Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

An analysis of the factors affecting customer commitment in a New Zealand financial institution.

A thesis presented in partial fulfilment of the requirements for the Masters of

Applied Statistics

at Massey University, Albany, New Zealand

Helen Blayney

2003

Abstract

This thesis presents the results of the analysis of data collected in a postal and email survey of personal customers of the financial institution. The objective of the research is to identify various variables, which are significant in predicting commitment of a customer to their principal financial institution and to ascertain if the life stage variables contribute to the level of commitment. Two surveys to groups of personal customers a year apart provided data for analysis. The results indicate that the variables that contribute most to predicting commitment include the life stage variables. The results also point to the existence of quite different affective response rates for those customers who received an email questionnaire. No significant difference in commitment level was identified for customers common to both surveys. Although these results represent a somewhat preliminary analysis of the influence of life stage on commitment level, they do indicate that there is much to be learned about this relationship.

Table of Contents

List of tables

List of Figures

Chapter 1 Introduction

This research focuses on the analysis of factors affecting consumer commitment in the financial institution. Understanding customer commitment has emerged as the way forward for organisations in the future. Hofmeyr and Rice (2000) predict, "... in the years ahead the understanding and use of commitment will be a standard every day practice." Interest in customer commitment evolved initially from organisations' focus on customer relationship management (CRM) in the 1980's.

The next sections outline the main developments in organisations' understanding of customers over the last few decades. In particular customer relationship management, customers loyalty, commitment, satisfaction, the 'Conversion' model, profitability and life stage.

The chapter ends by describing the theoretical stance for this research and then listing the research questions that will be addressed.

1.1 The development of CRM

In the early days, businesses were small and producers knew their customers well. As industries grew, mass marketing evolved and the focus in the marketing industry was on developing distribution channels. It was at this point that producers lost contact with customers. What resulted was an abundance of services and products but a lack of good customers for them. It was not until the 1980's that marketers realised that producers ' should find products for good customers rather than good customers for the product' (Storbacka and Lehtinen, 2001). A slowing economy reduced disposable income and greater price sensitivity had made it more difficult and costly for companies to maintain existing relationships or acquire new ones. As a result of these issues, customer relationship management (CRM) was established with many companies embracing it with the hope of reversing recent trends. 'One of the most significant developments in the practice of marketing is the shift in emphasis to customer relationship management from a transaction orientation' Wyner (1999), the key word here being relationship.

Central to understanding and implementing CRM was a thorough understanding of the customer value creation process. 'The aim is not to maximise the revenue from single transactions but rather to build a lasting relationship with the customer.' (Storbacka and Lehtinen, 2001). At this point, marketing managers had to begin to consider the individual customer and the relationship between the customer and company. They needed to know the actual relationship that they had with each customer and from this there evolved a mass of research on the level of customer relationships and how to measure these relationships with emphasis on customer loyalty, commitment and satisfaction and their inter-relationships.

Loyalty has been defined as share of requirements for a particular product (Hofmeyr and Rice, 2002). A customer whose purchase of brand A is more than 67% of their total spend in a product category is considered a loyal to that brand. Day (1969) said loyalty is a combination of commitment and buyer behaviour. Satisfaction with a brand, although an essential component to commitment, is not sufficient to explain loyalty (Oliva, Oliver and McMillan, 1992).

1.2 Customer Loyalty

Much of the early research saw customer loyalty as a unidimensional issue, of the visible loyalty and omitted to assess customer's psychological feelings towards the brand – their intentions. There was also no way of knowing what proportion of the early 'brand loyalty' literature employed measures of actual brand loyalty, repeat purchasing behaviour or both. It was not until Day (1969) that loyalty was seen as a two-dimension concept, a combination of psychological feeling and actual behaviour. Previously, only the proportion-of-purchase was considered for a measure of the relationship between a customer and a firm. It soon became obvious that this definition was not enough as it was not possible to determine the duration of a customer's relationship or whether the customer was at risk of defection. Also no link could be found between loyal behaviour and demographics or effects of pricing or advertising and so on. Day criticised these early loyalty measures, as they did not 'distinguish between true or intentional loyalty and spurious loyalty associated with consistent purchasing of one brand.'

Day went on to say that spuriously loyal buyers lack any attachment to brand attributes thus they can easily be captured by another brand, which offers a better deal. Simply using socio-economic and demographic variables won't really be specific enough for describing brand loyalty. 'These descriptive variables are at best remote proxies for important individual differences in buying styles, decision process or sensitivity to promotional influences.'

Some researchers, including Ehrenberg (1988), believe that loyalty doesn't even exist - consumers are "promiscuous" as they have a set of brands to which they are loyal, so are not really loyal to one brand at all. Schultz (1998) even suggests that we don't really know what loyalty is or even how to measure it and 'Brand loyalty is even worse.'

Other researchers have linked inertia with loyalty. Papatla and Krishnamurthi (1992) believe that past behaviour as well as marketing variables affect future choice decisions. They presented a 'model of dynamic choice behaviour' which involves four characteristics, namely loyalty, short term reluctance to change brands, long term effect on inertia of repetitive consumption and, finally, choice alternatives. They established a probit model using these characteristics, which allows the tracking of, repeat purchases and also the tracking of switches quite well.

1.3 Commitment

There have been varying views on the relationship between commitment and loyalty. Day's (1969) thinking was that commitment leads to loyalty and others agree that this concept could 'provide the essential basis for distinguishing and assessing the relative degrees of brand loyalty' (Jacoby and Kyner, 1973). These authors believe that the early single unidimensional measures of brand loyalty are insufficient, as brand loyalty is a complex multidimensional phenomenon. To establish their true or representative relationship with customers, marketers need to focus on the reasons for the behaviour not just the outcome. However, this view is not universally held, as others suggest that the two constructs are either not related (Oliva et al) or that they are one and the same and represent each other (Assael, 1987). The common view is that the two are related, yet by definition are distinct, with commitment leading to loyalty (Beatty, Kahle and Homer, 1988). They comment that commitment is an 'emotional or psychological attachment to a brand,' that develops before a customer is able to determine that their repeat purchase behaviour is from a sense of loyalty.

'Everyone seems to believe that frequent purchases are equal to loyalty' (Schultz, 1998). Although there are some researchers who have different views of loyalty, the main theme in the marketing field is that loyalty can be defined as a combination of brand attitude and behaviour. Pritchard, Havitz & Howard sum up the advantage of a two-dimensional loyalty measure in that we can 'move beyond the traditional dependence on "purchase only" measures of brand loyalty to capture the strength of consumers' commitments to a particular brand,' thus enabling marketers to distinguish between those who buy strictly from habit or from convenience from those whose repeat purchase behaviour is based on genuine attachment (commitment). Wyner (1999) concludes that 'Attitudinal measures of loyalty are used to capture preferences that do not translate directly into behaviour' and 'Behavioural measures alone would not give a complete picture'. So proportion-of-purchase definitions, when used alone, are inadequate for assessing brand loyalty and to get the benefits of both attitudinal and behavioural measures, we must combine them.

A distinction can be made between committed loyalty and loyalty without commitment. It is possible to buy a brand loyally and yet not be attached to it, when there is a systematic external reason why consumers cannot get the brand they want, e.g. the preferred brand is not readily available. Behavioural loyalty without psychological commitment cannot really go on too long, because another more desirable product could potentially entice the customer to defect. It is equally possible to be committed but seldom buy a product. A customer may be committed to BMWs but never purchase one because they are outside of his/her price range.

Throughout the marketing literature, there is much debate about the best method for measuring loyalty. A common finding is that, as commitment increases, so loyalty (defined as share of requirements) increases. At an aggregate level, commitment will correlate with loyalty, but at the individual consumer level, there will always be loyal customers who are not committed. Identifying customers who display loyalty without commitment is the same as identifying high value customers at risk. Commitment will be maximised when there are a high number of perceived distinguishing attributes among brands and a high level of salience is attached to these attributes (Robertson, 1976).

Jacoby and Kyner thought that a true measure of brand loyalty was more complex than first thought and that it is actually a combination of six different conditions. The conditions they suggest can be summed up as a biased, behavioural response expressed over time, by some decision-making unit, with respect to one or more alternative brands, which is a function of psychological processes. Results of the study supported their hypothesis that brand loyalty and simple repeat purchasing behaviour had different underlying dynamics. If a customer did not satisfy all 6 conditions then they were not brand loyal. It was not really until Pritchard, Havitz & Howard (1999) researched the link between commitment and loyalty that they found 'the tendency to resist changing preference to be a key precursor to loyalty'. So understanding commitment is a prerequisite to understanding true loyalty. By studying loyalty behaviour companies know which customers to ignore (those who are not loyal) in order to avoid a waste of resources (Reichheld, 1993).

1.4 Satisfaction

Much had been written about satisfaction and dissatisfaction but little on the relationship between satisfaction and loyalty. Oliva, Oliver & McMillan (1992) found that the complexity of the link between satisfaction and behaviour had until that time been underestimated. Their research looked at the service industry, in particular an electricity supplier, but much of this research is relevant to the financial industry, as service is a large portion of the relationship experienced by the customer. Their research investigated the non-linear relationship between satisfaction and behaviour and its multi threshold characteristics. They found that when satisfaction rose above a certain threshold, customer loyalty climbed rapidly but when satisfaction was below a certain threshold, loyalty declined just as rapidly. Between the thresholds, loyalty was particularly flat.

Some researchers believe that the psychological side of loyalty is a consequence of satisfaction – but others (Schultz, 1998) found it debatable that satisfaction was enough, and Oliva, Oliver & McMillan (1992) suggest that loyalty is much more complicated than that. Oliva et al (1992) state that 'Dissatisfaction must be extreme or repetitive to dislodge a previously loyal customer under high involvement.' They go on to say that, 'neither commitment nor satisfaction is linearly related to loyalty.' They believe that a single dissatisfying transaction will not necessary change a customer from being brand loyal, as loyalty response behaviour lags behind satisfaction.

