlvl;

run;



* Logistic Regression Analysis Comparison to Step2 V1-V40 -> Rating2 *;
* 7 Factors - Oblique*;

proc logistic data=thesis.factors_7_PM;
   model rating2 = Factor1-Factor7  / selection=backward
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
```

```
   title 'Predictive ability of BACKWARD selection model - 7 Factors -
Oblique';
   table rating2*_predlvl;

run;


proc logistic data=thesis.factors_7_PM;
   model rating2 = Factor1-Factor7  / selection=forward
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of FORWARD selection model - 7 Factors -
Oblique';
   table rating2*_predlvl;

run;


proc logistic data=thesis.factors_7_PM;
   model rating2 = Factor1-Factor7  / selection=stepwise
                                      slentry=.2
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;
```

```
proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of STEPWISE selection model - 7 Factors -
Oblique';
   table rating2*_predlvl;

run;



* Logistic Regression Analysis Comparison to Step2 V1-V40 -> Rating2 *;
* 10 Factors - Orthoginal*;

proc logistic data=thesis.factors_10_VM;
   model rating2 = Factor1-Factor10 / selection=backward
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (   data=thesis.predictR2,
              response=rating2,
              p=_pred,
              predname=_predlvl,
              out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD selection model - 10 Factors -
Orthoginal';
   table rating2*_predlvl;

run;


proc logistic data=thesis.factors_10_VM;
   model rating2 = Factor1-Factor10 / selection=forward
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (   data=thesis.predictR2,
              response=rating2,
```

```
               p=_pred,
               predname=_predlvl,
               out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of FORWARD selection model - 10 Factors -
Orthoginal';
   table rating2*_predlvl;

run;


proc logistic data=thesis.factors_10_VM;
   model rating2 = Factor1-Factor10 / selection=stepwise
                                       slentry=.2
                                       lackfit
                                       rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of STEPWISE selection model - 10 Factors -
Orthoginal';
   table rating2*_predlvl;

run;



* Logistic Regression Analysis Comparison to Step2 V1-V40 -> Rating2 *;
* 10 Factors - Oblique*;

proc logistic data=thesis.factors_10_PM;
   model rating2 = Factor1-Factor10 / selection=backward
                                       lackfit
                                       rsq;

   output out=thesis.predictR2 predicted=_pred;
run;
```

```
* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD selection model - 10 Factors -
Oblique';
   table rating2*_predlvl;

run;


proc logistic data=thesis.factors_10_PM;
   model rating2 = Factor1-Factor10 / selection=forward
                                      lackfit
                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of FORWARD selection model - 10 Factors -
Oblique';
   table rating2*_predlvl;

run;


proc logistic data=thesis.factors_10_PM;
   model rating2 = Factor1-Factor10 / selection=stepwise
                                      slentry=.2
                                      lackfit
                                      rsq;
```

```
   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of STEPWISE selection model - 10 Factors -
Oblique';
   table rating2*_predlvl;

run;



* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* 7 Factors - Orthoginal*;

* Create the training dataset *;
data Estimate7vm;
  set thesis.factors_7_VM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test7vm;
  set thesis.factors_7_VM;
  if year=93;
run;


* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> Q1 model *;
* Use the Variables determined in Step 1 - Backward Elimination*;

proc logistic data=Estimate7vm outest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;
```

```
* Use parameters on test data *;

proc logistic data=Test7vm inest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                      maxiter=0
                                      lackfit
                                      rsq
                                      ctable
                                      pprob=(0 to 1 by 0.01);

run;

* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> Q1 model *;
* Use the Variables determined in Step 1 - Forward Elimination*;

proc logistic data=Estimate7vm outest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                      lackfit
                                      rsq
                                      ctable
                                      pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test7vm inest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                      maxiter=0
                                      lackfit
                                      rsq
                                      ctable
                                      pprob=(0 to 1 by 0.01);

run;

* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> Q1 model *;
* Use the Variables determined in Step 1 - Stepwise Elimination*;

proc logistic data=Estimate7vm outest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                      lackfit
                                      rsq
                                      ctable
                                      pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;
```

```
proc logistic data=Test7vm inest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                        maxiter=0
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;




* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* 7 Factors - Oblique*;

* Create the training dataset *;
data Estimate7pm;
  set thesis.factors_7_PM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test7pm;
  set thesis.factors_7_PM;
  if year=93;
run;


* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> Q1 model *;
* Use the Variables determined in Step 1  - Backward Elimination*;

proc logistic data=Estimate7pm outest=parms descend;
   model Q1 = Factor1 Factor3 Factor7 / selection=none
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test7pm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor7 / selection=none
                                        maxiter=0
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);
```

```
run;
* Use the Variables determined in Step 1  - Forward Elimination*;

proc logistic data=Estimate7pm outest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test7pm inest=parms descend;
   model Q1 = Factor1-Factor3 Factor7 / selection=none
                                        maxiter=0
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;
* Use the Variables determined in Step 1  - Stepwise Elimination*;

proc logistic data=Estimate7pm outest=parms descend;
   model Q1 = Factor1 Factor3 Factor7 / selection=none
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test7pm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor7 / selection=none
                                        maxiter=0
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;


* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* 10 Factors - Orthoginal*;

* Create the training dataset *;
```

```
data Estimate10vm;
  set thesis.factors_10_VM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test10vm;
  set thesis.factors_10_VM;
  if year=93;
run;


* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> Q1 model *;
* Use the Variables determined in Step 1 - Backward Elimination*;

proc logistic data=Estimate10vm outest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 Factor7 / selection=none
                                                lackfit
                                                rsq
                                                ctable
                                                pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test10vm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 Factor7 / selection=none
                                                maxiter=0
                                                lackfit
                                                rsq
                                                ctable
                                                pprob=(0 to 1 by 0.01);

run;
* Use the Variables determined in Step 1 - Forward Elimination*;

proc logistic data=Estimate10vm outest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 Factor7 / selection=none
                                                lackfit
                                                rsq
                                                ctable
                                                pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test10vm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 Factor7 / selection=none
                                                maxiter=0
```

```
                                                      lackfit
                                                      rsq
                                                      ctable
                                                      pprob=(0 to 1 by 0.01);

run;
* Use the Variables determined in Step 1 - Stepwise Elimination*;

proc logistic data=Estimate10vm outest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 Factor7 / selection=none
                                                      lackfit
                                                      rsq
                                                      ctable
                                                      pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test10vm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 Factor7 / selection=none
                                                      maxiter=0
                                                      lackfit
                                                      rsq
                                                      ctable
                                                      pprob=(0 to 1 by 0.01);

run;


* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* 10 Factors - Oblique*;

* Create the training dataset *;
data Estimate10pm;
  set thesis.factors_10_PM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test10pm;
  set thesis.factors_10_PM;
  if year=93;
run;


* Estimate the Parameters for Logistic Regression Analysis *;
* V1 - V40 --> Q1 model *;
* Use the Variables determined in Step 1 - Backward Elimination*;

proc logistic data=Estimate10pm outest=parms descend;
```

```
   model Q1 = Factor1 Factor3 Factor5 / selection=none
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test10pm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 / selection=none
                                        maxiter=0
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;

* Use the Variables determined in Step 1 - Forward Elimination*;

proc logistic data=Estimate10pm outest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 / selection=none
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;

* Use parameters on test data *;

proc logistic data=Test10pm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 / selection=none
                                        maxiter=0
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;

* Use the Variables determined in Step 1 - Stepwise Elimination*;

proc logistic data=Estimate10pm outest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 / selection=none
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;
```

```
* Use parameters on test data *;

proc logistic data=Test10pm inest=parms descend;
   model Q1 = Factor1 Factor3 Factor5 / selection=none
                                        maxiter=0
                                        lackfit
                                        rsq
                                        ctable
                                        pprob=(0 to 1 by 0.01);

run;



************************************************************************
*************;

* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* F1-F7 --> R2 Model Comparison to V1-V40 --> R2 Model*;
* 7 Factors - Orthoginal*;

* Create the training dataset *;
data Estimate7vm;
  set thesis.factors_7_VM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test7vm;
  set thesis.factors_7_VM;
  if year=93;
run;

* Estimate the Parameters for Logistic Regression Analysis *;
* F1 - F13 --> R2 model *;
* Use the Variables determined in Step 2 Comparison -
Backward/Forward/Stepwise Elimination*;

proc logistic data=Estimate7vm outest=parms descend;
   model rating2 = Factor1-Factor5 Factor7 / selection=none
                                             lackfit
                                             rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;
```

```
%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 7 Factors - Orthoginal - Estimate';
   table rating2*_predlvl;

run;



* Use parameters on test data *;

proc logistic data=Test7vm inest=parms descend;
   model rating2 = Factor1-Factor5 Factor7 / selection=none
                                             maxiter=0
                                             lackfit
                                             rsq;

   output out=thesis.predictR2 predicted=_pred;
run;

* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 7 Factors - Orthoginal - Test';
   table rating2*_predlvl;

run;


* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* F1-F7 --> R2 Model Comparison to V1-V40 --> R2 Model*;
* 7 Factors - Oblique*;

* Create the training dataset *;
data Estimate7pm;
```

```
  set thesis.factors_7_PM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test7pm;
  set thesis.factors_7_PM;
  if year=93;
run;

* Estimate the Parameters for Logistic Regression Analysis *;
* F1 - F13 --> R2 model *;
* Use the Variables determined in Step 2 Comparison -
Backward/Forward/Stepwise Elimination*;

proc logistic data=Estimate7pm outest=parms descend;
   model rating2 = Factor1-Factor3 Factor5-Factor7 / selection=none
                                                     lackfit
                                                     rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (   data=thesis.predictR2,
              response=rating2,
              p=_pred,
              predname=_predlvl,
              out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 7 Factors - Oblique - Estimate';
   table rating2*_predlvl;

run;



* Use parameters on test data *;

proc logistic data=Test7pm inest=parms descend;
   model rating2 = Factor1-Factor3 Factor5-Factor7 / selection=none
                                                     maxiter=0
                                                     lackfit
                                                     rsq;

   output out=thesis.predictR2 predicted=_pred;
```

```
run;

* Use the CLASSIFY macro to determine the predictions *;

%classify (   data=thesis.predictR2,
              response=rating2,
              p=_pred,
              predname=_predlvl,
              out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 7 Factors - Oblique - Test';
   table rating2*_predlvl;

run;


* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* F1-F7 --> R2 Model Comparison to V1-V40 --> R2 Model*;
* 10 Factors - Orthoginal*;

* Create the training dataset *;
data Estimate10vm;
  set thesis.factors_10_VM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test10vm;
  set thesis.factors_10_VM;
  if year=93;
run;

* Estimate the Parameters for Logistic Regression Analysis *;
* F1 - F13 --> R2 model *;
* Use the Variables determined in Step 2 Comparison -
Backward/Forward/Stepwise Elimination*;

proc logistic data=Estimate10vm outest=parms descend;
   model rating2 = Factor1-Factor7 Factor9-Factor10 / selection=none
                                                      lackfit
                                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;
```

```
%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 10 Factors - Orthoginal - Estimate';
   table rating2*_predlvl;

run;



* Use parameters on test data *;

proc logistic data=Test10vm inest=parms descend;
   model rating2 = Factor1-Factor7 Factor9-Factor10 / selection=none
                                                      maxiter=0
                                                      lackfit
                                                      rsq;

   output out=thesis.predictR2 predicted=_pred;
run;

* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 10 Factors - Orthoginal - Test';
   table rating2*_predlvl;

run;


* Logistic Regression Analysis Comparison to Step4 - Out of Sample
Prediction Tests *;
* F1-F7 --> R2 Model Comparison to V1-V40 --> R2 Model*;
* 10 Factors - Oblique*;

* Create the training dataset *;
```

```
data Estimate10pm;
  set thesis.factors_10_PM;
  if year >= 89;
  if year <= 92;
run;

* Create the test dataset *;
data Test10pm;
  set thesis.factors_10_PM;
  if year=93;
run;

* Estimate the Parameters for Logistic Regression Analysis *;
* F1 - F13 --> R2 model *;
* Use the Variables determined in Step 2 Comparison -
Backward/Forward/Stepwise Elimination*;

proc logistic data=Estimate10pm outest=parms descend;
   model rating2 = Factor1-Factor3 Factor5 / selection=none
                                             lackfit
                                             rsq;

   output out=thesis.predictR2 predicted=_pred;
run;


* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 10 Factors - Oblique - Estimate';
   table rating2*_predlvl;

run;



* Use parameters on test data *;

proc logistic data=Test10pm inest=parms descend;
   model rating2 = Factor1-Factor3 Factor5 / selection=none
                                             maxiter=0
                                             lackfit
                                             rsq;
```