'Customer satisfaction has come to represent an important cornerstone for customer-orientated business practices across the multitude of companies operating in diverse industries.' (Szymanski and Henard, 2001). For a company to truly understand their customers they need to first concentrate on commitment with satisfaction being only one facet of a commitment measurement, before trying to assess loyalty. 'It is not sufficient for a company to merely satisfy customer needs, nor is it enough that the customer is satisfied' (Storbacka and Lehtinen, 2001). The widely used 'Conversion model' (Hofmeyr and Rice, 2001), which through four simple questions is able to identify a customer's commitment level to a particular brand, includes satisfaction as one of four components.

Reichheld (1993) said that building a highly loyal customer base cannot be done as an add on, it must be integral to a company's basic business strategy. He believed that 'designing and managing this self reinforcing system is the key to achieving outstanding customer loyalty'. He goes on to say that as customer loyalty and market share increases, the cost of acquiring and servicing customers goes down. However others researchers do not agree with this (Reinartz and Kumar, 2002) as research has found that loyal customers may cost more as they often expect something in return for their loyalty.

A customers' socio-economic and demographic grouping have been consistently shown to influence levels of satisfaction, irrespective of the product that customers are evaluating (Bryant and Cha, 1996). The American Customer Satisfaction Index (ACSI) has shown that 'significant differences based on sex, age, socio-economic status, and metro or non-metro residence.' The ACSI was started in 1994 and is a national, cross industry, cross company measure of satisfaction. It measures 7 economic sectors, 38 industries (including e-commerce and e-business) and more than 200 companies and federal or local government agencies. In addition to the company-level satisfaction scores, the ACSI produces scores for the causes and consequences of customer satisfaction and their relationships. (American Customer Satisfaction Index website, 2003).

The satisfaction component of commitment has been split further to include strength of relationship (Barnes, 1997). He measured strength of a customer's relationship using three variables: percentage of their business with the financial institution, their intention to continue the relationship and whether they are prepared to recommend the financial institution to others. He said that relationships with customers could not be built on behaviour alone, as this does not consider their feelings of closeness or attachment.

To summarise as even highly satisfied customers defect and buy other brands, while those who say they are not satisfied, continue to purchase. So it appears that satisfaction alone is not enough to influence consumer behaviour.

1.5 The Conversion model

Commitment has for some time been seen as a very important aspect of buyer behaviour and much work has gone into being able to measure a customer's commitment accurately. Out of all this research a very successful and widely used model is the 'Conversion Model' (Hofmeyr and Rice, 2000). The Conversion model uses four dimensions to measure commitment.

1. satisfaction with the brand
2. perception of the alternatives
3. the importance of brand choice
4. degree of ambivalence

For customers to be classified as strongly committed, brand choice has to matter a great deal; they have to be convinced that the brand is significantly better than any competitive brand and there should be no ambivalence. Users of the brand are segmented into four segments and an allocation algorithm is used to classify people into the appropriate segment. The segments are defined as follows:

1. Entrenched: users unlikely to switch brands in the foreseeable future
2. Average: users unlikely to change in short term, but with some possibility of change in the medium term.
3. Shallow: users who have a lower commitment than average with some of the consumers already actively considering alternatives.
4. Convertible: Users who are most likely to defect.

Committed customers are 'entrenched' and 'average', while uncommitted customers are 'shallow' and 'convertible'. Commitment is possible even when people are dissatisfied as long as the relationship is something that they care about. This is because people tend to try to fix an important relationship, which is why satisfaction does not correlate with defection but commitment levels do (Hofmeyr and Rice, 2000). Uncommitted customers are a consequence of one of three instances - they don't care which product or brand they use; they are so dissatisfied that it doesn't matter how much they do care or, thirdly, something else appeals to them causing ambivalence. Satisfaction measuring systems are simply not designed to provide insight into how many customers stay loyal to the company for how long (Reichheld, 1993). Instead the commitment level of the customer needs to be established. Managing committed and uncommitted customers requires totally different strategies. Committed customers need reinforcement of their beliefs about the brand and maintaining the desires that drive the brand – this is achieved with the use of advertising. Uncommitted, uncaring customers require a strengthening of the brands market presence.

Hofmeyr and Rice (2000) found consistently over 2800 studies, that as commitment increases, loyalty (defined as share of requirements) increases. "No matter what the brand, product category or country, when commitment is measured as we measure it, the more committed a consumer is to a brand, the bigger the 'share' of that consumer's 'requirements' the brand will fill." (Hofmeyr and Rice, 2000).

They also link commitment level and defection, "... what we can say with confidence is that, at every level of commitment, a person identified as more committed will have a significantly lower likelihood of defection than someone identified as less committed." There is a moderating factor though – the bigger brands retain their uncommitted customers more easily than smaller brands. In other words the rate at which uncommitted customers defect is a function of brand size. A possible explanation of this occurrence is that there are constant market place incentives to buy a large brand. So if a customer becomes uncommitted, there is still plenty of market pressure favouring that brand. Hofmeyr and Rice say, "... the relationships that people have with brands are not static," indicating that an uncommitted customer can recover.

1.6 Profitability

Once we have established the commitment level and behaviour of a customer it is essential to assess their profitability, as being loyal does not necessarily mean that the customer is profitable. Those customers who have been with the financial institution for several years and only have a transaction account are rarely profitable. Almquist, Heaton and Hall (2002) state that companies need to improve customer relationships that enhance the long-term flow of profits to the company and those relationships that do this are satisfaction and commitment and hopefully loyalty. Understanding customer's behaviour, needs and purchase occasions creates value opportunities. They recommend that the goal should not be to increase customer loyalty across the board but rather to acquire, retain and develop the most valuable customers.

'The best customers, we're told, are the loyal ones' states Werner (2002), he goes on to say that 'What we've found is that the relationship between loyalty and profitability is much weaker – and subtler – than the proponents of loyalty programs claim.' Also 'Instead of measuring loyalty alone, companies will have to find ways to measure the relationship between loyalty and profitability so that they can better identify which customers to focus on and which to ignore. We believe that no company should ever take for granted the idea that managing customers for loyalty is the same as managing them for profits'. Reichheld (1993) stresses that firms need to attract the right customers – not those easiest to attract or most profitable in short term, but those who are likely to do business with the company over time, in other words those who are committed. Dowling and Uncles (1997) caution that 'In short, the contention that loyal customers are always more profitable is a gross oversimplification'. So a key use of studying commitment is the need to be able to identify those people whose long-term loyalty can be developed.

One definition of a loyal customer is a customer who maintains a long-term relationship with a company and some authors (Morgan and Hunt, 1994: Sheth and Parvatiyar, 1995) believe that this long-term relationship in it self will result in a profitable relationship. However, Reinartz and Kumar (2002) found that this is not necessarily the case due to normal fluctuations in customer activity over time. They go on to say that companies "... may learn more from the seasonal and life stage variations in customer buying patterns than from the specific number of years the customer has been on the database." It seems that life stage could also be an important indicator of customer value, as Reinartz and Kumar point out "What appears on the surface to be dormancy may actually be a naturally occurring pattern that will trigger purchasing when the next cycle comes around."

1.7 Life stage

Existing research seems to suggest that customer value can be assessed and may change with the transition from one life stage to the next. The Council on Financial Competition (2000) highlighted an example of a large North American the financial institution, which segments customers by a combination of behaviour, life stage and demographics. In order to maximise lifetime profitability as customers migrate from one-stage (and set of needs) to another, the financial institution develops products and systems that not only meet customers' current needs profitably but also help them in the transition from one life stage to the next. Apparently the financial institution has had much success with this segmentation policy, it uses five life stage segments based on attitudinal and behavioural characteristics, current and potential profitability, expected purchasing behaviour, vulnerability to defect and channel preferences.

The segments are:

1. youth
2. early adulthood (18-30 very transient, high job turnover)
3. borrowers
4. asset gathers
5. asset preservers

'The objective is to identify prospects before or as their needs change in order to approach them with compelling offers when they're the most receptive.' By using life stage as a method of segmentation there is a risk to carry, as the financial institution will loose money on some customers, especially younger customers who do not have the balances to offset acquisition costs. However this should be offset by future potential profitability of customers.

There have been mixed results with regard to the impact of life stages on a consumers values and satisfaction. Callingham and Baker (2002) examined whether a person's circumstances (life stage) rather than personality predicted behaviour. They went on to say that it is circumstance that predicates behaviour that in turn predicates thinking. The authors found that in reality people are too diverse to be categorized so simply using life stage. They found that demographics have done well in describing differences in the data and that the other classification systems, such as life stage, have not added much.

Lansing and Morgan (1955) on the other hand described how income, expenditures on durable goods, assets and debts and subjective feelings about their financial position differ at different life cycle stages; while Wells and Gubar (1966) found that life stages are related to important changes in purchasing behaviour. The changes in income and purchasing at different life stages occur with 'the changes in satisfaction with one's financial position that occur from stage to stage in the life cycle' (Andreasen, 1984). An interesting finding from Andreasen's study was that the demands placed on income by durable goods purchases are very much related to life cycle position. An increase in durable goods, especially expensive ones (boats, cars and so on) could also signal a need for loans. 'Lifecycle would, therefore, seem to be a more sensitive indicator of the family's financial situation than would chronological age'.

In today's increasingly fragmented society, expectations of behaviour by certain age groups are not clear-cut as people marry and remarry. Segmenting by life stage, allows companies to offer customers what they need, taking into account their ability and willingness to pay.

1.8 Theoretical stance for this Research

In this research the position was taken that life stage contributes to and significantly influences the level of customer commitment and ultimately loyalty in the financial industry. There appears to be a need to distinguish between matters relating to commitment, loyalty and life stage events. The relationships between life stage and commitment must be established if marketing managers are to effectively meet customers' existing and potential needs. So far loyalty has been seen as a function of commitment and predicted behaviour but it is not as simple as that because external forces would also affect the loyalty of a customer. It is proposed that life stage not only affects commitment but loyalty too.

1.9 Research

This research examines whether life stage is significant in regard to commitment level and if any particular life stage has more stable commitment levels than others.
This analysis is performed for a rather special research population, namely the personal customer segment of the financial institution. These customers are managed by personal financial advisors or have a household income greater than \$100,000 and footings of between \$125,000 and \$500,000 (footings are the absolute value of the summed assets and liabilities). Questionnaires designed to measure the commitment level of customers, using of a number of indirect questions, were sent to a sample of personal customers in 2002 and 2003.

Chapter 2 explains the analysis performed in this research, which includes SEM, factor analysis, data mining techniques, chi-square tests t-tests and analysis of variance. Chapter 3 details the results from this analysis. Chapter 4 outlines the conclusions that come out of this research and recommendations for future work. The appendices contain more detailed information on variables used in the analysis.

The specific research questions for this research are as follows:

- Test whether the 2002 and 2003 questionnaire data fit the Conversion model of Hofmeyr and Rice (2000) using structural equation modelling (SEM)?
- Can the Conversion model be improved upon? Test an extension of the Conversion model, based on research by Barnes (1997) on strength of relationship and the link between inertia and loyalty (Papatla and Krishnamurthi, 1992), again using SEM for 2002 and 2003 customers.
- Create commitment levels for the 2002 and 2003 respondents using a combination of the 'best' structural equation model and factor analysis.
- Can we predict these commitment levels using only the information from the financial institution's database? Is the prediction model for 2002 commitment also valid for 2003 customers?
- Determine whether demographic variables are significant in predicting commitment levels.
- Determine whether Hofmeyr and Rice's theories apply to the financial institution's personal customers, namely commitment level is significant to product uptake and defection. We also test the theory that commitment and profitability, and therefore revenue, are related. (Werner, 2002; Dowling and Uncles, 1997).
- There were two methods of delivery used for the 2003 questionnaire, namely postal and email. Was the type of delivery influential on commitment level? Which method had the highest commitment level?

- There were 223 customers common to both the 2002 and 2003 surveys. Was there a significant change in commitment level for the common customers? Is there any significant difference between those common customers who decreased commitment in 2003 from those who did not?

Chapter 2 Methodology

2.1 Introduction

This chapter describes the methodology used for this research. In particular structural equation modelling was used for the assessment of two theory-based models. Factor analysis was used to establish commitment scores for the customers in the analyses, by reducing the dimensionality of the multivariate dataset for the 2002 averaged questionnaire scales. Data mining was used to predict the commitment level using transactional and demographic data. In particular logistic regression, classification trees and neural networks were utilised. T-tests, chi-squared tests and ANOVA were used to assess significant differences between the 2002 and 2003 results.

2.2 Customers used in the analyses

This research is based on the financial institution personal customers. An attempt to model and assess personal customer commitment at a household level was undertaken using a sample of personal customers. Household level was used rather than individual level as firstly, the personal division is dominated (97%) by joint account customers and, secondly the research focused on life stage, so it was logical to look at the household level. The purpose of this research was to identify whether any variables are significant in predicting commitment level. This information will allow the financial institution not only to tailor contacts with the customer to suit their commitment level, but also to use commitment level as an input variable for defection modelling. Personal customers were selected for this research because they potentially earn the financial institution more revenue per customer over their lifetime than do retail customers, while Private customers are already very closely managed.

Internal research conducted in February 2002 by the financial institution on the profitability of customers by division found that the private and personal segments appeared to be more profitable than retail customers. Personal customers accounted for 10.5% of the total customer base and 14.5% of total profit. Retail customers accounted for 71.5% of the total customers but only 26.5% of total profit. Private customers accounted for 3.3% of the customer base and 9.9% of the total profit. Long-term relationships with Personal customers were considered desirable because of their potential profitability.

The defection rate for the year December 2001 – December 2002 for the combined private and personal customers was 1.2% total defection and 16.2% partial defection (when some but not all accounts are closed). Defection for retail customers during the same period was 11.7% total defection and 9.6% partial defection. So it seemed that the partial defection issue was of greater concern than total defection in the case of the personal customer segment and needed to be addressed.

True attitudinal and behavioural loyalty should be established so customers have little interest in competitive offerings (Dick & Basu, 1994 and Jacoby & Chestnut, 1978). In other words commitment and loyalty have been linked with customer retention. Therefore an understanding of customer commitment may allow managers to increase retention of desirable customers.

As well as an important input into a defection model, knowing customers commitment level is useful for advertising purposes (Hofmeyr and Rice, 2000). Those customers who are already committed are most receptive to advertising. They look at the advertisement as reinforcement that they have made the right decision to be committed to a company or brand. Those who are not committed find that the advertising will remind them of the reasons they are dissatisfied with the company or brand.

It is important that a company knows what proportions of their customers are uncommitted. If there is a large proportion uncommitted then the reasons for this low commitment need to be understood and steps taken to manage the relationship appropriately in order to build commitment (Hofmeyr and Rice, 2000).

In this study the hypothesis tested was that life stage influences commitment level. The most well known model used for commitment is the 'Conversion' model (Hofmyer and Rice, 2000), which combined four components, namely perception of alternatives, ambivalence, importance of brand choice and satisfaction with brand. Other researchers have linked inertia with loyalty (Papatla and Krishnamurthi, 1992) and the satisfaction component has been split further to include strength of relationship (Barnes, 1997). Barriers to defection such as relationship termination costs and continued relationship benefits (Morgan and Hunt, 1994) also trust and shared values have been found to be significant in explaining satisfaction. Satisfaction is a component of commitment and research suggests that customer satisfaction ratings vary because of differences in customer characteristics including the life stage variables of sex, age, and socio-economic level (Mittal & Kamakura, 2001 and Bryant & Cha, 1996). This research investigates the significance of life stage for predicting commitment levels.

This research used two questionnaires over a one-year period, using two groups of Personal customers. The 2002 questionnaire was sent to 9564 customers, while 8990 customers received the 2003 questionnaire. There were 2995 customers common to both surveys. The aim was to establish whether commitment levels are significantly different across life stages and also if there were any life stage whose commitment levels was more stable than others. Life stage for this research was based on the maximum age of the couple and kids indicator (see later) developed by the financial institution.

The next section describes the questionnaire design.

2.3 The Questionnaire

The questionnaire was designed by the marketing department prior to this research being implemented. As we lacked input to the questionnaire development process, the details of the questions and nature of the measures are not as precise as we might have desired. Even so we believe that the information gained from the questionnaire is useful.

The research population for this part of the study consisted of the 2002 survey for each personal financial advisor.

The first questionnaire, from June 2002, was used to form commitment levels and these levels were then modelled using data mining techniques. The second, from June

2003, was used to validate the 2002 analysis and to compare the commitment level of customers from both surveys. The 2002 questionnaire was posted to a representative of each of the 9564 members of the research population of personal customers in June 2002. A judgemental sample was selected with each personal financial institutioner contributing his/her 85 personal managed groups with highest footings. A covering letter was included with the questionnaire stating that the financial institution valued their customers' views and telling customers that if they returned the questionnaire they would be entered into a draw for 5 chances to win $1000. No follow-up letter or second questionnaire was sent out but a thank you letter was sent to those who responded. The cut-off date for returning questionnaire was the end of July 2002.

An almost identical questionnaire was sent out in June 2003 to 8990 personal customers. The reduction in the research population was due to a reduction in the number of personal financial advisors in employment with the financial institution at the time of the second questionnaire. The only change to the 2003 questionnaire was the omission of two statements within two questions and the addition of one open-ended question (see appendix 1). The questions used in the 2002 analysis were unaffected by these changes and we felt that such changes would not significantly affect responses in 2003. In 2003 the financial institution used two methods of delivery, 5198 questionnaires were sent by post and 3792 by email. The email customers received their covering letter by post two weeks prior to the questionnaire being sent via email. The letter for both postal and email customers contained the same information as in 2002 with the exception that there were only 3 prizes of $1000 to be won by those who responded. Again no follow-up letter or questionnaire was sent. The financial institution chose to email the questionnaires to all those customers with a reliable email address on the database because this distribution method was considered to be more economical.

Both questionnaires were designed to measure the performance of the financial institution's personal financial advisors and comprised 20 questions, three of which, positioned toward the end of the questionnaire, were designed to measure customers' commitment levels to the financial institution. The three questions for commitment consisted of varying length point scales, seeking respondents' views towards perception of alternatives, inertia and ambivalence.

The perception of alternatives question was a simple ten-point assessment scale for six New Zealand registered financial institutions. The question was worded 'Even if you have never used a particular financial institution or the financial institution, you can have a general impression based on what you have seen or heard about them. Please tell us how you feel about each of the following financial institution, regardless of whether you have used them or not.' Response categories ranged from (10) 'perfect in every way' to (1) 'completely unsatisfactory'.

The inertia question gave three statements, each with a five-point agree/disagree scale, ranging from (5) 'strongly agree' to (1) 'strongly disagree'. Some of the statements were worded negatively to reduce response bias. The statements used in this analysis were 'Changing financial institutions is a real hassle', 'I have thought recently about changing financial institutions' and 'I like to look around for the best interest rates'.

Finally, the ambivalence question consisted of three options designed to identify how undecided the customer feels about continuing to use the financial institution as a provider of financial services. The question was worded 'Please tick one of the three options to tell us how you feel about continuing to use the financial institution as your provider of financial services.' The options were 'there are many good reasons to continue using the financial institution, and no good reasons to change', 'there are many good reasons to continue using the financial institution, but also many good reasons to change' and 'there are few reasons to continue using the financial institution, and many good reasons to change'.