```
   output out=thesis.predictR2 predicted=_pred;
run;

* Use the CLASSIFY macro to determine the predictions *;

%classify (  data=thesis.predictR2,
             response=rating2,
             p=_pred,
             predname=_predlvl,
             out=thesis.predictlvlR2 )
run;


proc freq data=thesis.predictlvlR2;
   title 'Predictive ability of BACKWARD/FORWARD/STEPWISE selection model
- 10 Factors - Oblique - Test';
   table rating2*_predlvl;

run;


* Period 0 - Step 1 V1-V40 → Q1 *;

* Set up data *;

data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=0;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;
  if proto=1 then delete;
run;


proc logistic data=tob0 descend;
   model Q1 = v2-v3 v5-v7 v11-v32 v34-v36 v38-v40 / selection=backward
                                                     lackfit
                                                     rsq
                                                     ctable
                                                     pprob=(0 to 1
by 0.01);

run;


proc logistic data=tob0 descend;
   model Q1 = v2-v3 v5-v7 v11-v32 v34-v36 v38-v40 / selection=forward
                                                     lackfit
                                                     rsq
                                                     ctable
```

```
                                                                 pprob=(0 to 1
by 0.01);
run;


proc logistic data=tob0 descend;
   model Q1 = v2-v3 v5-v7 v11-v32 v34-v36 v38-v40 / selection=stepwise
                                                                 slentry=.2
                                                                 lackfit
                                                                 rsq
                                                                 ctable
                                                                 pprob=(0 to 1
by 0.01);
run;


* Logistic Regression Analysis Comparison to Step1 V1-V40 -> Q1 *;

proc logistic data=thesis.factors_5_vm_p0 descend;
   model Q1 = Factor1-Factor5 / selection=backward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_5_pm_p0 descend;
   model Q1 = Factor1-Factor5 / selection=backward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_7_vm_p0 descend;
   model Q1 = Factor1-Factor7 / selection=backward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;


proc logistic data=thesis.factors_7_pm_p0 descend;
   model Q1 = Factor1-Factor7 / selection=backward
                                 lackfit
                                 rsq
                                 ctable pprob=(0 to 1 by 0.01);

run;
```

```
* Run initial factor analysis on Period 0 data *;

data tob0;
  set thesis.ciicadj;
*  if known=0;
  if period=0;
  if q71=1 then q1=1;
  if q72=1 then q1=1;
  if q73=1 then q1=1;

  if rating2='A' then rating3=10;
  if rating2='B' then rating3=8;
  if rating2='C' then rating3=6;
  if rating2='D' then rating3=4;
  if rating2='E' then rating3=2;

  if proto=1 then delete;
  run;


* Perform Orthoginal Rotations using Varimax on nfactors for 5, 7 and 10
Factors*;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=5
                  out=thesis.factors_5_VM_P0
                  scree
                  msa
                  preplot
                  plot
                  score
                  round
                  flag=.5;
   title " Factor Analysis for Period I Varimax with 5 Factors";
   var v2-v3 v5-v7 v11-v32 v34-v36 v38-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=7
                  out=thesis.factors_7_VM_P0
                  scree
                  msa
                  preplot
```

```
                  plot
                  score
                  heywood;
   title " Factor Analysis for Period I Varimax with 7 Factors";
   var v2-v3 v5-v7 v11-v32 v34-v36 v38-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=varimax
                  nfactors=10
                  out=thesis.factors_10_VM_P0
                  scree
                  msa
                  preplot
                  plot
                  score
                  heywood;
   title " Factor Analysis for Period I Varimax with 10 Factors";
   var v2-v3 v5-v7 v11-v32 v34-v36 v38-v40;
run;


* Perform Oblique Rotations using Promax on nfactors for 5, 7, and 10
Factors*;


proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=5
                  out=thesis.factors_5_PM_P0
                  scree
                  msa
                  preplot
                  plot
                  score;
   title " Factor Analysis for Period I Promax with 5 Factors";
   var v2-v3 v5-v7 v11-v32 v34-v36 v38-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=7
                  out=thesis.factors_7_PM_P0
                  scree
                  msa
                  preplot
                  plot
```

```
                  score
                  heywood;
   title " Factor Analysis for Period I Promax with 7 Factors";
   var v2-v3 v5-v7 v11-v32 v34-v36 v38-v40;
run;