We also used several other questions from the questionnaire to supplement these three. This allowed us to also analyse importance of brand choice, satisfaction with brand and strength of relationship.

There was no specific question to assess importance of brand choice to the customer, so a surrogate statement 'All financial institutions are the same' was used. This statement allowed a five-point agree/disagree scale response option ranging from (5) 'strongly agree' to (1) 'strongly disagree'. It was assumed that a customer who strongly agreed that all financial institutions are the same does not find the choice of brand important, while a customer who strongly disagrees is more likely to think that the choice of brand is important.

Satisfaction with brand was assessed using two 5-point scale questions with response options ranging from (5) excellent to (1) poor. The first question related to the overall quality of service provider by "your personal financial advisor". The second question was 'How would you rate your overall satisfaction with the financial institution as a whole?'

Strength of relationship was assessed using the question 'Who would you consider to be your main the financial institution?' We also used 'Would you recommend the financial institution in the future to friends or relatives for their financial needs?' this was a 5 point scale ranging from (5) definitely would to (1) definitely would not.

The next section describes the data collection procedure.

2.4 Data collection

The population studied for this research were the upper portion of the financial institution personal managed groups, based on the group's footings. The majority of managed groups in the personal division contain households and immediate family, but there are occasions when a financial advisor groups customers together purely for his convenience e.g. all Asian customers together. Also just because a customer has a household income greater than $100k or footings of $125k - $500k does not mean they have to be in a managed group. If the customer does not want to deal with a personal financial advisor then he will be dropped to the retail division and would not therefore be part of the sample population even though he may qualify on footings. So managed groups take many forms. The research population of 9564 in 2002 consisted of the top 85 managed groups (based on total group footings over $100,000) for each of the 113 personal financial advisors. Note that three of the personal financial advisors did not have 85 managed groups which qualified for selection, only 73, 67 and 74 managed groups could be selected for these financial advisors. The research population of 8990 in 2003 was selected using a slightly different approach. As there were only 97 personal financial advisors in 2003, the financial institution decided to select all managed groups which qualified for selection up to a maximum of 100 per personal financial advisor. There was an overlap of the 2002 and 2003 samples of 2995 customers.

A judgemental sampling method was designed to select the most important member of each of the managed groups, to receive a questionnaire. The two questionnaires used different selection methods but both selected one account from each group and holders of that account received a questionnaire. A hierarchical selection process for 2002 gave priority to the joint account customers with the highest footings. If there were no joint accounts, the non-personal account holder with highest footings was selected. If there were no joint accounts or non-personal accounts in a managed group then the individual with the highest footings was selected to receive the questionnaire. Following this procedure the 2002 questionnaire was sent to 9284 joint account customers, 128 non-personal account holders and 152 individuals. The total response rate in 2002 was 38.70%, giving a total of 3701 that returned the questionnaire. These 3701 responders consisted of 3604 joint accounts, 52 individuals and 45 non-personal accounts. A chi-squared goodness-of-fit test was used to determine whether these response categories match the actual group breakdown (see Table 2.1).

Table 2.1 2002 questionnaire respondents versus non respondents

	JOINT ACCOUNTS	**INDIVIDUALS**	**NON PERSONAL**	**TOTAL**
Responder	3604	52	45	3701
Non responder	5680	100	83	5863
TOTAL	9284	152	128	**9564**

The chi-square test was not significant (p-value = >0.25) so there is no evidence of response bias in the 2002 responders.

The purpose of using the questionnaire for this research was to establish a level of commitment for each customer allowing tests for relationships between commitment and life stage. An issue with the way in which the financial institution sent out the questionnaires was that 97% were addressed to couples, so we could not be sure which member of the couple completed the questionnaire. Commitment is the strength of a psychological bond with a product or company, which is unique to the individual (Beatty, Kahle and Homer, 1988) so the questionnaires should ideally have been sent to individuals. However, with the personal customer segment being dominated by joint accounts, it was assumed that the commitment scores obtained reflected the commitment levels of these couples.

As we were interested in life stage and commitment, it would not make sense to consider the life stage of flatmates or even a mother/daughter relationship. The decision was therefore made to predict the commitment of a couple only if they were in a 'married-type' relationship (e.g. married, engaged, de facto and partners). This means that the research population was restricted to Personal customers in 'married-type' relationships. Individuals without 'married-type' relationships were not included because they were so few in number.

We were confident that the only real problem in defining life stage for couples would be age. As we were not certain how best to represent a couple's age, we used three age variables: minimum age, maximum age and a flag for those couples who had an age gap greater than 15 years. All other variables, namely life stage and financial activity variables, we felt would be unaffected by using couples as the analysis unit, as long as the financial activity variables included the joint and all personal accounts for each couple.

So, taking the original joint account for the 'married-type' couples who responded, we also included in the analysis any individual accounts that the couple had in order to get a complete picture of their transaction patterns for the financial activity variables. We did not simply omit the 52 individuals who returned a questionnaire in 2002. If the individual also had a joint account in a 'married-type' relationship, then we included this customer's joint account and any individual accounts for this couple when determining values for the financial activity variables. This enabled us to retain 22 individuals as their data was now in the current form. If the original individual did not have a 'married-type' joint account then the individual was rejected because this group was so small (30). Joint accounts were not found for the individuals who responded in 2003 due to lack of time. We also rejected any customer who did not respond to all of the questions needed to form a commitment level. As commitment level is central to this research we needed to be sure of its accuracy. The final sample of customers used in the 2002 research consisted of 2453 responders as shown in table 2.2

Table 2.2 2002 Customer Selection Procedure

TARGET	QUESTIONNAIRES SENT	RESPONSE RECEIVED	INCLUDED IN THE ANALYSIS
'Married-type' couples	8590	3386	2431
Other joint accounts	694	218	0
Individuals	152	52	22*
Non-personal accounts	128	45	0
TOTAL	9564	3701	2453

(* these individuals had 'married-type' joints, which were included in the analysis)

the financial institution decided to use a different selection method for recipients of the 2003 questionnaire. The account chosen to represent the managed group was the account with highest footings regardless of account type. Questionnaires were then addressed to the customers whose account was selected. In the case of joint accounts being chosen, the questionnaire was addressed to both parties of the joint account. Fewer questionnaires were sent in 2003, as there were fewer personal financial advisors in than 2002.

Due to the change in selection criteria and smaller sample in 2003 only 2995 from the original survey of 9564 in 2002 were selected to receive questionnaires in 2003, of these only 845 had been used in the 2002 analyses. The first sample of customers we analysed from the second questionnaire consisted of all those 'married-type' couples who responded in 2003. The second sample consisted of those customers who responded in both 2002 and 2003 (see table 2.3). Couples in both of these samples had to have answered all the relevant commitment questions. Although the response rate was only slightly lower for 2003 the percentage of useable response was obviously a lot lower.

Table 2.3 Response rates in 2002 and 2003

	2002	**2003**
Questionnaires sent	9564	8990
Responses received	3701 (38.70%)	3354 (37.36%)
Usable responses	2453	852
Common usable responses	223	223

The 2003 survey consisted of 58% postal and 42% emailed questionnaires. Traditionally, researchers have relied extensively on postal questionnaires to collect the information from respondents (Dillman, 1978). As email becomes more accessible, the possibility of collecting data from large sample email surveys grows more attractive. A few researchers (Kiesler & Sproull, 1986; Walsh et al 1992) have suggested that email may become a routine research tool in the future.

Important factors that influence the choice of method of questionnaire delivery are: speed; cost; sample control; convenience; response rate and quality of data (Dillon, Madden and Firtle, 1993). Metha and Sivadas (1995) believe that Email questionnaires may be more difficult to complete. "It is much easier to circle a number in a paper-pencil exercise than to use the cursor keys to position a cursor and then enter a response on an electronic survey."

Another difference between mail and email is the tone of response aspect of email. Kiesler & Sproull (1986) found that electronic respondents tend to be "both more self-absorbed and uninhibited" so they generally write longer answers to open-ended questions. However, email has less anonymity to the respondent as the receiver of an email can in most cases identify the respondent from the header (Metha and Sivadas, 1995). The main disadvantage with email is that not everybody uses email. This tends to limit its usage to middle-to-upper class respondents. However, with the exponential growth of the Internet and email users (Tse, 1998) email has potential. There are also several advantages to email:

- No need for prepaid return envelopes
- Eliminates time and effort stuffing envelopes

- Cost of delivery is no longer dependant on the size of the survey
- Transmission of questionnaires by email is extremely fast
- An emailed survey might imply urgency and avoid being perceived as junk mail

The email/postal breakdown for the 2003 sample is summarised below. It is important to compare response rates by delivery type in order to evaluate the sample design.

Table 2.4 Comparison of 2003 postal and email response rates

	POSTAL	EMAIL	TOTAL
Selected	5198	3792	8990
Response	1737	1622	3359
Response rate	33.42%	42.77%	37.36%
Useable responses	573	279	852
Useable response rate	11.02%	7.36%	9.48%
Common (2002/2003) useable response	161	62	223
Common usable response rate	3.10%	1.64%	2.48%

Clearly although response rate were significantly higher for email questionnaires (p-value < 0.00005) usable response rates were lower for email questionnaires (p-value < 0.00005). These figures indicate some bias in favour of postal questionnaire in our 2003 survey.

Next we explain the statistical methods used in the research.

2.5 Statistical Methods

In this section we explain the main statistical methods used for the research, namely multiple iteration for missing values, Cronbach's Alpha for scale reliability, structural equation modelling, factor analysis, artificial neural networks, logistic regression and classification trees.

2.5.1 Multiple iteration

Multiple iteration is a simulation method used to impute missing data. The base SAS procedure for multiple imputation in Release 8.1 creates multiply imputed data sets for incomplete p-dimensional multivariate data. It offers three methods for creating the imputed data sets: the regression method, the propensity score method, and the Markov Chain Monte Carlo (MCMC) method. The output dataset contains 5 imputed versions of the original data with the missing values replaced by imputed values. The MCMC method was used in these analyses using the single chain option for all imputations, as this allows the specification of the initial estimates for this method.