proc factor data=tob0
                  corr
                  method=ml
                  rotate=promax
                  nfactors=10
                  out=thesis.factors_10_PM_P0
                  scree
                  msa
                  preplot
                  plot
                  score
                  heywood;
   title " Factor Analysis for Period I Promax with 10 Factors";
   var v2-v3 v5-v7 v11-v32 v34-v36 v38-v40;
run;
```

Appendix C
Additional Factor Loadings

Factor Loadings for 10 Factors (Post 1989 Theoretical Structure / Post 1989 Data Empirical Results)

		Orthoganaly Rotated Factors										Oblique Rotated Factor Pattern									
Post 1989 Theoretical Structure		F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Technical Feasibility	V1	**0.73**	0.15	0.02	0.19	0.04	0.04	0.09	-0.11	0.07	0.02	**0.80**	0.02	-0.12	0.07	0.08	0.04	-0.01	-0.09	-0.04	-0.04
Functional Performance	V2	**0.63**	0.32	0.15	0.10	0.01	0.02	0.14	-0.11	0.11	0.19	**0.60**	0.25	0.06	-0.05	-0.03	-0.01	0.01	-0.13	0.03	0.11
Research & Development	V3	**0.69**	0.09	0.13	0.24	-0.07	0.03	0.14	0.09	0.06	0.09	**0.73**	-0.09	0.06	0.08	-0.08	0.02	0.01	0.13	-0.01	0.02
Technology Significance	V4	-0.06	0.10	0.13	0.02	**0.52**	0.15	0.05	0.10	0.04	0.09	-0.08	0.00	0.01	0.04	**0.53**	0.11	0.00	0.04	0.06	0.11
Safety	V5	0.05	0.29	0.03	0.17	0.02	-0.05	0.07	0.29	0.00	0.35	-0.03	0.27	-0.01	0.12	0.00	-0.07	-0.16	0.24	0.02	0.35
Environmental Impact	V6	-0.10	0.18	0.06	0.19	0.07	0.00	0.04	**0.50**	-0.09	0.15	-0.13	0.10	0.05	0.13	0.04	-0.05	-0.09	**0.48**	-0.03	0.15
Technology Production	V8	0.36	0.18	0.03	**0.48**	-0.09	-0.05	0.00	0.09	0.03	0.00	0.30	0.11	-0.05	**0.43**	-0.07	-0.04	-0.06	0.05	0.00	-0.05
Tooling Cost	V9	0.19	0.02	0.18	**0.57**	0.02	-0.02	0.08	-0.02	0.02	0.09	0.05	-0.07	0.11	**0.59**	0.02	0.01	0.03	-0.10	0.00	0.08
Cost of Production	V10	0.14	0.24	0.09	**0.59**	0.07	0.08	0.06	0.12	-0.03	-0.08	0.02	0.17	-0.02	**0.60**	0.06	0.08	0.02	0.02	-0.02	-0.11
Need	V11	-0.05	0.31	0.22	0.01	**0.43**	0.12	0.20	0.21	-0.19	-0.02	-0.08	0.23	0.09	-0.04	0.37	0.00	0.15	0.12	-0.17	-0.05
Potential Market	V12	-0.05	0.28	0.00	0.08	0.14	0.16	0.01	0.22	-0.31	-0.07	-0.03	0.26	-0.08	0.07	0.11	0.08	-0.02	0.14	-0.29	-0.10
Trend of Demand	V13	0.03	0.11	0.17	0.06	0.27	**0.40**	0.07	0.22	-0.21	-0.02	0.07	-0.05	0.10	0.04	0.21	0.36	-0.01	0.16	-0.20	-0.03
Durations of Demand	V14	-0.01	0.19	0.04	-0.04	0.13	**0.97**	0.03	0.06	0.05	0.03	-0.03	0.05	0.01	0.02	0.01	**1.02**	-0.06	-0.04	0.10	0.03
Demand Predictability	V15	0.06	0.09	0.09	-0.04	0.05	0.15	0.22	**0.58**	0.01	0.09	0.09	-0.08	0.07	-0.17	0.00	0.08	0.19	**0.62**	0.07	0.05
Product Line Potential	V16	-0.03	0.13	0.06	-0.01	0.29	0.03	0.07	0.38	-0.01	0.07	-0.01	0.03	-0.02	-0.07	0.29	-0.05	0.04	0.38	0.02	0.07
Societal Benefits	V18	0.11	0.09	0.04	-0.10	0.18	0.10	0.21	0.11	-0.07	**0.54**	0.08	-0.01	-0.08	-0.12	0.19	0.07	0.13	0.09	-0.07	**0.55**
Compatibility	V19	0.07	**0.45**	0.27	-0.01	0.08	0.10	0.19	0.04	-0.10	0.29	-0.07	**0.44**	0.23	-0.09	-0.05	0.03	0.08	-0.07	-0.10	0.24
Learning	V20	0.23	**0.44**	0.00	0.13	-0.26	0.08	0.21	0.11	-0.07	0.14	0.12	**0.46**	-0.04	0.04	-0.34	0.04	0.17	0.04	-0.06	0.06
Visibility	V21	0.12	**0.55**	0.09	0.02	0.06	0.14	0.16	0.07	-0.06	0.13	0.01	**0.57**	0.00	-0.05	-0.04	0.07	0.10	-0.04	-0.05	0.07
Appearance	V22	0.13	**0.48**	0.20	0.07	0.13	0.02	0.05	0.09	0.02	0.01	0.04	**0.50**	0.16	-0.04	0.03	-0.05	-0.03	0.01	0.01	-0.04
Function	V23	0.34	**0.43**	0.16	0.06	0.31	-0.01	0.02	-0.03	0.30	-0.01	0.28	**0.42**	0.05	-0.05	0.28	-0.05	-0.07	-0.07	0.26	-0.05
Durability	V24	0.11	**0.50**	0.01	0.22	0.12	0.06	0.00	0.18	0.09	-0.01	0.00	**0.51**	-0.08	0.17	0.07	0.02	-0.04	0.09	0.12	-0.04
Service	V25	0.04	0.35	0.07	0.28	0.00	0.02	0.14	0.12	-0.03	0.11	-0.10	0.36	-0.01	0.25	-0.04	-0.02	0.11	0.03	-0.01	0.07
Price	V26	0.11	0.05	0.22	0.28	0.22	-0.01	0.19	-0.05	-0.03	0.02	0.03	-0.03	0.12	0.27	0.21	-0.04	0.16	-0.11	-0.05	-0.01
Existing Competition	V27	0.01	0.07	0.14	0.09	0.16	-0.02	0.05	0.06	**0.56**	-0.01	-0.12	0.05	0.14	0.06	0.14	0.01	0.07	0.07	**0.60**	-0.01
New Competition	V28	0.11	0.06	-0.05	-0.06	0.01	0.01	0.08	-0.06	**0.57**	0.04	0.05	-0.06	-0.06	-0.07	0.03	0.07	0.12	0.00	**0.58**	0.04
Marketing Research	V29	0.19	0.18	0.07	0.20	-0.08	0.07	**0.50**	0.19	0.02	-0.07	0.09	0.09	-0.04	0.12	-0.13	0.00	**0.55**	0.17	0.05	-0.15
Promotion Cost	V30	0.09	0.13	0.22	0.04	0.13	0.01	**0.53**	-0.09	0.07	0.11	-0.06	0.08	0.10	0.00	0.07	-0.05	**0.55**	-0.14	0.08	0.05
Distribution	V31	0.08	0.06	0.06	0.09	0.09	0.00	**0.45**	0.13	0.09	0.10	-0.01	-0.02	-0.06	0.05	0.08	-0.06	0.49	0.12	0.12	0.05
Legality	V32	0.16	0.09	-0.07	0.12	-0.12	-0.04	0.02	0.11	0.13	**0.51**	0.08	0.06	-0.12	0.10	-0.09	0.00	-0.05	0.10	0.13	**0.52**
Development Risks	V33	**0.50**	0.05	0.27	0.27	0.03	-0.07	0.13	0.06	0.00	0.21	**0.49**	-0.12	0.22	0.13	0.01	-0.08	0.00	0.07	-0.07	0.16
Dependence	V34	0.08	0.21	0.04	0.06	0.14	0.17	0.00	0.04	-0.15	0.06	0.08	0.17	-0.04	0.05	0.12	0.14	-0.07	-0.02	-0.16	0.04
Protection	V35	0.04	0.02	0.10	0.05	**0.57**	0.04	0.03	0.06	0.13	-0.04	0.06	-0.07	-0.05	0.06	**0.63**	0.00	0.01	0.04	0.12	-0.02
Investment Costs	V37	0.27	0.09	0.34	**0.52**	0.13	0.00	0.26	0.01	-0.01	0.15	0.12	-0.06	0.23	**0.47**	0.10	-0.01	0.17	-0.08	-0.03	0.11
Potential Sales	V38	0.13	0.29	**0.57**	0.11	0.28	0.06	0.13	0.16	-0.07	-0.06	0.06	0.18	**0.59**	-0.05	0.11	-0.03	0.00	0.08	-0.09	-0.12
Payback Period	V39	0.14	0.09	**0.70**	0.31	0.10	0.07	0.20	0.05	0.08	0.03	-0.03	-0.05	**0.78**	0.18	-0.09	0.05	0.07	-0.02	0.07	-0.02
Profitability	V40	0.14	0.18	**0.81**	0.18	0.21	0.06	0.08	0.13	0.11	-0.03	0.02	0.03	**0.94**	-0.01	-0.02	0.02	-0.11	0.07	0.09	-0.07

Factor Loadings for 10 Factors (Pre 1989 Theoretical Structure / Post 1989 Data Empirical Results)

		Orthogonally Rotated Factors										Oblique Rotated Factor Pattern									
Variables - PIES Structure		F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Legality	V32	0.16	0.09	-0.07	0.12	-0.12	-0.04	0.02	0.11	0.13	**0.51**	0.08	0.06	-0.12	0.10	-0.09	0.00	-0.05	0.10	0.13	**0.52**
Safety	V5	0.05	0.29	0.03	0.17	0.02	-0.05	0.07	0.29	0.00	0.35	-0.03	0.27	-0.01	0.12	0.00	-0.07	-0.16	0.24	0.02	0.35
Societal Benefits	V18	0.11	0.09	0.04	-0.10	0.18	0.10	0.21	0.11	-0.07	**0.54**	0.08	-0.01	-0.08	-0.12	0.19	0.07	0.13	0.09	-0.07	**0.55**
Environmental Impact	V6	-0.10	0.18	0.06	0.19	0.07	0.00	0.04	**0.50**	-0.09	0.15	-0.13	0.10	0.05	0.13	0.04	-0.05	-0.09	**0.48**	-0.03	0.15
Technical Feasibility	V1	**0.73**	0.15	0.02	0.19	0.04	0.04	0.09	-0.11	0.07	0.02	**0.80**	0.02	-0.12	0.07	0.08	0.04	-0.01	-0.09	-0.04	-0.04
Functional Performance	V2	**0.63**	0.32	0.15	0.10	0.01	0.02	0.14	-0.11	0.11	0.19	**0.60**	0.25	0.06	-0.05	-0.03	-0.01	0.01	-0.13	0.03	0.11
Technology Production	V8	0.36	0.18	0.03	**0.48**	-0.09	-0.05	0.00	0.09	0.03	0.00	0.30	0.11	-0.05	**0.43**	-0.07	-0.04	-0.06	0.05	0.00	-0.05
Tooling Cost	V9	0.19	0.02	0.18	**0.57**	0.02	-0.02	0.08	-0.02	0.02	0.09	0.05	-0.07	0.11	**0.59**	0.02	0.01	0.03	-0.10	0.00	0.08
Cost of Production	V10	0.14	0.24	0.09	**0.59**	0.07	0.08	0.06	0.12	-0.03	-0.08	0.02	0.17	-0.02	**0.60**	0.06	0.08	0.02	0.02	-0.02	-0.11
Development Risks	V33	**0.50**	0.05	0.27	0.27	0.03	-0.07	0.13	0.06	0.00	0.21	**0.49**	-0.12	0.22	0.13	0.01	-0.08	0.00	0.07	-0.07	0.16
Investment Costs	V37	0.27	0.09	0.34	**0.52**	0.13	0.00	0.26	0.01	-0.01	0.15	0.12	-0.06	0.23	**0.47**	0.10	-0.01	0.17	-0.08	-0.03	0.11
Payback Period	V39	0.14	0.09	**0.70**	0.31	0.10	0.07	0.20	0.05	0.08	0.03	-0.03	-0.05	**0.78**	0.18	-0.09	0.05	0.07	-0.02	0.07	-0.02
Profitability	V40	0.14	0.18	**0.81**	0.18	0.21	0.06	0.08	0.13	0.11	-0.03	0.02	0.03	**0.94**	-0.01	-0.02	0.02	-0.11	0.07	0.09	-0.07
Marketing Research	V29	0.19	0.18	0.07	0.20	-0.08	0.07	**0.50**	0.19	0.02	-0.07	0.09	0.09	-0.04	0.12	-0.13	0.00	**0.55**	0.17	0.05	-0.15
Research & Development	V3	**0.69**	0.09	0.13	0.24	-0.07	0.03	0.14	0.09	0.06	0.09	**0.73**	-0.09	0.06	0.08	-0.08	0.02	0.01	0.13	-0.01	0.02
Potential Market	V12	-0.05	0.28	0.00	0.08	0.14	0.16	0.01	0.22	-0.31	-0.07	-0.03	0.26	-0.08	0.07	0.11	0.08	-0.02	0.14	-0.29	-0.10
Potential Sales	V38	0.13	0.29	**0.57**	0.11	0.28	0.06	0.13	0.16	-0.07	-0.06	0.06	0.18	**0.59**	-0.05	0.11	-0.03	0.00	0.08	-0.09	-0.12
Trend of Demand	V13	0.03	0.11	0.17	0.06	0.27	**0.40**	0.07	0.22	-0.21	-0.02	0.07	-0.05	0.10	0.04	0.21	0.36	-0.01	0.16	-0.20	-0.03
Demand Predictability	V15	0.06	0.09	0.09	-0.04	0.05	0.15	0.22	**0.58**	0.01	0.09	0.09	-0.08	0.07	-0.17	0.00	0.08	0.19	**0.62**	0.07	0.05
Durations of Demand	V14	-0.01	0.19	0.04	-0.04	0.13	**0.97**	0.03	0.06	0.05	0.03	-0.03	0.05	0.01	0.02	0.01	**1.02**	-0.06	-0.04	0.10	0.03
Product Line Potential	V16	-0.03	0.13	0.06	-0.01	0.29	0.03	0.07	0.38	-0.01	0.07	-0.01	0.03	-0.02	-0.07	0.29	-0.05	0.04	0.38	0.02	0.07
Compatibility	V19	0.07	**0.45**	0.27	-0.01	0.08	0.10	0.19	0.04	-0.10	0.29	-0.07	**0.44**	0.23	-0.09	-0.05	0.03	0.08	-0.07	-0.10	0.24
Learning	V20	0.23	**0.44**	0.00	0.13	-0.26	0.08	0.21	0.11	-0.07	0.14	0.12	**0.46**	-0.04	0.04	-0.34	0.04	0.17	0.04	-0.06	0.06
Need	V11	-0.05	0.31	0.22	0.01	**0.43**	0.12	0.20	0.21	-0.19	-0.02	-0.08	0.23	0.09	-0.04	0.37	0.00	0.15	0.12	-0.17	-0.05
Dependence	V34	0.08	0.21	0.04	0.06	0.14	0.17	0.00	0.04	-0.15	0.06	0.08	0.17	-0.04	0.05	0.12	0.14	-0.07	-0.02	-0.16	0.04
Visibility	V21	0.12	**0.55**	0.09	0.02	0.06	0.14	0.16	0.07	-0.06	0.13	0.01	**0.57**	0.00	-0.05	-0.04	0.07	0.10	-0.04	-0.05	0.07
Promotion Cost	V30	0.09	0.13	0.22	0.04	0.13	0.01	**0.53**	-0.09	0.07	0.11	-0.06	0.08	0.10	0.00	0.07	-0.05	**0.55**	-0.14	0.08	0.05
Distribution	V31	0.08	0.06	0.06	0.09	0.09	0.00	**0.45**	0.13	0.09	0.10	-0.01	-0.02	-0.06	0.05	0.08	-0.06	0.49	0.12	0.12	0.05
Service	V25	0.04	0.35	0.07	0.28	0.00	0.02	0.14	0.12	-0.03	0.11	-0.10	0.36	-0.01	0.25	-0.04	-0.02	0.11	0.03	-0.01	0.07
Appearance	V22	0.13	**0.48**	0.20	0.07	0.13	0.02	0.05	0.09	0.02	0.01	0.04	**0.50**	0.16	-0.04	0.03	-0.05	-0.03	0.01	0.01	-0.04
Durability	V24	0.11	**0.50**	0.01	0.22	0.12	0.06	0.00	0.18	0.09	-0.01	0.00	**0.51**	-0.08	0.17	0.07	0.02	-0.04	0.09	0.12	-0.04
Function	V23	0.34	**0.43**	0.16	0.06	0.31	-0.01	0.02	-0.03	0.30	-0.01	0.28	**0.42**	0.05	-0.05	0.28	-0.05	-0.07	-0.07	0.26	-0.05
Price	V26	0.11	0.05	0.22	0.28	0.22	-0.01	0.19	-0.05	-0.03	0.02	0.03	-0.03	0.12	0.27	0.21	-0.04	0.16	-0.11	-0.05	-0.01
Existing Competition	V27	0.01	0.07	0.14	0.09	0.16	-0.02	0.05	0.06	**0.56**	-0.01	-0.12	0.05	0.14	0.06	0.14	0.01	0.07	0.07	**0.60**	-0.01
New Competition	V28	0.11	0.06	-0.05	-0.06	0.01	0.01	0.08	-0.06	**0.57**	0.04	0.05	-0.06	-0.06	-0.07	0.03	0.07	0.12	0.00	**0.58**	0.04
Protection	V35	0.04	0.02	0.10	0.05	**0.57**	0.04	0.03	0.06	0.13	-0.04	0.06	-0.07	-0.05	0.06	**0.63**	0.00	0.01	0.04	0.12	-0.02
Technology Significance	V4	-0.06	0.10	0.13	0.02	**0.52**	0.15	0.05	0.10	0.04	0.09	-0.08	0.00	0.01	0.04	**0.53**	0.11	0.00	0.04	0.06	0.11

Factor Loadings for 7 Factors (Pre 1989 Theoretical Structure / Pre 1989 Data Empirical Results)

		Orthoganaly Rotated Factors							Oblique Rotated Factor Pattern						
Pre 1989 Theoretical Structure		F1	F2	F3	F4	F5	F6	F7	F1	F2	F3	F4	F5	F6	F7
Safety	V5	**0.55**	0.32	0.13	-0.01	0.06	**0.53**	0.00	**0.47**	0.08	0.05	-0.14	-0.06	**0.52**	-0.03
Environmental Impact	V6	**0.56**	0.35	0.02	0.16	0.22	0.33	0.05	**0.49**	0.16	-0.17	0.07	0.11	0.23	0.03
Societal Benefits	V18	**0.57**	0.27	0.17	0.17	0.04	0.25	0.07	**0.53**	0.11	0.01	0.08	-0.10	0.17	0.04
Legality	V32	0.20	0.29	0.11	0.09	0.07	**0.64**	0.08	-0.05	0.02	0.09	0.02	0.00	**0.71**	0.06
Functional Feasibility	V2	0.30	**0.66**	0.28	0.06	-0.01	0.12	0.12	0.13	**0.73**	0.13	-0.04	-0.18	-0.05	0.10
Research & Development	V3	0.21	**0.76**	0.21	0.16	0.14	0.22	0.04	-0.05	**0.83**	0.02	0.08	-0.02	0.04	0.02
Production Feasibility	V7	0.30	**0.66**	0.18	0.18	0.16	0.11	-0.06	0.14	**0.75**	-0.02	0.10	0.02	-0.08	-0.08
Stage of Development	V17	**0.45**	**0.55**	0.29	0.00	0.27	0.16	0.01	0.31	**0.50**	0.14	-0.14	0.16	-0.02	-0.02
Marketing Research	V29	0.18	0.30	0.24	**0.54**	0.12	0.05	0.05	-0.01	0.25	0.06	**0.56**	0.00	-0.04	0.02
Investment Costs	V36	0.10	**0.42**	**0.62**	0.20	0.13	0.10	0.03	-0.16	0.35	**0.58**	0.11	0.04	0.03	-0.03
Payback Period	V39	0.20	0.31	**0.66**	0.20	0.11	0.20	0.03	-0.03	0.14	**0.64**	0.09	0.02	0.17	-0.04
Profitability	V40	0.19	0.23	**0.73**	0.12	0.15	0.07	0.00	0.02	0.07	**0.75**	0.00	0.09	0.03	-0.07
Potential Market	V12	**0.72**	0.23	0.13	0.15	0.20	0.17	0.01	**0.76**	0.04	-0.07	0.04	0.08	0.04	-0.02
Trend of Demand	V13	**0.63**	0.08	0.32	0.24	0.03	0.04	0.03	**0.69**	-0.10	0.18	0.16	-0.10	-0.05	-0.01
Demand Life Cycle	V14	0.35	-0.03	0.36	0.33	-0.04	-0.01	0.10	0.34	-0.19	0.27	0.31	-0.14	-0.04	0.07
Demand Predictability	V15	**0.49**	0.16	0.14	0.31	0.26	0.12	0.05	**0.43**	-0.02	-0.05	0.26	0.18	0.02	0.03
Product Line Potential	V16	**0.71**	0.18	0.18	0.02	0.19	0.04	0.04	**0.80**	0.02	0.02	-0.11	0.09	-0.11	0.02
Potential Sales	V38	**0.53**	0.16	**0.58**	0.06	0.19	0.11	0.10	**0.47**	-0.10	**0.53**	-0.09	0.11	0.04	0.05
Need	V11	**0.68**	0.17	0.27	0.24	0.10	-0.03	0.05	**0.74**	0.04	0.07	0.16	-0.04	-0.18	0.02
Compatibility	V19	**0.60**	0.33	0.24	0.16	0.05	0.18	0.23	**0.54**	0.16	0.06	0.05	-0.11	0.05	0.22
Learning	V20	**0.53**	**0.47**	0.13	0.14	0.11	0.29	-0.01	**0.44**	0.37	-0.04	0.04	-0.04	0.17	-0.04
Visibility	V21	**0.53**	0.33	0.36	0.20	0.20	0.09	0.10	**0.45**	0.18	0.19	0.09	0.07	-0.06	0.07
Appearnance	V22	**0.55**	0.30	0.13	0.18	0.17	0.22	**0.41**	**0.41**	0.07	-0.07	0.08	0.04	0.09	**0.44**
Promotion	V30	**0.49**	0.22	**0.42**	**0.42**	0.25	0.14	-0.01	0.37	0.01	0.25	0.34	0.14	0.05	-0.06
Distribution	V31	**0.45**	0.28	0.26	**0.49**	0.26	0.18	0.03	0.30	0.09	0.05	**0.45**	0.14	0.07	0.00
Dependence	V34	**0.52**	0.22	0.11	0.01	0.14	0.20	0.07	**0.51**	0.06	-0.01	-0.09	0.05	0.11	0.06
Function	V23	**0.47**	**0.48**	0.31	0.07	0.16	0.16	0.38	0.28	0.33	0.14	-0.06	0.03	-0.01	0.39
Durability	V24	**0.50**	**0.45**	0.09	0.10	0.27	0.23	0.33	0.33	0.28	-0.12	-0.01	0.15	0.07	0.35
Service	V25	0.14	**0.51**	0.20	0.20	0.16	0.19	0.20	-0.13	**0.46**	0.06	0.14	0.06	0.08	0.20
Price	V26	**0.45**	0.36	0.26	0.09	0.21	0.09	-0.01	0.39	0.28	0.12	-0.01	0.11	-0.04	-0.04
Existing Competition	V27	0.13	0.21	**0.42**	0.18	**0.53**	0.04	0.14	-0.12	0.01	0.35	0.08	**0.55**	-0.05	0.12
New Competition	V28	0.33	0.19	0.17	0.12	**0.54**	0.15	-0.01	0.20	-0.03	0.06	0.02	**0.56**	0.06	-0.03
Protection	V35	0.19	0.06	0.34	0.01	0.30	-0.03	0.06	0.12	-0.08	0.32	-0.08	0.31	-0.09	0.04

Factor Loadings for 7 Factors (Post 1989 Theoretical Structure / Pre 1989 Data Empirical Results)

		Orthoganaly Rotated Factors							Oblique Rotated Factor Pattern						
Post 1989 Theoretical Structure		F1	F2	F3	F4	F5	F6	F7	F1	F2	F3	F4	F5	F6	F7
Functional Feasibility	V2	0.30	**0.66**	0.28	0.06	-0.01	0.12	0.12	0.13	**0.73**	0.13	-0.04	-0.18	-0.05	0.10
Research & Development	V3	0.21	**0.76**	0.21	0.16	0.14	0.22	0.04	-0.05	**0.83**	0.02	0.08	-0.02	0.04	0.02
Safety	V5	**0.55**	0.32	0.13	-0.01	0.06	**0.53**	0.00	**0.47**	0.08	0.05	-0.14	-0.06	**0.52**	-0.03
Environmental Impact	V6	**0.56**	0.35	0.02	0.16	0.22	0.33	0.05	**0.49**	0.16	-0.17	0.07	0.11	0.23	0.03
Production Feasibility	V7	0.30	**0.66**	0.18	0.18	0.16	0.11	-0.06	0.14	**0.75**	-0.02	0.10	0.02	-0.08	-0.08
Need	V11	**0.68**	0.17	0.27	0.24	0.10	-0.03	0.05	**0.74**	0.04	0.07	0.16	-0.04	-0.18	0.02
Potential Market	V12	**0.72**	0.23	0.13	0.15	0.20	0.17	0.01	**0.76**	0.04	-0.07	0.04	0.08	0.04	-0.02
Trend of Demand	V13	**0.63**	0.08	0.32	0.24	0.03	0.04	0.03	**0.69**	-0.10	0.18	0.16	-0.10	-0.05	-0.01
Demand Life Cycle	V14	0.35	-0.03	0.36	0.33	-0.04	-0.01	0.10	0.34	-0.19	0.27	0.31	-0.14	-0.04	0.07
Demand Predictability	V15	**0.49**	0.16	0.14	0.31	0.26	0.12	0.05	**0.43**	-0.02	-0.05	0.26	0.18	0.02	0.03
Product Line Potential	V16	**0.71**	0.18	0.18	0.02	0.19	0.04	0.04	**0.80**	0.02	0.02	-0.11	0.09	-0.11	0.02
Societal Benefits	V18	**0.57**	0.27	0.17	0.17	0.04	0.25	0.07	**0.53**	0.11	0.01	0.08	-0.10	0.17	0.04
Compatibility	V19	**0.60**	0.33	0.24	0.16	0.05	0.18	0.23	**0.54**	0.16	0.06	0.05	-0.11	0.05	0.22
Learning	V20	**0.53**	**0.47**	0.13	0.14	0.11	0.29	-0.01	**0.44**	0.37	-0.04	0.04	-0.04	0.17	-0.04
Visibility	V21	**0.53**	0.33	0.36	0.20	0.20	0.09	0.10	**0.45**	0.18	0.19	0.09	0.07	-0.06	0.07
Appearnance	V22	**0.55**	0.30	0.13	0.18	0.17	0.22	**0.41**	**0.41**	0.07	-0.07	0.08	0.04	0.09	**0.44**
Function	V23	**0.47**	**0.48**	0.31	0.07	0.16	0.16	0.38	0.28	0.33	0.14	-0.06	0.03	-0.01	0.39
Durability	V24	**0.50**	**0.45**	0.09	0.10	0.27	0.23	0.33	0.33	0.28	-0.12	-0.01	0.15	0.07	0.35
Service	V25	0.14	**0.51**	0.20	0.20	0.16	0.19	0.20	-0.13	**0.46**	0.06	0.14	0.06	0.08	0.20
Price	V26	**0.45**	0.36	0.26	0.09	0.21	0.09	-0.01	0.39	0.28	0.12	-0.01	0.11	-0.04	-0.04
Existing Competition	V27	0.13	0.21	**0.42**	0.18	**0.53**	0.04	0.14	-0.12	0.01	0.35	0.08	**0.55**	-0.05	0.12
New Competition	V28	0.33	0.19	0.17	0.12	**0.54**	0.15	-0.01	0.20	-0.03	0.06	0.02	**0.56**	0.06	-0.03
Marketing Research	V29	0.18	0.30	0.24	**0.54**	0.12	0.05	0.05	-0.01	0.25	0.06	**0.56**	0.00	-0.04	0.02
Promotion	V30	**0.49**	0.22	**0.42**	**0.42**	0.25	0.14	-0.01	0.37	0.01	0.25	0.34	0.14	0.05	-0.06
Distribution	V31	**0.45**	0.28	0.26	**0.49**	0.26	0.18	0.03	0.30	0.09	0.05	**0.45**	0.14	0.07	0.00
Legality	V32	0.20	0.29	0.11	0.09	0.07	**0.64**	0.08	-0.05	0.02	0.09	0.02	0.00	**0.71**	0.06
Dependence	V34	**0.52**	0.22	0.11	0.01	0.14	0.20	0.07	**0.51**	0.06	-0.01	-0.09	0.05	0.11	0.06
Protection	V35	0.19	0.06	0.34	0.01	0.30	-0.03	0.06	0.12	-0.08	0.32	-0.08	0.31	-0.09	0.04
Potential Sales	V38	**0.53**	0.16	**0.58**	0.06	0.19	0.11	0.10	**0.47**	-0.10	**0.53**	-0.09	0.11	0.04	0.05
Payback Period	V39	0.20	0.31	**0.66**	0.20	0.11	0.20	0.03	-0.03	0.14	**0.64**	0.09	0.02	0.17	-0.04
Profitability	V40	0.19	0.23	**0.73**	0.12	0.15	0.07	0.00	0.02	0.07	**0.75**	0.00	0.09	0.03	-0.07

Factor Loadings for 7 Factors (Post 1989 Theoretical Structure / Post 1989 Data Empirical Results)

		Orthoganaly Rotated Factors							Oblique Rotated Factor Pattern						
Post 1989 Theorectical Structure		F1	F2	F3	F4	F5	F6	F7	F1	F2	F3	F4	F5	F6	F7
Technical Feasibility	V1	**0.65**	-0.03	0.05	0.27	0.12	-0.30	0.15	**0.65**	-0.14	0.12	0.13	0.00	-0.31	0.05
Functional Performance	V2	**0.51**	0.11	0.01	**0.43**	0.27	-0.22	0.19	0.38	0.05	-0.03	0.36	0.14	-0.22	0.10
Research & Development	V3	**0.64**	0.10	-0.03	0.22	0.21	-0.12	0.08	**0.62**	0.02	-0.03	0.09	0.12	-0.12	0.00
Technology Significance	V4	-0.04	0.15	**0.49**	-0.02	0.09	0.09	0.24	-0.06	0.04	**0.49**	-0.10	0.11	0.08	0.23
Safety	V5	0.13	0.01	0.04	0.25	0.23	0.38	0.04	0.06	-0.06	-0.05	0.26	0.23	**0.43**	0.07
Environmental Impact	V6	0.04	0.06	0.18	0.09	0.11	**0.50**	-0.07	0.03	-0.01	0.10	0.10	0.14	**0.52**	-0.02
Technology Production	V8	**0.59**	0.03	-0.07	0.17	-0.05	0.18	0.02	**0.62**	-0.06	-0.09	0.10	-0.11	0.18	0.00
Tooling Cost	V9	**0.54**	0.21	-0.04	-0.05	0.01	0.21	0.02	**0.59**	0.15	-0.07	-0.16	-0.01	0.19	-0.01
Cost of Production	V10	**0.52**	0.13	0.17	0.12	-0.14	0.32	-0.03	**0.57**	0.02	0.12	0.05	-0.18	0.28	-0.05
Need	V11	-0.01	0.26	**0.52**	0.19	0.08	0.17	-0.01	-0.09	0.19	**0.44**	0.16	0.05	0.12	-0.03