Recent research advocates that the preferable method to impute missing data for factor analysis is a maximum likelihood – based approach. Kamakura & Wedel (2000) found that the models recovered the true factor structure in synthetic data quite well even when 50% of observations were missing. However, direct maximum-likelihood methods are computationally complicated and require a special implementation for each new type of model. Multiple imputation on the other hand, is a general technique that can be applied to a wide variety of modelling problems (Shafer and Olsen, 1998).

Multiple imputation (MI) is a Monte Carlo approach, which solves an incomplete-data problem by repeatedly solving a complete-data version. Each missing value is replaced by a set of $m>1$ plausible values drawn from their predictive distribution. The observations in the dataset which have complete data are used to predict the missing values. The variability among the results of the multiple imputation analysis provides a measure of the uncertainty due to missing data. After performing MI there are m apparently complete datasets, each of which can be analysed. The estimates and standard errors of the m datasets are then combined using simple averaging rules provided by Rubin (1997) to produce overall estimates and standard errors that reflect missing-data uncertainty.

In multiple imputation a very small number of imputations will suffice, 3 are often adequate and 5 are certainly ample. The base SAS procedure for multiple imputation returns 5 imputations.

Multiple imputation is fairly robust to departures from the imputation model. Variables, which are heavily skewed, should be transformed before using this method and then transformed back. No prior distribution for parameters is required in the case of large samples. However, missing data needs to be missing at random (MAR) e.g. probabilities of missing values may depend on data values that are observed but not ones that are missing. It is not possible to relax the MAR assumption. The alternative to assuming MAR is to propose a formal probability model for response and carry out an analysis under that model, which is a difficult task. In the vast majority of studies, principled methods assume MAR will tend to perform better than ad hoc procedures such as list wise deletion or imputation methods (Hair et al, 1998).

2.5.2 Cronbach's Alpha

Cronbach's alpha (Cronbach, 1951) was used to determine the reliability of the scales created from the questionnaire. Cronbach's alpha measures how well a set of items measures a single unidimensional latent construct. It is a function of the number of test items and the shared variance of these items. Cronbach's alpha is defined as

$$\alpha = \frac{p}{p-1}\left[1-\frac{\sum_{j=1}^{p} Var(X_j)}{Var\left(\sum_{j=1}^{p} X_j\right)}\right]$$

where p = number of items

If the inter-item correlations are high, then there is evidence that the items are measuring the same construct and therefore the scales are reliable. If the alphas for the standardised variables are greater then 0.7 then they are acceptable. If this is true for both the raw variables and the standardised variables, we can conclude that the scales show good reliability (Nunnaly, 1978). Although Nunnaly has indicated that 0.7 is acceptable, lower thresholds are sometimes used in the literature (Santos, 1999).

2.5.3 Structural equation models

These were used to test the commitment models of 'Conversion model' and the 'Extended model' for commitment.

Essentially Structural Equation Modelling (SEM) allows researchers to check if a given theoretical model is supported by the data. SEM is a confirmatory method used to test whether the data supports a theoretical model. SEM is a straightforward and efficient method of dealing with multiple relationships simultaneously. All SEM techniques are distinguished by two characteristics (i) estimation of multiple and interrelated dependence relationships and (ii) ability to represent unobserved concepts in these relationships and account for measurement error in the estimation process. SEM allows for causal relationships in which the change in one variable is assumed to result in a change in another.

Path diagrams can be constructed within the AMOS 4.0 environment, a structural equation programme (Arbuckle, 1997). These path diagrams are then translated into a series of structural equations by the programme. Ellipses are used to indicate latent constructs, rectangles are used to indicate observed (manifest) variables, straight lines indicate causal links and curved lines indicate correlations. A variable may be an indicator for more than one construct, but this situation needs to have a strong theoretical rationale and should generally be avoided.

Amos uses Ordinary Pearson correlations to estimate the proposed model, although they tend to produce reliability estimates which are too low when the manifest variables are measured using a discrete ordinal scale (Olsson, 1979; Rigdon, 1991), But Babakus et al (1987) have found that the goodness of fit deteriorates when the alternative Polychoric and Tobit correlations are used.

Structural equation modelling was used in this study to combine aspects of multiple regression (examining dependence relationships) and factor analysis (representing unmeasured concepts – factors with multiple variables) to estimate a series of interrelated dependence relationships simultaneously. First we need to draw upon theory, prior experience and the research objectives to distinguish which independent variables predict each dependent variable. This type of modelling is ideal for our study because it can incorporate variables which are not directly measured (i.e. latent variables). Commitment could not be directly measured in our study but we had a series of unrelated questions that were indicators of commitment. Using this model decreases the problem of how to formulate a theoretical concept question (e.g. attitudes towards the financial institution). By asking simple questions which consumers should have no problem answering we get better representation and reduce measurement error. The path diagram should not be extensively modified to force a good fit, as this is a confirmatory method guided more by theory than by empirical results.

When using structural equation modelling the default method of estimation is maximum likelihood. The maximum likelihood estimate is defined as that estimate of the parameter that results in the maximum likelihood or probability of observing the given sample data. It is the value of the parameter for which the sample data will occur most often.

These estimates depend upon the assumption of a multivariate normal distribution. One method to diagnose the presence of distributional problems is to use bootstrap simulations. Firstly the SEM is run using the maximum likelihood estimation method and then the same model is re-run using 1000 bootstrap simulations. If there are significant differences between the standard errors of the maximum likelihood method and the bootstrapping then this would indicate that there are distribution concerns and that alternative methods of estimation, like asymptotically distribution-free (ADF) estimation should be used instead of maximum likelihood.

AMOS calculates fit measures for each model specified by the theoretical model and two additional models – saturated and independent. The saturated model has no constraints placed on the population moments; it is the most general model possible and guaranteed to fit any data perfectly. Any AMOS model is a constrained version of the saturated model. The saturated model is needed to compute chi-squared and derived statistics such as RMSEA.

The independent model is the opposite extreme. The observed variables are assumed to be uncorrelated with each other. It is so constrained that you would expect it to provide a poor fit to any interesting set of data. This model is required for calculating several fit indices such as CFI and NFI. So these two models are the extremes between which specified models lie.

There are various measures to assess the goodness of fit of the data to the model. One of the most popular measurements used for the assessment of the SEM is the chi-square statistic. However, this measure is not reliable when a sample size is large. Joreskog (1993) sounded an early warning about over interpretation of the chi-square test statistic. He noted that ' in large samples even trivial deviations of a model from the actual structure could be detected and could lead to a rejection of the null hypothesis'. Beardon et al (1982) and Bentler (1990) also regarded the chi-square test as being too sensitive to sample size. When sample size exceeds 200 this test will tend to always reject fitted models. For this reason the chi-square goodness-of-fit test for SEM was not used in this study.

We also could not use the PMIN, P or FMIN measures as they are based upon the chi-square statistic. They all measure some aspect of discrepancy: FMIN is the minimum value, estimated F, of the sample discrepancy, defined as the fitted chi-squared value, calculated using sample moments from the multivariate normal likelihood (see Arbuckle and Wothke (1997) page 390 for the formulae). Most models are useful approximations that do not fit perfectly in the population (Arbuckle, 1997). In other words, the null hypothesis of perfect fit is not creditable to begin with and will in the end be accepted only if the sample is not allowed to get too big. "If the sample is small then the χ^2 test will show that the data are 'not significantly different from' quite a wide range of very different theories, while if the sample is large, the χ^2 test will show that the data are significantly different from those expected on a given theory even though the difference may be so very slight as to be negligible or unimportant on other criteria" (Gulliksen and Tukey, 1958). So in other words we are not able to rely on any of the measures, which involve the minimum sample discrepancy function.

For larger samples goodness-of-fit measures based on the population discrepancy (F_0) measures are preferred. This value is estimated by fitting a model to the population moments (F_0) rather than to the sample moments ($\hat{F}$). Low values are indicative of a good fit.

$\hat{F}_0 = \max\left(\dfrac{\hat{C}-d}{n}, 0\right)$ where $\hat{C}$ = minimum discrepancy, n$\hat{F}$, d = degrees of freedom and n = total number of observations in all groups combined minus the number of groups, N-r.

We start our discussion of goodness-of-fit measures used in this study by describing the absolute fit measures, which are measures of overall goodness-of-fit for both the structural and measurement models collectively. In this study we use only four absolute measures, namely RMSEA, GFI, PCLOSE and MECVI.

RMSEA – root mean square of approximation

The RMSEA measure attempts to correct for the tendency of the chi-square statistic to reject models when the sample size is greater than 200 while applying a penalty factor for more complex models. RMSEA is the discrepancy per degree of freedom, measured in terms of the population. A RMSEA of about 0.05 or less would indicate a close fit of the model in relation to the degrees of freedom (Arbuckle, 1997). Values ranging from 0.05 to 0.08 are deemed acceptable (Hair et al). They also said that RMSEA is best suited in a confirmatory or competing models strategy with larger samples, therefore ideal for this study. RMSEA will favour a simpler model if the fit is adequate. The formulae for RMSEA is:

$$\text{RMSEA} = \sqrt{\frac{\hat{F}_0}{d}}$$

GFI – Goodness-of-fit index

This is a nonstatistical measure ranging in value from 0 (poor fit) to 1.0 (perfect fit). It compares the squared residuals from prediction to the actual data, but it is not adjusted for degrees of freedom. Higher values indicate a better fit, but no absolute threshold levels for acceptability have been established, as this measure is for comparing models.

$$GFI = 1 - \frac{\hat{F}}{\hat{F}_b}$$

Where $\hat{F}_0$ is minimum value of the population discrepancy function and $\hat{F}_b$ is obtained by evaluating F under the independence model.