Potential Market	V12	0.03	0.00	0.39	0.20	-0.07	0.20	-0.27	0.04	-0.08	0.38	0.20	-0.10	0.16	-0.27
Trend of Demand	V13	0.07	0.16	**0.55**	0.06	0.06	0.06	-0.15	0.09	0.07	**0.56**	-0.03	0.03	0.00	-0.20
Durations of Demand	V14	-0.03	0.03	**0.49**	0.19	0.08	-0.03	-0.05	-0.06	-0.06	**0.51**	0.15	0.05	-0.06	-0.08
Demand Predictability	V15	0.03	0.10	0.26	0.10	0.27	0.26	-0.03	-0.01	0.03	0.21	0.07	0.29	0.27	-0.02
Product Line Potential	V16	-0.03	0.07	0.33	0.06	0.14	0.26	0.10	-0.06	-0.02	0.29	0.03	0.16	0.28	0.12
Societal Benefits	V18	0.02	0.02	0.20	0.10	**0.59**	0.06	0.02	-0.04	-0.05	0.21	0.01	**0.61**	0.11	0.01
Compatibility	V19	0.06	0.27	0.20	**0.42**	0.30	0.10	-0.05	-0.13	0.26	0.05	**0.43**	0.23	0.08	-0.08
Learning	V20	0.27	-0.01	-0.02	**0.47**	0.22	0.12	-0.18	0.17	-0.04	-0.10	**0.48**	0.13	0.13	-0.20
Visibility	V21	0.11	0.09	0.23	**0.52**	0.17	0.11	-0.03	-0.06	0.04	0.11	**0.56**	0.08	0.10	-0.05
Appearance	V22	0.13	0.20	0.19	**0.45**	0.01	0.13	0.10	-0.04	0.17	0.04	**0.50**	-0.09	0.10	0.08
Function	V23	0.25	0.15	0.20	**0.43**	0.00	-0.03	**0.47**	0.08	0.05	0.10	**0.43**	-0.10	-0.03	**0.45**
Durability	V24	0.21	0.02	0.19	**0.42**	-0.04	0.30	0.16	0.10	-0.08	0.08	**0.47**	-0.11	0.31	0.18
Service	V25	0.23	0.10	0.09	0.27	0.09	0.29	-0.04	0.17	0.04	0.00	0.28	0.05	0.29	-0.03
Price	V26	0.30	0.26	0.19	-0.03	0.05	0.04	0.05	0.31	0.21	0.17	-0.13	0.02	0.00	0.00
Existing Competition	V27	0.06	0.17	0.00	0.02	0.01	0.10	**0.50**	-0.03	0.13	-0.08	0.02	0.02	0.13	**0.52**
New Competition	V28	0.03	-0.04	-0.12	-0.02	0.10	-0.11	**0.44**	-0.02	-0.06	-0.12	-0.03	0.11	-0.06	**0.45**
Marketing Research	V29	0.32	0.14	0.10	0.17	0.19	0.10	-0.06	0.28	0.09	0.06	0.10	0.15	0.09	-0.09
Promotion Cost	V30	0.15	0.29	0.11	0.11	0.30	-0.05	0.09	0.05	0.29	0.04	0.03	0.26	-0.07	0.05
Distribution	V31	0.17	0.12	0.11	0.02	0.30	0.09	0.08	0.14	0.07	0.09	-0.06	0.30	0.11	0.06
Legality	V32	0.18	-0.11	-0.17	0.12	**0.42**	0.19	0.10	0.14	-0.16	-0.19	0.10	**0.44**	0.28	0.13
Development Risks	V33	**0.54**	0.26	-0.02	0.12	0.26	-0.02	0.07	**0.49**	0.21	-0.06	-0.01	0.20	-0.03	0.00
Dependence	V34	0.11	0.01	0.29	0.17	0.04	0.03	-0.09	0.10	-0.06	0.29	0.14	0.00	0.01	-0.11
Protection	V35	0.06	0.13	**0.44**	-0.09	-0.01	0.03	0.35	0.07	0.00	**0.47**	-0.18	0.01	0.01	0.34
Investment Costs	V37	**0.57**	0.39	0.10	0.03	0.17	0.15	0.04	**0.55**	0.33	0.03	-0.12	0.12	0.11	-0.02
Potential Sales	V38	0.18	**0.58**	0.34	0.24	0.01	0.08	0.05	0.01	**0.59**	0.17	0.20	-0.08	-0.02	-0.01
Payback Period	V39	0.33	**0.73**	0.11	0.07	0.08	0.07	0.08	0.18	**0.78**	-0.08	-0.02	0.00	-0.03	-0.01
Profitability	V40	0.23	**0.78**	0.22	0.16	0.02	0.07	0.17	0.03	**0.83**	-0.01	0.11	-0.08	-0.05	0.09

Factor Loadings for 7 Factors (Pre 1989 Theoretical Structure / Post 1989 Data Empirical Results)

		Orthoganaly Rotated Factors							Oblique Rotated Factor Pattern						
Pre 1989 Theoretical Structure		F1	F2	F3	F4	F5	F6	F7	F1	F2	F3	F4	F5	F6	F7
Safety	V5	0.13	0.01	0.04	0.25	0.23	0.38	0.04	0.06	-0.06	-0.05	0.26	0.23	**0.43**	0.07
Environmental Impact	V6	0.04	0.06	0.18	0.09	0.11	**0.50**	-0.07	0.03	-0.01	0.10	0.10	0.14	**0.52**	-0.02
Societal Benefits	V18	0.02	0.02	0.20	0.10	**0.59**	0.06	0.02	-0.04	-0.05	0.21	0.01	**0.61**	0.11	0.01
Legality	V32	0.18	-0.11	-0.17	0.12	**0.42**	0.19	0.10	0.14	-0.16	-0.19	0.10	**0.44**	0.28	0.13
Functional Performance	V2	**0.51**	0.11	0.01	**0.43**	0.27	-0.22	0.19	0.38	0.05	-0.03	0.36	0.14	-0.22	0.10
Research & Development	V3	**0.64**	0.10	-0.03	0.22	0.21	-0.12	0.08	**0.62**	0.02	-0.03	0.09	0.12	-0.12	0.00
Technology Production	V8	**0.59**	0.03	-0.07	0.17	-0.05	0.18	0.02	**0.62**	-0.06	-0.09	0.10	-0.11	0.18	0.00
Tooling Cost	V9	**0.54**	0.21	-0.04	-0.05	0.01	0.21	0.02	**0.59**	0.15	-0.07	-0.16	-0.01	0.19	-0.01
Cost of Production	V10	**0.52**	0.13	0.17	0.12	-0.14	0.32	-0.03	**0.57**	0.02	0.12	0.05	-0.18	0.28	-0.05
Marketing Research	V29	0.32	0.14	0.10	0.17	0.19	0.10	-0.06	0.28	0.09	0.06	0.10	0.15	0.09	-0.09
Payback Period	V39	0.33	**0.73**	0.11	0.07	0.08	0.07	0.08	0.18	**0.78**	-0.08	-0.02	0.00	-0.03	-0.01
Profitability	V40	0.23	**0.78**	0.22	0.16	0.02	0.07	0.17	0.03	**0.83**	-0.01	0.11	-0.08	-0.05	0.09
Potential Market	V12	0.03	0.00	0.39	0.20	-0.07	0.20	-0.27	0.04	-0.08	0.38	0.20	-0.10	0.16	-0.27
Trend of Demand	V13	0.07	0.16	**0.55**	0.06	0.06	0.06	-0.15	0.09	0.07	**0.56**	-0.03	0.03	0.00	-0.20
Durations of Demand	V14	-0.03	0.03	**0.49**	0.19	0.08	-0.03	-0.05	-0.06	-0.06	**0.51**	0.15	0.05	-0.06	-0.08
Demand Predictability	V15	0.03	0.10	0.26	0.10	0.27	0.26	-0.03	-0.01	0.03	0.21	0.07	0.29	0.27	-0.02
Product Line Potential	V16	-0.03	0.07	0.33	0.06	0.14	0.26	0.10	-0.06	-0.02	0.29	0.03	0.16	0.28	0.12
Potential Sales	V38	0.18	**0.58**	0.34	0.24	0.01	0.08	0.05	0.01	**0.59**	0.17	0.20	-0.08	-0.02	-0.01
Need	V11	-0.01	0.26	**0.52**	0.19	0.08	0.17	-0.01	-0.09	0.19	**0.44**	0.16	0.05	0.12	-0.03
Compatibility	V19	0.06	0.27	0.20	**0.42**	0.30	0.10	-0.05	-0.13	0.26	0.05	**0.43**	0.23	0.08	-0.08
Learning	V20	0.27	-0.01	-0.02	**0.47**	0.22	0.12	-0.18	0.17	-0.04	-0.10	**0.48**	0.13	0.13	-0.20
Visibility	V21	0.11	0.09	0.23	**0.52**	0.17	0.11	-0.03	-0.06	0.04	0.11	**0.56**	0.08	0.10	-0.05
Appearance	V22	0.13	0.20	0.19	**0.45**	0.01	0.13	0.10	-0.04	0.17	0.04	**0.50**	-0.09	0.10	0.08
Promotion Cost	V30	0.15	0.29	0.11	0.11	0.30	-0.05	0.09	0.05	0.29	0.04	0.03	0.26	-0.07	0.05
Distribution	V31	0.17	0.12	0.11	0.02	0.30	0.09	0.08	0.14	0.07	0.09	-0.06	0.30	0.11	0.06
Dependence	V34	0.11	0.01	0.29	0.17	0.04	0.03	-0.09	0.10	-0.06	0.29	0.14	0.00	0.01	-0.11
Function	V23	0.25	0.15	0.20	**0.43**	0.00	-0.03	**0.47**	0.08	0.05	0.10	**0.43**	-0.10	-0.03	**0.45**
Durability	V24	0.21	0.02	0.19	**0.42**	-0.04	0.30	0.16	0.10	-0.08	0.08	**0.47**	-0.11	0.31	0.18
Service	V25	0.23	0.10	0.09	0.27	0.09	0.29	-0.04	0.17	0.04	0.00	0.28	0.05	0.29	-0.03
Price	V26	0.30	0.26	0.19	-0.03	0.05	0.04	0.05	0.31	0.21	0.17	-0.13	0.02	0.00	0.00
Existing Competition	V27	0.06	0.17	0.00	0.02	0.01	0.10	**0.50**	-0.03	0.13	-0.08	0.02	0.02	0.13	**0.52**
New Competition	V28	0.03	-0.04	-0.12	-0.02	0.10	-0.11	**0.44**	-0.02	-0.06	-0.12	-0.03	0.11	-0.06	**0.45**
Protection	V35	0.06	0.13	**0.44**	-0.09	-0.01	0.03	0.35	0.07	0.00	**0.47**	-0.18	0.01	0.01	0.34

Factor Loadings for 5 Factors (Pre 1989 Theoretical Structure / Pre 1989 Data Empirical Results)

		Orthoganaly Rotated Factors					Oblique Rotated Factor Pattern				
Pre 1989 Theoretical Structure		F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
Safety	V5	**0.59**	**0.47**	0.10	0.05	0.03	**0.59**	0.39	-0.05	-0.10	-0.08
Environmental Impact	V6	**0.60**	**0.44**	0.00	0.18	0.17	**0.56**	0.32	-0.21	0.06	0.09
Societal Benefits	V18	**0.59**	0.34	0.17	0.05	0.16	**0.58**	0.21	0.02	-0.11	0.08
Legality	V32	0.29	**0.47**	0.06	0.06	0.07	0.21	**0.48**	-0.06	-0.05	0.01
Functional Feasibility	V2	0.28	**0.66**	0.30	0.03	0.05	0.12	**0.68**	0.20	-0.14	-0.05
Research & Development	V3	0.20	**0.79**	0.21	0.13	0.17	-0.05	**0.84**	0.06	-0.03	0.10
Production Feasibility	V7	0.27	**0.65**	0.20	0.13	0.21	0.07	**0.65**	0.04	-0.01	0.14
Stage of Development	V17	**0.43**	**0.57**	0.28	0.27	0.04	0.28	**0.47**	0.14	0.15	-0.09
Marketing Research	V29	0.16	0.30	0.26	0.13	**0.52**	-0.04	0.23	0.10	0.01	**0.53**
Investment Costs	V36	0.07	**0.44**	**0.62**	0.18	0.19	-0.19	0.37	**0.59**	0.05	0.11
Payback Period	V39	0.20	0.37	**0.64**	0.15	0.18	0.00	0.26	**0.60**	0.01	0.08
Profitability	V40	0.17	0.25	**0.72**	0.21	0.11	-0.01	0.09	**0.73**	0.10	0.00
Potential Market	V12	**0.72**	0.27	0.13	0.18	0.18	**0.74**	0.06	-0.06	0.05	0.08
Trend of Demand	V13	**0.61**	0.08	0.35	0.06	0.24	**0.65**	-0.14	0.24	-0.10	0.16
Demand Life Cycle	V14	0.34	-0.03	0.38	0.01	0.30	0.33	-0.20	0.33	-0.11	0.26
Demand Predictability	V15	**0.49**	0.19	0.13	0.25	0.32	**0.43**	0.00	-0.05	0.16	0.27
Product Line Potential	V16	**0.69**	0.18	0.19	0.18	0.05	**0.75**	-0.04	0.06	0.07	-0.08
Potential Sales	V38	**0.53**	0.19	**0.57**	0.26	0.05	**0.48**	-0.06	**0.52**	0.14	-0.10
Need	V11	**0.64**	0.15	0.30	0.12	0.24	**0.66**	-0.08	0.16	-0.02	0.15
Compatibility	V19	**0.62**	0.39	0.25	0.09	0.12	**0.61**	0.24	0.10	-0.07	0.00
Learning	V20	**0.53**	**0.53**	0.13	0.08	0.16	**0.46**	**0.46**	-0.04	-0.08	0.07
Visibility	V21	**0.52**	0.34	0.36	0.23	0.20	**0.43**	0.16	0.23	0.10	0.09
Appearnance	V22	**0.59**	0.39	0.13	0.22	0.11	**0.56**	0.24	-0.04	0.10	0.00
Promotion	V30	**0.48**	0.26	0.41	0.26	**0.44**	0.34	0.03	0.24	0.12	0.37
Distribution	V31	**0.45**	0.32	0.25	0.25	**0.51**	0.29	0.14	0.04	0.11	**0.48**
Dependence	V34	**0.54**	0.28	0.10	0.13	0.02	**0.56**	0.15	-0.02	0.02	-0.09
Function	V23	**0.49**	**0.52**	0.30	0.24	0.03	0.38	**0.40**	0.18	0.12	-0.11
Durability	V24	**0.54**	**0.52**	0.08	0.28	0.06	**0.46**	**0.41**	-0.10	0.18	-0.06
Service	V25	0.16	**0.56**	0.19	0.19	0.16	-0.05	**0.56**	0.07	0.10	0.09
Price	V26	**0.43**	0.37	0.27	0.20	0.12	0.35	0.24	0.15	0.08	0.01
Existing Competition	V27	0.12	0.23	0.36	**0.60**	0.17	-0.12	0.02	0.26	**0.63**	0.08
New Competition	V28	0.33	0.23	0.12	**0.51**	0.17	0.20	0.03	-0.04	**0.51**	0.09
Protection	V35	0.17	0.05	0.31	0.36	0.01	0.09	-0.13	0.29	0.37	-0.07

Factor Loadings for 5 Factors (Post 1989 Theoretical Structure / Pre 1989 Data Empirical Results)

Post 1989 Theoretical Structure		Orthoganaly Rotated Factors					Oblique Rotated Factor Pattern				
		F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
Functional Feasibility	V2	0.28	**0.66**	0.30	0.03	0.05	0.12	**0.68**	0.20	-0.14	-0.05
Research & Development	V3	0.20	**0.79**	0.21	0.13	0.17	-0.05	**0.84**	0.06	-0.03	0.10