PCLOSE

Another measure of population discrepancy, which we can use, is PCLOSE. This measure is for testing the null hypothesis that the RMSEA is no greater than 0.05. If this is significant then it means that the RMSEA is less than 0.05, indicating that the model is a close fit to the data.

MECVI

Finally, we shall use the information-theoretic measure of MECVI. This statistic creates a composite measure of badness of fit ($\hat{F}$) and the complexity (q) of the model by forming a weighted sum of the two. Simple models that fit well receive low scores according to such a criterion. Complicated, poorly fitting models get high scores. The MECVI statistic is not to be used to evaluate one model; it needs to be used to compare several models.

$$\text{MECVI.} = \frac{\hat{F} + 2q\dfrac{p(p+3)}{N-p-2}}{p(p+3)}$$

Where q = the number of parameters

p = the number of sample moments in all groups combined

N = sample size

Next we describe the incremental fit statistics, which are measures of goodness-of-fit that compares the current model to a specified null model to determine the degree of improvement over the null model. In our case the null model is the independence model. In this study we use only four incremental fit statistics, namely AGFI, CFI, NFI and TLI.

AGFI - Adjusted goodness-of-fit index

This measure takes into account the degrees of freedom (d_b) for the null (independence) model.

$$\text{AGFI} = 1 - (1 - GFI)\frac{d_b}{d}$$

Recommended acceptance values are greater than or equal to 0.90 (Hair et al).

CFI - comparative fit index

$$\text{CFI} = 1 - \frac{\max(\hat{C} - d, 0)}{\max(\hat{C}_b - d_{b,} 0)}$$

Where $\hat{C}_b = n\hat{F}_b$ is the minimum discrepancy of the independence model to see how large the discrepancy becomes. $\hat{C} = n\hat{F}$ is minimum discrepancy of the model being evaluated.

NFI – normed fit index

Measures range from 0 (no fit at all) to 1.0 (perfect fit). NFI is a relative comparison of the proposed model to the null model (independence model).

$$\text{NFI} = 1 - \frac{\hat{C}}{\hat{C}_b}$$

TLI – Tucker-Lewis Index

This combines a measure of parsimony into a comparative index between the proposed and null models resulting in a range 0 to 1.0 where greater than or equal to 0.9 is acceptable.

$$\text{TLI} = \frac{\frac{\hat{C}_b}{d_b} - \frac{\hat{C}}{d}}{\frac{\hat{C}_b}{d_b} - 1}$$

Where d = degrees of freedom

Bentler and Bonett (1980) referred to the NFI and TLI measures as this: 'In our experience, models with overall fit indices of less than 0.9 can usually be improved substantially.'

The goodness of fit measures in SEM are not rigorous and are best used to compare models rather than to assess an independent model. '...one must rely upon a judicious interpretation of statistical and non statistical rules-of-thumb along with conceptual and philosophical criteria. The process inherently contains elements of subjectivity and social consensus/discensus. But these are essential inputs to any creative process and make research the frustrating, yet fascinating, endeavour it is'. (Bagozzi, 1988)

2.5.4 Factor analysis

Factor analysis is a dimensionality reduction procedure, based on proven statistical techniques. It uses the covariances between the original variables to create latent (common) factors. It explains only the common or shared variation so is likely to be more useful when measurement accuracy is low, or discrete ordinal scales are used to collect information.

Assumptions:

1. the common factors are independent and are standardized variables with zero mean and unit variance.
2. the specific variates (residual information unexplained by the latent factor) are independent with zero mean but non-zero variances.
3. common factors and specific variates are independent of each other.

Normality is not assumed unless hypothesis testing is required. When performing factor analysis you are able choose the number of common factors (m), with $m \leq \frac{1}{2}(p-1)$ where p = number of variables.

There are an infinite number of alternative solutions to factor analysis, which allows the use of factor rotation in order to produce simpler factors that are more easily interpreted. There are a couple of ways to create the factors:

(i) maximum likelihood factor analysis
- assumes normality
- scale invariant
- uses formal tests of significance to determine the number of common factors

(ii) Iterated principal components factor analysis
- sensitive to different measurement methods
- does not require normality
- uses a reduced covariance/correlation matrix

If we choose only one factor then the total variation within the data is to be explained entirely by this one factor, shown by a cumulative value greater than 1 for the eigenvalues. This was found to be appropriate in this study.

Since the common factor is unobservable, we cannot measure it directly, however we can measure the indicators of the unobservable factor and compute the correlation matrix containing the correlations among the indicators. Now given the computed correlation matrix among the indicators, the purposes of factor analysis is to

(i) identify the common factor that is responsible for the correlations among the indicators; and
(ii) estimate the pattern and structure loadings, communalities, shared variates and the unique variances

In summary, factor analysis is a technique, which can be '... used to develop scales for the various unobserved constructs such as attitudes, image, intelligence, personality and patriotism.' (Sharma, 1996). In our case the unobserved construct is commitment.

2.5.5 Artificial neural networks

Artificial neural networks are often referred to as a 'black box' approach to data mining. This is because although the output often has business value, there is no way of knowing how the output was derived. They are used to model output using non-linear functions of input variables. These models can be used to predict the output variable. In this study neural networks are used to classify, predicting levels for discrete output variables. However, these models can also be used to predict values for continuous interval outputs.

A neural network consists of input nodes, hidden nodes and output nodes. Each of the hidden nodes takes many inputs and generates an output that is a non-linear function of the weighted sum of the inputs. The weights assigned to each of the inputs are obtained during a training process (often back-propagation) in which outputs generated by the net are compared with target outputs.

Backward propagation is an iterative method for estimating the weights in a neural network. This method takes the training data and uses the existing weights in the neural network to calculate the output. Errors are then calculated by taking the difference between the calculated output result and the expected result. The error is then fed back into the neural network and the weights are adjusted to minimise the error. This continues until a near optimum solution is reached. A danger for this method is local optima, where we can produce a model which accurately classifies the current training set in a local sense but is not the best solution on a global scale. Validation data is used to prevent over-fitting of the model. A model improvement produced using the training data is only acceptable if it also means an improvement for the validation data. It is possible for the neural network to learn the training data too well, therefore modelling is 'stopped' at the point where the validation data has the lowest error. The weights at this point are used in the final model.

The number of inputs, hidden nodes, outputs, and the weighted algorithms for the connections between nodes determine the complexity of a neural network. There is a drawback to neural networks in that the resultant training weights do not provide reasons for a particular prediction, so they are very difficult to interpret and explain.

Inputs are combined into a single output value using an activation function, which has two parts. The first part is the combination function that combines the inputs into a single weighted summed value (Berry and Linoff, 1997). The second part is the transfer function which transforms the value of the combination function to the output. There are three transfer functions which are commonly used, namely sigmoid, linear and hyperbolic. The most common is the S-shaped sigmoid function, which is a nonlinear function that transfers the combination function into a value between 0 and 1.

The sigmoid function formula is defined as:

$$\text{Sigmoid(x)} = \frac{1}{1+e^{-x}}$$

The predictive ability of a neural network depends in part on the quality of the training data. It is also important for the analyst to have some knowledge of the subject matter, especially for selecting appropriate inputs and choosing an appropriate error function. Experienced neural network users typically try several architectures to determine the best network for a specific data set. The design process and the training process are both iterative.

The strengths of neural networks are that they handle a wide range of problems, are able to produce good results for complicated problems and can handle categorical and continuous variables. However, they require inputs in the range 0 to 1 to be most powerful, it is difficult to explain the predictions and they require large amounts of data, particularly if discrete ordinal variables are used. When an input variable is discrete the neural network creates a series of binary variables, (one per discrete level) which can escalate the number of input nodes required if there are a large number of discrete variables used. Neural networks are not good when a large number of input variables are used with relatively few observations. A minimum number, approximately 10, of observations per input node is required for the training set. Also a large number of input variables make it difficult for the network to find patterns and trees are the preferred method of prediction in this situation.

2.5.6 Logistic regression

Logistic regression is an alternative for classifying data. Logistic regression '... does not make any assumptions regarding the distribution of the independent variables and is therefore preferred when the independent variables are a combination of categorical and continuous variables because in such cases the multivariate normality is clearly violated.' (Sharma, 1996). He goes on to say that logistic regression is also preferred when the independent variables are continuous, but are not normally distributed.

The logistic regression model takes natural logs of the odds of the event happening.

$$\text{Odds} = \frac{p}{1-p} \qquad \text{where p = Prob(event occurring)}$$

The log of the odds, called logit (p) is modelled using a linear function of the independent variable when p ranges from 0 to 1.

Logistic regression finds the best fitting equation as does linear regression, however the relationship between X and the response variable Y is in the form of an S-shaped curve. The dependent variable in logistic regression is qualitative, either a binary variable or an ordinal variable with several categories. In the binary situation rather than predicting the value of the response variable, it models the probabilities that the response takes one of these two values. Instead of minimising the sum of the squared deviations for best fit, it usually uses a maximum likelihood method, which maximises the probability of getting the observed data given the fitted regression coefficients. In logistic regression OLS regression is replaced by measures of deviance, as a measure of fit. This deviance is a measure of lack of fit of the data to the logistic regression model. These deviances are derived from ratios of maximum likelihood's under different models and are referred to collectively as likelihood ratio tests. (Cohen et al, 2003).

Maximum likelihood estimation

Below we assume that the dependent variable in the logistic regression is binary. Let Y be the random binary variable whose value is zero or one. The probability P(Y =1) is given by

$$P(Y = 1) = p = \frac{e^{\beta X}}{1+e^{\beta X}}$$

Where β is the vector of coefficients and X is the vector of independent variables. More generally for non-binary discrete dependent variables let π_i be the proportion of cases for category i.

A non-linear relationship that is often used for proportions is the logistic curve. It models the probability that the dependent variable Y falls in category i using the equation.

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_1 + ...)}{1+\exp(\beta_0 + \beta_1 x_1 + ...)}$$

This curve is always between 0 and 1. The logistic relationship is intrinsically linear because we can linearise the relationship using the transformation

$$\ln = \left[\frac{\pi_i}{1-\pi_i}\right] = \beta_0 + \beta_1 x_1 + ...$$

Three methods of multiple logistic regression are available for sequential variable selection. These are iterative approaches and involve considering only addition or deletion of a single variable at each step.

(i) Backward elimination

- start by fitting the model with all available explanatory variables
- next determine the best model by dropping one variable. The model with the minimum residual sum of squares in the best model at this stage. The variable to be deleted is the variable whose coefficient has a t-statistic closest to zero.
- This process continues until the t-ratio for the variable being dropped would be significant at some predetermined level – usually 0.01 or 0.05.

- Once a variable is dropped from the model using this method it cannot re-enter.

(ii) Forward selection

- this procedure starts with an initial model, which contains only the intercept and no explanatory variables.
- One variable is added to the model at a time.
- The first model contains the intercept and the single explanatory variable, which creates the model with the minimum residual sum of squares.
- Next the remaining variables are added, to see which one creates the model with the intercept and two explanatory variables with the minimum residual sum of squares.
- This continues until the t-ratio for adding another variable is not significant at a predetermined level.
- Once a variable enters the model, it cannot be removed at a later stage.

(iii) Stepwise regression

- This is a combination of backward elimination and forward selection. The procedure is the same as forward selection but after each step of forward selection, a step of backward elimination is attempted. A higher significance level is normally used for adding variables than for deleting them, in order to keep useful variables in the model.

The main advantage of forward selection over backward elimination is that the initial models that are fitted only contain small numbers of explanatory variables. So only models with small numbers of explanatory variables ever need to be fitted. However each step of forward selection requires more models to be fitted at each step. Backward elimination is generally to be preferred if it is computationally feasible, as the inter relationships of all the variables can be assessed.

2.5.7 Classification Trees

Classification trees provide a third classification method. A classification tree can be likened to a set of hierarchical questions. A record enters the tree at the root node, then a test is applied which will determine which node the record proceeds to, another test is applied and the record moves to the next node. This continues until the record reaches a leaf node and all the records, which end up in a particular node, are classified in the same way.

Classification trees are used to predict membership of objects in the classes of a categorical dependent variable using their measurements on one or more predictor variables. Classification trees are flexible in terms of types of variables and are useful for exploratory analysis. A tree is a hierarchy of questions about the objects, which are 'asked', and the final decision made depends on the answers to the previous questions. The tree consists of branches and leaf nodes and can be used for categorical or continuous predictors. However, in this research we use them only for classifications.

There are three methods, which are commonly used to split the nodes into further nodes, namely the CART, CHAID and C4.5 (the latest version being C5.0) algorithms.

CART, - one of the most popular methods was first published in 1984 by L. Brieman et al. This method builds a binary tree by splitting the records at each node according to a function of a single input field (Berry and Linoff, 1997) thus producing a binary tree. This is also known as the Gini index, which is defined as

$$\text{Gini} = 1 - \text{sum of } P_k^2$$

where P_k^2 is the proportion of objects classified to be in group k.

C4.5 – developed by J. Ross Quinlan it is very similar to CART but produces trees with varying numbers of branches per node. This method produces one branch for every value taken on by a categorical variable. This is also known as the Entropy index which calculates the amount of information needed to decide if an arbitrary object in the sample belongs to one of the groups is defined as

$$I = -\left[p_1 \log(p_1) + p_2 \log(p_2) + \ldots\ldots + p_k \log(p_k)\right]$$

Both of the above methods overfit the data then prune back some branches.

CHAID (Chi-square automatic interaction detection) – developed by J.A. Hartigan (1975) is the most widely used method. This method attempts to stop growing the tree before overfitting occurs. This method is restricted to categorical variables, so continuous variable need to be redefined into classes. The splitting criteria for this method are based on the chi-square test in the following manner. A chi-square test is performed on the cross-tabulation between the target and each of the independent variables. The p-values for each independent variable are then ranked. The variable with the largest chi-square statistic indicates the strongest association and is therefore chosen for the split. This method is used at each split.

Classification trees are only as good as the choice of analysis option used to produce them. For finding models that predict well, there is no substitution for a thorough understanding of the nature of the relationship between the predictor and dependent variables. There are two methods of pruning trees to avoid over fitting.

(i) Pre-pruning (Bonsai-ing) – tree construction is halted early by setting a threshold for goodness-of-fit measures, if the result of splitting a node means the measure falls below this threshold then splitting is halted. The problem with this method is that it is difficult to choose an appropriate threshold.

(ii) Post-pruning (or just 'pruning') – this method removes branches from a 'fully grown' tree. The validation data is used to decide which is the 'best pruned tree'. The size of the tree can be decided at the point where the validation data is at its lowest error rate.

Trees are flexible as they have the ability to examine the effects of predictor variables one at a time, rather than all at once.

Classification trees are a method of supervised learning used in Data Mining. Most trees are less than perfect, as variables don't completely predict the outcome. Also data is noisy and may be incomplete, although trees can deal with missing data. We need to determine the best tree without overfitting or underfitting the data.

The main strengths of trees are that they are able to generate understandable rules, not much computation is involved, and they provide a clear indication of which variable are important to the decision. However, trees are not as effective when the target variable is continuous.

2.6 Data Analysis

Two surveys, one year apart, were used to develop commitment levels for each couple for 2002 and 2003. Data mining techniques were used to predict these commitment levels. Finally, comparisons were made between the two questionnaires particularly in regard to differences in commitment level.

2.6.1 Commitment Level

To develop a commitment level for each couple in each year we:

(i) summarised the scales from the 2002 questionnaire into six input components for the two theory-based theoretical models to be tested. Cronbach's Alpha was used to test the reliability of any scale with more than two items.

(ii) used SEM to establish if the data fitted the theory-based models, namely the 'Conversion' model and an extended version of the model

(iii) used the standardised regression weights from the 'best' SEM to produce logical ad hoc commitment levels

(iv) used factor analysis to create a continuous commitment score

(v) compared both the ad hoc logical commitment levels and the continuous commitment score to confirm the reliability of the ad hoc method boundaries for four final commitment levels.

Figure 2.1 on the following page shows the steps taken to develop commitment levels.

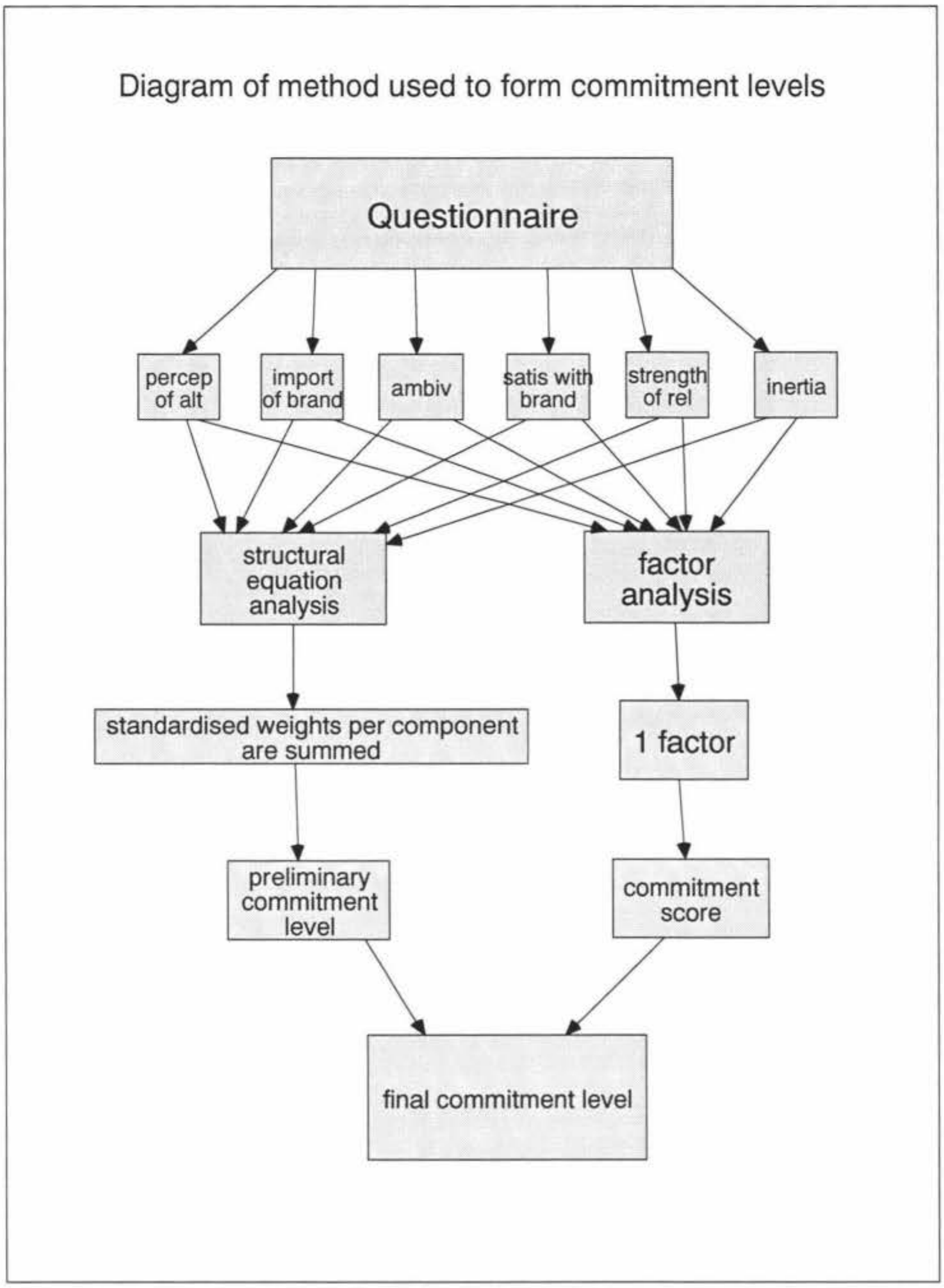

Figure 2.1 Outline of method used to create final commitment levels

The first stage of the research involved testing if two theory-based models for commitment were supported by the data from the 2002 questionnaire. This was done using structural equation modelling. The first model checked was the well known 'Conversion Model' (Hofmeyr and Rice, 2000), which combines four components: satisfaction with brand, perception of alternatives, ambivalence and importance of brand choice. The second commitment model tested combined the Conversion model components with strength of relationship (Barnes, 1997) and inertia (Papatla and

Krishnamurthi, 1992). Barnes found strength of relationship was influential on satisfaction. While Papatla and Krishnamurthi relate the 'dynamics of choice behaviour' (ambivalence) with inertia. We call this second model the 'Extended model'.