Safety	V5	**0.59**	**0.47**	0.10	0.05	0.03	**0.59**	0.39	-0.05	-0.10	-0.08
Environmental Impact	V6	**0.60**	**0.44**	0.00	0.18	0.17	**0.56**	0.32	-0.21	0.06	0.09
Production Feasibility	V7	0.27	**0.65**	0.20	0.13	0.21	0.07	**0.65**	0.04	-0.01	0.14
Need	V11	**0.64**	0.15	0.30	0.12	0.24	**0.66**	-0.08	0.16	-0.02	0.15
Potential Market	V12	**0.72**	0.27	0.13	0.18	0.18	**0.74**	0.06	-0.06	0.05	0.08
Trend of Demand	V13	**0.61**	0.08	0.35	0.06	0.24	**0.65**	-0.14	0.24	-0.10	0.16
Demand Life Cycle	V14	0.34	-0.03	0.38	0.01	0.30	0.33	-0.20	0.33	-0.11	0.26
Demand Predictability	V15	**0.49**	0.19	0.13	0.25	0.32	**0.43**	0.00	-0.05	0.16	0.27
Product Line Potential	V16	**0.69**	0.18	0.19	0.18	0.05	**0.75**	-0.04	0.06	0.07	-0.08
Societal Benefits	V18	**0.59**	0.34	0.17	0.05	0.16	**0.58**	0.21	0.02	-0.11	0.08
Compatibility	V19	**0.62**	0.39	0.25	0.09	0.12	**0.61**	0.24	0.10	-0.07	0.00
Learning	V20	**0.53**	**0.53**	0.13	0.08	0.16	**0.46**	**0.46**	-0.04	-0.08	0.07
Visibility	V21	**0.52**	0.34	0.36	0.23	0.20	**0.43**	0.16	0.23	0.10	0.09
Appearnance	V22	**0.59**	0.39	0.13	0.22	0.11	**0.56**	0.24	-0.04	0.10	0.00
Function	V23	**0.49**	**0.52**	0.30	0.24	0.03	0.38	**0.40**	0.18	0.12	-0.11
Durability	V24	**0.54**	**0.52**	0.08	0.28	0.06	**0.46**	**0.41**	-0.10	0.18	-0.06
Service	V25	0.16	**0.56**	0.19	0.19	0.16	-0.05	**0.56**	0.07	0.10	0.09
Price	V26	**0.43**	0.37	0.27	0.20	0.12	0.35	0.24	0.15	0.08	0.01
Existing Competition	V27	0.12	0.23	0.36	**0.60**	0.17	-0.12	0.02	0.26	**0.63**	0.08
New Competition	V28	0.33	0.23	0.12	**0.51**	0.17	0.20	0.03	-0.04	**0.51**	0.09
Marketing Research	V29	0.16	0.30	0.26	0.13	**0.52**	-0.04	0.23	0.10	0.01	**0.53**
Promotion	V30	**0.48**	0.26	0.41	0.26	**0.44**	0.34	0.03	0.24	0.12	0.37
Distribution	V31	**0.45**	0.32	0.25	0.25	**0.51**	0.29	0.14	0.04	0.11	**0.48**
Legality	V32	0.29	**0.47**	0.06	0.06	0.07	0.21	**0.48**	-0.06	-0.05	0.01
Dependence	V34	**0.54**	0.28	0.10	0.13	0.02	**0.56**	0.15	-0.02	0.02	-0.09
Protection	V35	0.17	0.05	0.31	0.36	0.01	0.09	-0.13	0.29	0.37	-0.07
Potential Sales	V38	**0.53**	0.19	**0.57**	0.26	0.05	**0.48**	-0.06	**0.52**	0.14	-0.10
Payback Period	V39	0.20	0.37	**0.64**	0.15	0.18	0.00	0.26	**0.60**	0.01	0.08
Profitability	V40	0.17	0.25	**0.72**	0.21	0.11	-0.01	0.09	**0.73**	0.10	0.00

Appendix D

Variable Relationships: Pre 1989 to Post 1989

Post 89 Theoretical		Pre 89 Theoretical
Technical Feasibility	V1	
Functional Performance	V2	Functional Feasibility
Research and Development	V3	Research and Development
Technology Significance	V4	
Safety	V5	Safety
Environmetal Impact	V6	Environmental Impact
	V7	Production Feasibility
Technology of Production	V8	
Tooling Cost	V9	
Cost of Production	V10	
Need	V11	Need
Potential Market	V12	Potential Market
Trend of Demand	V13	Trend of Demand
Duration of Demand	V14	Demand Life Cycle
Demand Predictability	V15	Demand Predictability
Product Line Potential	V16	Product Line Potential
	V17	Stage of Development
Societal Benefits	V18	Societal Benefits
Compatibility	V19	Compatibility
Learning	V20	Learning
Visibility	V21	Visibility
Appearnance	V22	Appearnance
Function	V23	Function
Durability	V24	Durability
Service	V25	Service
Price	V26	Price
Existing Competition	V27	Existing Competition
New Competition	V28	New Competition
Marketing Research	V29	Marketing Research
Promotion Cost	V30	Promotion
Distribution	V31	Distribution
Legality	V32	Legality
Development Risks	V33	
Dependence	V34	Dependence
Protection	V35	Protection
	V36	Investment Costs
Size of Investment	V37	
Potential Sales	V38	Potential Sales
Payback Period	V39	Payback Period
Profitability	V40	Profitability

To	TIE7701 21	TRE7808 05	WAL7902 06	SET8009 09	RIE8608 28	BOU8907 06	ARC9412 20
From	HAR7609 13	LUC7702 02	GEO7808 06	HEL7901 16	PAT8009 11	LEP8608 38	CRI8907 36
Excep-tion	Evaluation of BUR7701 05 was not found		HEL7901 16 and PRE7902 03 use summary sheet D. Evaluation of KRY7901C01 was not found.	LAW7901 20 and WAL7902 06 use summary sheet C. Evaluation of KRY7901C01 was not found.			
	A	B	C	D	E	F	G
1	Legality	Legality	Legality	Legality	Legality	Legality * #30 in G	Technical Feasibility * New item
2	Safety	Safety	Safety	Safety	Safety	Safety * #5 in G	Functional Performance
3	Social Benefit * #4 in B	Environmental Impact	Environmental Impact	Environmental Impact	Environmental Impact	Environmental Impact * #6 in G	Research & Development
4	Environmental Impact * #3 in B	Societal Impact	Societal Impact	Societal Impact * Equivalent to #4 in E	Societal Benefits	Societal Benefits * #16 in G	Technology Significance * New item
5	Functional Feasibility	Functional Feasibility	Functional Feasibility	Functional Feasibility	Functional Feasibility	Functional Feasibility * Equivalent to # 2 in G	Safety
6	Production Feasibility	Production Feasibility	Production Feasibility	Production Feasibility	Production Feasibility	Production Feasibility * Replaced by #7, #8, and #9 in G	Environmental Impact
7	Stage Of Development	Stage Of Development	Stage Of Development	Stage Of Development	Stage Of Development	Stage Of Development * Dropped from G	Technology Of Production * New item
8	Investment Costs	Investment Costs	Investment Costs	Investment Costs	Investment Costs * #10 in F	Marketing Research * #27 in G	Tooling Cost * New item
From	HAR7609 13	LUC7702 02	GEO7808 06	HEL7901 16	PAT8009 11	LEP8608 38	CRI8907 36
To	TIE7701 21	TRE7808 05	WAL7902 06	SET8009 09	RIE8608 28	BOU8907 06	ARC9412 20
Excep-tion	Evaluation of BUR7701 05 was not found		HEL7901 16 and PRE7902 03 use summary sheet D. Evaluation of KRY7901C01 was not found.	LAW7901 20 and WAL7902 06 use summary sheet C. Evaluation of KRY7901C01 was not found.			
9	Profitability * #10 in B	Payback Period * New item	Payback Period	Payback Period	Payback Period * #11 in F	Research And Development * #3 in G	Cost Of Production * New item
10	Research * Replaced by #11 and #12 in B	Profitability	Profitability	Profitability	Profitability * #12 in F	Investment Costs * Dropped; Not same as # 34 in G	Need
11	Market Size * Equivalent to #13 in B	Marketing Research * New item	Marketing Research	Marketing Research	Marketing Research * #8 in F	Payback Period * #36 in G	Potential Market
12	Trend Of Demand * #15 in B	Research And Development * New item	Research And Development	Research And Development	Research And Development * #9 in F	Profitability * #37 in G	Trend Of Demand
13	Stability Of Demand * #16 in B	Potential Market	Potential Market	Potential Market	Potential Market	Potential Market * #11 in G	Duration of Demand
14	Product Life Cycle * #17 in B	Potential Sales * New item	Potential Sales	Potential Sales	Potential Sales	Potential Sales * #35 in G	Demand Predictability
15	Compatibility * #19 in B	Trend Of Demand	Trend Of Demand	Trend Of Demand	Trend Of Demand	Trend Of Demand * #12 in g	Product Line Potential
16	Learning * #20 in B	Stability Of Demand * Equivalent to #17 in C	Product Life Cycle * #17 in D	Demand Predictability	Demand Predictability	Demand Predictability * #14 in G	Societal Benefits
From	HAR7609 13	LUC7702 02	GEO7808 06	HEL7901 16	PAT8009 11	LEP8608 38	CRI8907 36
To	TIE7701 21	TRE7808 05	WAL7902 06	SET8009 09	RIE8608 28	BOU8907 06	ARC9412 20
Excep-tion	Evaluation of BUR7701 05 was not found		HEL7901 16 and PRE7902 03 use summary sheet D. Evaluation of KRY7901C01 was not found.	LAW7901 20 and WAL7902 06 use summary sheet C. Evaluation of KRY7901C01 was not found.			
17	Need * #21 in B	Product Life Cycle * #16 in C	Demand Predictability * #16 in D	Product Life Cycle * Equivalent to #17 in E	Demand Life Cycle	Demand Life Cycle * Equivalent to #13 in G	Compatibility
18	Market Dependence * #22 in B	Product Line Potential * New item	Product Line Potential	Product Line Potential	Product Line Potential	Product Line Potential * #15 in G	Learning
19	Visibility * #23 in B	Compatibility	Compatibility	Compatibility	Compatibility	Compatibility * #17 in G	Visibility
20	Promotion * #24 in B	Learning	Learning	Learning	Learning	Learning * #18 in G	Appearance
21	Distribution	Need	Need	Need	Need	Need	Function

	* #25 in B					* #10 in G	
22	Service * #26 in B	Dependence	Dependence	Dependence	Dependence	Dependence * # 32 in G	Durability
23	Durability * #29 in B	Visibility	Visibility	Visibility	Visibility	Visibility * #19 in G	Service
24	Appearance * #27 in B	Promotion	Promotion	Promotion	Promotion	Promotion * Equivalent to #28 in G	Price
25	Function * #28 in B	Distribution	Distribution	Distribution	Distribution	Distribution * #29 in G	Existing Competition
26	Cost * Dropped: Not same as # 30 in B	Service	Service	Service	Service * #29 in F	Appearance * #20 in G	New Competition
From	HAR7609 13	LUC7702 02	GEO7808 06	HEL7901 16	PAT8009 11	LEP8608 38	CRI8907 36
To	THE7701 21	TRE7808 05	WAL7902 06	SET8009 09	RIE8608 28	BOU8907 06	ARC9412 20
Excep-tion	Evaluation of BUR7701 05 was not found		HEL7901 16 and PRE7902 03 use summary sheet D. Evaluation of KRY7901C01 was not found.	LAW7901 20 and WAL7902 06 use summary sheet C. Evaluation of KRY7901C01 was not found.			
27	Existing Competition * #31 in B	Appearance	Appearance	Appearance	Appearance * #26 in F	Function * #21 in G	Marketing Research
28	New Competition * #32 in B	Function	Function	Function	Function * #27 in F	Durability * #22 in G	Promotion Cost
29	Protection * #33 in B	Durability	Durability	Durability	Durability * #28 in F	Service * #23 in G	Distribution
30		Price * New item	Price	Price	Price	Price * #24 in G	Legality
31		Existing Competition	Existing Competition	Existing Competition	Existing Competition	Existing Competition * #25 in G	Development Risks * New item
32		New Competition	New Competition	New Competition	New Competition	New Competition * #26 in G	Dependence
33		Protection	Protection	Protection	Protection	Protection	Protection
34							Size of Investment * New item
35							Potential Sales
36							Payback Period
37							Profitability

Appendix E

Pre 1989 and Post 1989 Together

Post 1989 Theorectical Structure

Factor	Variable		F1	F2	F3	F4	F5	F6	F7
			Oblique Rotated Factor Pattern						
	Technical Feasibility	V1	0.65	-0.14	0.12	0.13	0.00	-0.31	0.05
	Functional Performance	V2	0.38	0.05	-0.03	0.36	0.14	-0.22	0.10
Technical	Research & Development	V3	0.62	0.02	-0.03	0.09	0.12	-0.12	0.00
	Technology Significance	V4	-0.06	0.04	0.49	-0.10	0.11	0.08	0.23
	Safety	V5	0.06	-0.06	-0.05	0.26	0.23	0.43	0.07
	Environmental Impact	V6	0.03	-0.01	0.10	0.10	0.14	0.52	-0.02
	Technology Production	V8	0.62	-0.06	-0.09	0.10	-0.11	0.18	0.00
Production	Tooling Cost	V9	0.59	0.15	-0.07	-0.16	-0.01	0.19	-0.01
	Cost of Production	V10	0.57	0.02	0.12	0.05	-0.18	0.28	-0.05
	Need	V11	-0.09	0.19	0.44	0.16	0.05	0.12	-0.03
	Potential Market	V12	0.04	-0.08	0.38	0.20	-0.10	0.16	-0.27
Market	Trend of Demand	V13	0.09	0.07	0.56	-0.03	0.03	0.00	-0.20
Demand	Durations of Demand	V14	-0.06	-0.06	0.51	0.15	0.05	-0.06	-0.08
	Demand Predictability	V15	-0.01	0.03	0.21	0.07	0.29	0.27	-0.02
	Product Line Potential	V16	-0.06	-0.02	0.29	0.03	0.16	0.28	0.12
	Societal Benefits	V18	-0.04	-0.05	0.21	0.01	0.61	0.11	0.01
	Compatibility	V19	-0.13	0.26	0.05	0.43	0.23	0.08	-0.08
	Learning	V20	0.17	-0.04	-0.10	0.48	0.13	0.13	-0.20
Market	Visibility	V21	-0.06	0.04	0.11	0.56	0.08	0.10	-0.05
Acceptability	Appearance	V22	-0.04	0.17	0.04	0.50	-0.09	0.10	0.08
	Function	V23	0.08	0.05	0.10	0.43	-0.10	-0.03	0.45
	Durability	V24	0.10	-0.08	0.08	0.47	-0.11	0.31	0.18
	Service	V25	0.17	0.04	0.00	0.28	0.05	0.29	-0.03
Market	Price	V26	0.31	0.21	0.17	-0.13	0.02	0.00	0.00
Competition	Existing Competition	V27	-0.03	0.13	-0.08	0.02	0.02	0.13	0.52
	New Competition	V28	-0.02	-0.06	-0.12	-0.03	0.11	-0.06	0.45
Market	Marketing Research	V29	0.28	0.09	0.06	0.10	0.15	0.09	-0.09
Effort	Promotion Cost	V30	0.05	0.29	0.04	0.03	0.26	-0.07	0.05
	Distribution	V31	0.14	0.07	0.09	-0.06	0.30	0.11	0.06
	Legality	V32	0.14	-0.16	-0.19	0.10	0.44	0.28	0.13
	Development Risks	V33	0.49	0.21	-0.06	-0.01	0.20	-0.03	0.00
	Dependence	V34	0.10	-0.06	0.29	0.14	0.00	0.01	-0.11
Risk	Protection	V35	0.07	0.00	0.47	-0.18	0.01	0.01	0.34
	Investment Costs	V37	0.55	0.33	0.03	-0.12	0.12	0.11	-0.02
	Potential Sales	V38	0.01	0.59	0.17	0.20	-0.08	-0.02	-0.01
	Payback Period	V39	0.18	0.78	-0.08	-0.02	0.00	-0.03	-0.01
	Profitability	V40	0.03	0.83	-0.01	0.11	-0.08	-0.05	0.09

Pre 1989 Theoretical Structure

Factor	Variable		F1	F2	F3	F4	F5	F6	F7
			Oblique Rotated Factor Pattern						
	Safety	V5	0.06	-0.06	-0.05	0.26	0.23	0.43	0.07
Societal	Environmental Impact	V6	0.03	-0.01	0.10	0.10	0.14	0.52	-0.02
	Societal Benefits	V18	-0.04	-0.05	0.21	0.01	0.61	0.11	0.01
	Legality	V32	0.14	-0.16	-0.19	0.10	0.44	0.28	0.13
	Functional Performance	V2	0.38	0.05	-0.03	0.36	0.14	-0.22	0.10
	Research & Development	V3	0.62	0.02	-0.03	0.09	0.12	-0.12	0.00
	Technology Production	V8	0.62	-0.06	-0.09	0.10	-0.11	0.18	0.00
Business	Tooling Cost	V9	0.59	0.15	-0.07	-0.16	-0.01	0.19	-0.01
Risk	Cost of Production	V10	0.57	0.02	0.12	0.05	-0.18	0.28	-0.05
	Marketing Research	V29	0.28	0.09	0.06	0.10	0.15	0.09	-0.09
	Payback Period	V39	0.18	0.78	-0.08	-0.02	0.00	-0.03	-0.01
	Profitability	V40	0.03	0.83	-0.01	0.11	-0.08	-0.05	0.09
	Potential Market	V12	0.04	-0.08	0.38	0.20	-0.10	0.16	-0.27
	Trend of Demand	V13	0.09	0.07	0.56	-0.03	0.03	0.00	-0.20
Demand	Durations of Demand	V14	-0.06	-0.06	0.51	0.15	0.05	-0.06	-0.08
Analysis	Demand Predictability	V15	-0.01	0.03	0.21	0.07	0.29	0.27	-0.02
	Product Line Potential	V16	-0.06	-0.02	0.29	0.03	0.16	0.28	0.12
	Potential Sales	V38	0.01	0.59	0.17	0.20	-0.08	-0.02	-0.01
	Need	V11	-0.09	0.19	0.44	0.16	0.05	0.12	-0.03
	Compatibility	V19	-0.13	0.26	0.05	0.43	0.23	0.08	-0.08
	Learning	V20	0.17	-0.04	-0.10	0.48	0.13	0.13	-0.20
Market	Visibility	V21	-0.06	0.04	0.11	0.56	0.08	0.10	-0.05
Acceptance	Appearance	V22	-0.04	0.17	0.04	0.50	-0.09	0.10	0.08
	Promotion Cost	V30	0.05	0.29	0.04	0.03	0.26	-0.07	0.05
	Distribution	V31	0.14	0.07	0.09	-0.06	0.30	0.11	0.06
	Dependence	V34	0.10	-0.06	0.29	0.14	0.00	0.01	-0.11
	Function	V23	0.08	0.05	0.10	0.43	-0.10	-0.03	0.45
	Durability	V24	0.10	-0.08	0.08	0.47	-0.11	0.31	0.18
	Service	V25	0.17	0.04	0.00	0.28	0.05	0.29	-0.03
Competitive	Price	V26	0.31	0.21	0.17	-0.13	0.02	0.00	0.00
	Existing Competition	V27	-0.03	0.13	-0.08	0.02	0.02	0.13	0.52
	New Competition	V28	-0.02	-0.06	-0.12	-0.03	0.11	-0.06	0.45
	Protection	V35	0.07	0.00	0.47	-0.18	0.01	0.01	0.34

Appendix F
Additional Tables for Chapter 7

Oblique (Promax Rotation) Factor Correlations 10 Factors

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9	Factor10
Factor1	1.00	0.38	0.36	<u>0.42</u>	0.01	0.03	0.37	-0.03	0.22	0.27
Factor2	0.38	1.00	0.39	0.28	0.31	0.29	0.32	0.36	-0.08	0.21
Factor3	0.36	0.39	1.00	0.35	<u>0.48</u>	0.15	<u>0.46</u>	0.15	-0.03	0.19
Factor4	<u>0.42</u>	0.28	0.35	1.00	-0.01	-0.07	0.18	0.26	0.03	0.09
Factor5	0.01	0.31	<u>0.48</u>	-0.01	1.00	0.26	0.19	0.14	-0.06	0.00
Factor6	0.03	0.29	0.15	-0.07	0.26	1.00	0.23	0.21	-0.19	0.01
Factor7	0.37	0.32	<u>0.46</u>	0.18	0.19	0.23	1.00	0.12	-0.10	0.30
Factor8	-0.03	0.36	0.15	0.26	0.14	0.21	0.12	1.00	-0.21	0.07
Factor9	0.22	-0.08	-0.03	0.03	-0.06	-0.19	-0.10	-0.21	1.00	-0.01
Factor10	0.27	0.21	0.19	0.09	0.00	0.01	0.30	0.07	-0.01	1.00

Ability to Predict Success for the 7 and 10 Factor Models

	7 Factors Varimax Rotation		10 Factors Varimax Rotation		7 Factors Promax Rotation		10 Factors Promax Rotation	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Correctly Predicts Success	47	0.73	47	0.73	48	0.75	45	0.70
Type I Error Actual success but model predicts failure	17	0.27	17	0.27	16	0.25	19	0.30
Correctly Predicts Failure	380	0.74	372	0.72	383	0.74	390	0.75
Type II Error Actual failure but model predicts success	137	0.26	145	0.28	134	0.26	127	0.25
Overall Predictive Ability		0.74		0.72		0.74		0.75

Ability to Predict Success for the 7 and 10 Factor Models (Split Sample Test)

	7 Factors Varimax Rotation		10 Factors Varimax Rotation		7 Factors Promax Rotation		10 Factors Promax Rotation	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Correctly Predicts Success	13	0.72	12	0.67	13	0.72	14	0.78
Type I Error Actual success but model predicts failure	5	0.28	6	0.33	5	0.28	4	0.22
Correctly Predicts Failure	118	0.77	110	0.71	118	0.77	116	0.75
Type II Error Actual failure but model predicts success	36	0.23	44	0.29	36	0.23	38	0.25
Overall Predictive Ability		0.76		0.71		0.76		0.76

Bibliography

Abratt, R. and A. A. Lombard (1993): "Determinants of Product Innovation in Specialty Chemical Companies," Industrial Marketing Management, Vol. 22, pp. 169-175.

Albaum, G. (1975): The Independent Inventor, Eugene, OR: University of Oregon.

Agresti, A. (1990): Categorical Data Analysis, New York: John Wiley.

Åstebro, T. (1998): "Basic Statistics on the Success Rate and Profits for Independent Inventors," Entrepreneurship Theory and Practice, 23, pp. 41-48.

Åstebro, T. (1999): "The Economics of Invention for Independent Inventors", manuscript, University of Waterloo.

Åstebro T. and I. Bernhardt (1999) "The Social Rate of Return to Canada's Inventor's Assistance Program," The Engineering Economist, XX, pp. XX-XX.

Åstebro T. and G. Sampson (1999): "The Statistical Accuracy of a Process to Evaluate, Ex Ante, the Ex Post Commercial Success of Inventions," in Robinson, P. (ed.) Entrepreneurship: Peaks and Valleys, Proceedings, 20th Canadian Conference for Small Business and Entrepreneurship.

Åstebro T. and Y. Gerchak (forthcoming): "Profitable Advice: The Value of Information Provided by Canada's Inventor's Assistance Program," Economics of Innovation and New Technology.

Åstebro T., J. Michela, and X. Zhang (2001): "The Survival of Innovations: Patterns and Predictors," manuscript, University of Waterloo.

Balachandra, R. and J. Friar (1997): "Factors for Success in R&D Projects and New Product Innovation: A Contextual Framework," IEEE Transactions on Engineering Management, 44, pp. 276-287.

Cattell, R.B. (1952): Factor Analysis, New York: Harper and Brothers.

Cattell, R.B. and K. Dickman (1962): "A Dynamic Model of Physical Influences Demonstrating the Necessity of Oblique Simple Structure," Psychological Bulletin, 59, pp. 389–400.

Cattell, R.B. and W. Sullivan (1962): "The Scientific Nature of Factors: A Demonstration by Cups of Coffee," Behavioral Science, 7, pp. 184-193.

Cattell, R.B. and R.L. Gorsuch (1963): "The Uniqueness and Significance of Simple Structure Demonstrated by Contrasting 'Natural Structure' and 'Random Structure' Data," Psychometrika, 28, pp. 55-67.

Cattell, R.B. (1966): "The Scree Test for the Number of Factors," Multivariate Behavioral Research, 1, pp. 245–276.

Cattell, R.B. and S. Vogelman (1977): "A Comprehensive Trial of the Scree and KG Criteria for Determining the Number of Factors," Multivariate Behavioral Research, 12, pp. 289–325.

Chrisman, J. J., Nelson, R. R., Hoy, F., and R. B. Robinson (1985): "The Impact of SBCD Consulting Activities," Journal of Small Business Management, Vol. 23, pp. 1-11.

CIC (1996): Annual Report, Canadian Industrial Innovation Centre, Waterloo, Canada.

Cliff, N. and C.D. Hamburger (1967): "The study of sampling errors in factor analysis by means of artificial experiments," Psychological Bulletin, 68, pp. 430-445.

Cliff, N. (1970): "The relation between sample size and population characteristic vectors," Psychometrika, 35, pp. 163-178.

Cliff, N. (1987): Analyzing Multivariate Data, San Diego: Harcourt Brace Jovanovich.

Constandse, W. J. (1971): "Why New Product Management Fails," Business Management, June, pp. 16-19.

Cooper, R. G. (1979): "The Dimensions of Industrial New Product Success and Failure," Journal of Marketing, 43, pp. 93-103.

Cooper, R. G. (1981): "An Empirically Derived New Product Project Selection Model," IEEE Transactions on Engineering Management, 28, pp. 54-61.

Cooper, R. G. (1983): "A Process Model for Industrial New Product Development," IEEE Transactions on Engineering Management, 30, pp. 2-11.

Cooper, R. G. (1984): "How New Product Strategies Impact on Performance," Journal of Product Innovation Management, 1, pp. 5-18.

Cooper, R. G. (1985): "Overall Corporate Strategies for New Product Programs," Industrial Marketing Management, 114, pp. 179-193.

Cooper R.G. (1985): "Selecting Winning New Product Projects: Using the NewProd System," Journal of Product Innovation Management, 2, pp. 34-45.

Cooper, R. G. and E. J. Kleinschmidt (1987): "New Products: What Separates Winners from Losers," Journal of Product Innovation Management, 4, pp.169-184.

Cooper R.G. and E.J. Kleinschmidt (1987): "What Makes a New Product a Winner: Success Factors at The Project Level," R&D Management, 17, pp. 175-187.

Cooper R.G. and E.J. Kleinschmidt (1987): "Success Factors in Product Innovation," Industrial Marketing Management, 16, pp. 215-224.

Cooper R.G. (1988): "Resource Allocation in the New Product Process," Industrial Marketing Management, 3, pp. 249-253.

Cooper R.G. (1992): "The NewProd System: The Industry Experience," Journal of Product Innovation Management, 9, pp. 113-128.

Cooper R.G., E.J. Kleinschmidt and J. Elko (1993): "Screening new products for potential winners," Long Range Planning, 26, pp. 74-82.

Cooper R.G. (1994): "Debunking the myths of new product development," Research Technology Management, 37, pp. 40-56.

Cooper R.G. and E.J. Kleinschmidt (2000): "New Product Performance: What distinguishes the star products," Australian Journal of Management, 25, pp. 17-45.

Cooper R.G. (2000): "Product innovation and technology strategy," Research Technology Management, 43, pp. 38-41.

Dahlin, K., M. Taylor and M. Fichman (1999): "Innovative Content as a Predictor of Success for Individual Inventors," manuscript, Carnegie Mellon University.

Davidson J.M., A. Clamen and R.A. Karol (1999): "Learning from the best New Product Developers," Research Technology Management, 42, pp. 12-18.

Dielman, T.E., R.B. Cattell, and R.K. Wagner (1972): "Evidence on the Simple Structure and Factor Invariance Achieved by Five Rational Methods on Four Types of Data," Multivariate Behavioral Research, 7, pp. 223-231.

Dillman, D. (1978): Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Evans, D. and B. Jovanovic (1989): "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints," Journal of Political Economy, pp. 802-27.

Gorsuch, R.L. (1974): Factor Analysis, Philadelphia: W.B. Saunders Co.

Griffin, A. (1997): "PDMA Research on New Product Development Practices: Updating Trends and Benchmarking best Practices," Journal of Product Innovation Management, Vol. 14, pp. 429-458.

Hopkins, D. S. (1981): "New Product Winners and Losers," Research Management, May, pp. 12-17.

Horn, J.L. and R. Engstrom (1979): "Cattell's Scree Test in Relation to Bartlett's Chi-Square Test and Other Observations on the Number of Factors Problem," Multivariate Behavioral Research, 14, pp. 283–300.

Jarmin, R. S. (1999): "Evaluating the Impact of Manufacturing Extension on Productivity Growth," Journal of Policy Analysis and Management, Vol. 18, No. 1, pp. 871-891.

Kaiser, H.F. (1970): "A Second Generation Little Jiffy," Psychometrika, 35, pp. 401-415.

Kaiser, H.F., and J. Rice (1974): "Little Jiffy Mark IV," Educational and Psychological Measurement, 34, pp. 111-117.

Kim, J.O. and C.W. Mueller (1978): Introduction to Factor Analysis: What It Is and How To Do It, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-013, Beverly Hills: Sage Publications.

Lavrakas, P.J. (1993): Telephone Survey Methods, 2nd ed., Newbury Park: SAGE Publications.

Lazo, H. (1965): "Finding a Key to Success in New Product Failures," Industrial Marketing Management, Vol. 50, pp. 74-77.

Lilien, G. and E. Yoon (1989): "Determinants of New Industrial Product Performance: A Strategic Reexamination of the Empirical Literature," IEEE Transactions on Engineering Management, 36, pp. 3-10.

Linn, R.L. (1968): "A Monte Carlo approach to the number of factors problem," Psychometrika, 33, pp. 37-72.

Maddala, G. S. (1983): Limited Dependent and Qualitative Variables in Econometrics. Cambridge, UK: Cambridge University Press.

Maidique, M. M. and B. J. Zirger (1984): A Study of the Success and Failure in Product Innovation: The Case of the U.S. Electronics Industry," IEEE Transactions on Engineering Management, Vol. 31, pp. 192-203.

Maidique, M. A. and B. J. Zirger (1985): "The New Product Learning Cycle," Research Report Series, Innovation & Entrepreneurship Institute, School of Business Administration, University of Miami, Coral Gables, FL, February 1985, pp. 85-101.

Mansfield, E. (1968): Industrial Research and Technological Innovation - An Econometric Analysis. New York: W. W. Norton.

McGrath RG. and I.C. MacMillan (2000): "Assessing Technology Projects Using Real Options Reasoning," Research Technology Management, 43, pp. 35-49.

Murphy S.A and V. Kumar (1997): "The front end of new product development: a Canadian Survey," R&D Management, 27, pp. 5-15.

New York Times (1994): "The Canadian Triangle Where High Tech Reins," April 13, Section C1, page C1 and C6, 1994.

Nunnally, J. (1978): Psychometric Theory, 2nd Edition, New York, McGraw-Hill.

Parker, R. G. Udell and L. Blades (1996): "The New Independent Inventor: Implications for Corporate Policy," Review of Business, Spring 1996, pp. 7-11, and page 34.

Perlitz M., P. Thorsten and R. Schrank (1999): "Real options valuation: the new frontier in R&D project evaluation?" R&D Management, 29, pp. 255-269.

Robinson, R. B. (1982): "The Importance of 'Outsiders' in Small Firm Strategic Planning," Academy of Management Journal, Vol. 25, pp. 80-93.

Rogers, E. (1995): The Diffusion of Innovation, 5th ed., Chicago, IL: The Free Press.

Rosenthal, R. and R. Rosnow (1991): Essentials of Behavioral Research: Methods and Data Analysis, 2nd edition. New York: McGraw Hill.

Rummel, R.J. (1970): "Understanding Factor Analysis," The Journal of Conflict Resolution, 11, pp. 444-479.

Souder, W. E. (1988): "Managing Relations between R&D and Marketing in New Product Development Projects" Journal of Product Innovation Management, Vol. 5 (1), pp. 6-19.

Stevens G. A. and J. Burley (1997): "3,000 Raw Ideas = 1 Commercial Success, " Research Technology Management, 40, pp. 16-27.

Stewart, D.W. (1981): "The Application and Misapplication of Factor Analysis in Marketing Research," Journal of Marketing Research, 18, pp. 51–62.

Thurstone, L.L. (1947): Multiple-Factor Analysis: a Development and Expansion of the Vectors of the Mind, Chicago, IL: The University of Chicago Press.

Tucker, L.B., R.F. Koopman, and R.L. Linn (1969): "Evaluation of factor analytic research procedures by means of simulated correlation matrices," Psychometrika, 34, pp. 421-459.

Udell, G. (1989): "Invention Evaluation Services: A Review of the State of the Art," Journal of Product Innovation Management, 6, pp. 157-68.

Udell, G., R. Bottin and D. Glass (1993): "The Wal-Mart Innovation Network: An experiment in Stimulating American Innovation," Journal of Product Innovation Management, 10, pp. 23-34.

USPTO, (1997): All Technologies Report, January 1963 -- December 1996, U.S. Patent and Trademark Office, Washington D.C., March 1997.

Voss, C. A. (1985): "Determinants of Success in the Development of Application Software," Journal of Product Innovation Management, Vol. 2, pp 122-129.

Wall Street Journal (1991): "More Inventors Try their Luck as Aid Increase," August 19, Section B1, 1991.

Yap, C. M. and W. E. Souder (1994): "Factors Influencing New Product Success and Failure in Small Entrepreneurial High-Technology Electronics Firms," Journal of Product Innovation Management, Vol. 11, pp. 418-432.

Yoon, E. and G. L. Lilien (1985): "New Industrial Product Performance, The Impact of Market Characteristics and Strategy," Journal of Product Innovation Management, Vol. 3, pp. 134-144.

Zoski, K.W. and S. Jurs (1996): "An objective counterpart to the visual scree test for actor analysis: The standard error scree," Educational and Psychological Measurement, 56, pp. 443-449.

Zwick, W.R., and W.F. Velicer (1982): Factors influencing four rules for determining the number of components to retain, Multivariate Behavioural Research, 17, pp. 253-269.

Zwick, W.R., and W.F. Velicer (1986): "Comparison of five rules for determining the number of components to retain," Psychological Bulletin, 99, pp. 432-442.