The questions used in each of the components discussed above are noted in table 2.6 below. Only those responders who answered all the relevant questions were included in the analyses. No imputation of missing values would be used in this stage of the analysis, as we wanted to be certain that the commitment levels were accurate. The 'Conversion' model has four components associated with it: satisfaction with brand, ambivalence, perception of alternatives and importance of brand choice. The 'Extended' model has six components associated with it: satisfaction with brand, ambivalence, perception of alternatives, importance of brand choice, inertia and strength of relationship.

Table 2.6 Questions from the 2002 questionnaire used for theoretical model components

Used in SEM models*	COMPONENT	Averaged Scale	QUESTIONS (2002)
1 & 2	Satisfaction with brand	yes	Q6_1, Q6_4, Q19
1 & 2	Perception of alternatives	no	Q14_1, Q14_2, Q14_3, Q14_4, Q14_5, Q14_6
1 & 2	Ambivalence	no	Q15
1 & 2	Importance of brand choice	no	Q16_3
2	Inertia	yes	Q16_1, Q16_2, Q16_5
2	Strength of relationship	yes	Q13, Q18

(* model 1 = 'Conversion' model, model 2 = 'Extended' model)

The six components for the two theory-based models were created from the scaled questions from the survey. The three average component scales were created and checked using Cronbach's Alpha. Coefficients greater than the cut-off of 0.7 are deemed to be reliable (Cronbach, 1951).

The perception of alternatives component (table 2.6) was not measured using an averaged scale. The idea behind the perception of alternatives set of questions was to compare the rating of the financial institution to five others in New Zealand. We created an alternative perception score using a simple count of the number of financial institutions the customer ranked as being equal or better then the financial institution. This gave us an idea of the relative attractiveness of other financial institutions to the customer. The importance of brand and ambivalence components were not averaged as they consisted of only one question so the actual raw discrete ordinal responses were used.

Structural equation modelling (SEM) was used to check whether the 2002 data supported the theoretical models. It was found that the extended model was the 'best'. Using the standardised regression weights for the six components from the 'best' SEM model, the component-averaged scales were used to form four commitment levels in a fairly ad-hoc manner that will be explained in the results sections. Principal components factor analysis was performed on the six component scores, and then used to create continuous commitment scores. Comparison of the SEM commitment levels and the continuous score allowed the confirmation of the reliability of the ad hoc method boundaries for the four commitment levels. This allows the conversion of customers answers to the relevant questions to one of four commitment levels, namely low, medium/low, medium/high and high.

The same scales and factor scores were used for 2002 and 2003 for the sake of consistency.

2.6.2 Predicting commitment

This stage of the research involved extracting all relevant demographic and transactional data from the financial institution's data warehouse to be used to predict commitment level using data mining modelling techniques. The aim was to establish whether life stage was a significant variable in predicting commitment level. In this stage any missing data was imputed using multiple iteration.

The data based model used five input components:

(i) life stage – age, income, kids

(ii) lifestyle – categories of expenditure, urban or rural area

(iii) the financial institution products held – number, recent changes, time since/to maturity

(iv) transaction patterns and transaction channels

(v) customer details

A large range of variables were extracted and formed for these components and details are given in the appendix 2.

The first input component, life stage, was calculated from age, income, probability of children at home, and a life stage variable (age and kids indicator combined). As the financial institution generally do not have information on whether customers have children, the 'kids' binary variable flag was created by the financial institution using automatic payments, direct debits, credit card and eftpos expenditure to find those customers who had child related expenditure. School fees, uniforms, increase in stationary at the beginning of school terms, toys and the size of the grocery bill all give an indication of family size.

The financial institution did not hold up-to-date income information. Unless a customer has a home loan with the financial institution, they do not check this information as Baycorp is used for credit checking purposes. The financial institution therefore estimate income by extrapolating a customer's account credits over the year. This method tends to inflate income as it double counts some credits to accounts such as transfers between customer's own accounts. Also irregular income, which happens to be credited to an account during the spot month used by the financial institution to calculate income, will be classed as regular income and added on each month there after. Income was therefore recalculated for this study by only including regular credits to accounts over a period of six months prior to the surveys.

The age variables for the couple took various different forms, maximum age and minimum age of the joint account customers and large age difference (if age difference was greater then 15 years). There were only 28 instances where the age gap is greater than 15 years, amongst the 2453 customers analysed in 2002. All were checked and appeared to be genuine married-type relationships based on address and relationship type indicators. Perhaps these were couples that were in a second relationship.

The second component was lifestyle, which was introduced to add another dimension to the life stages. A person's lifestyle is 'what they do' and 'where they live' so we used categories of expenditure and an urban/rural area variable as surrogates for this component. Ross Honeywill (2001) believes that consumers can be divided into two groups depending on their personalities and subsequent spending. 'I-cons' are more individual, high value customers who spend more on restaurants, books, clothes, personal wine, home improvements and so on. Traditionals tend to spend less in these areas potentially leading to lower spending patterns on credit cards and seeking fewer personal loans. This in turn would result in less interest earned by the financial institution from these customers. He believes that they have different spending patterns and can't be identified through traditional demographics. Lifestyle was assessed by looking at customers' expenditure in categories such as food, entertainment, health, home improvements and so on (see appendix 3). The expenditure information was gathered from customers' eftpos and credit card expenditure for a six-month period prior to the questionnaire. We took this as being indicative of their main expenditure as the majority (98%) of personal customers in the sample stated that the financial institution was their main financial institution.

The urban area data (city, town or rural) was derived from the 2001 NZ census using mesh blocks. Mesh blocks are grids 'laid' across the country consisting of 200 homes. These mesh blocks segment New Zealand census information. Only 70.39% of the financial institution customers used in the analysis had a mesh block. Also as mesh blocks were allocated to the financial institution data in December 2002, any customers who changed address before December 2002 will not have a valid mesh block for the time of the questionnaire (June 2002). Those who did change address were flagged and this variable was also included in the customer details component. Factor analysis was used to find a small number of powerful factors, to represent lifestyle, after imputing any missing values using multiple iteration.

The third component detailed the financial institution products held at the time of the questionnaire. We also looked at products held 3 months prior and noted any differences. We would expect that a decrease in the number of products would coincide with lower commitment levels and vice versa. Two variables were created for the number of accounts closed and the number of accounts opened during the previous three months. This section also included details of fixed home loans and any recent (three months prior) or imminent (three months post) maturity of such loans. The maturity of a home loan (whether amortization or a fixed term) gives the customer the chance to move financial institutions without penalty. This may lead to lower commitment levels at this potential repurchase stage. Similarly the recent or imminent maturity of long-term term deposits was flagged.

The fourth component noted transaction patterns, channel usage and average account balances per month. We considered the number (count) and the amount of transactions over the previous six months. A number of derived variables were formed: (i) ratios of transactions by transaction type, from May to June noting whether they increased, decreased or were unchanged; (ii) comparison of the maximum and minimum transaction types, both counts and amounts, for the first quarter compared to the second quarter, again noting increases, no change, decreases; (iii) the raw values (count and amount) for differences between May and June for each transactions type.

Customer details were the fifth and final component, which included information from the financial institution's contact log history between January and June, complaints, tenure, warnings, educational level and number of customers in the original managed group. Research shows that the more contact a customer has with the business then the closer the relationship between the two. Other authors, (Szymanski and Henard, 2001) also believe that dissatisfaction will increase as complaints increase. A limitation to using complaints to the financial institution is that a complaint is only logged if the official form is completed or if the contact centre is phoned. A complaint made to a teller will not be logged unless the form is filled out. These contacts are summed over the 6-month period at group level. The number of contacts for each group for the period January to June was used in this study.

As the financial institution did not hold information on a customer's educational level, we again used mesh blocks and census information to provide an indication of educational levels. The following rules were used to determine three educational levels, namely high school or less, higher school or vocational qualifications and finally university graduate. First we calculated the educational norms from the 2001 census data, creating proportions for each educational level across the New Zealand population (see appendix 4). The proportions for educational level were also calculated for each individual mesh block. Comparing the population norms to the mesh blocks, if the mesh block had a larger proportion than the population norm for a particular educational level then that is the level of education allotted to that particular mesh block. There was no occasion where more than one of the educational levels was greater than the norms for their mesh blocks. All customers who had mesh blocks were allocated an educational level using this method; multiple iteration was used to impute educational level for those 29.61% who did not have a mesh block. In 2003 22.65% did not have mesh blocks.

Once the data was extracted and manipulated the next stage was to predict commitment level using Enterprise Miner's modelling techniques. Decision trees, logistic regression and neural networks were used to classify responders into the appropriate commitment levels.

2.6.3 Comparative Analysis

There were three distinct groups of comparisons, which could be made using the 2002 and 2003 surveys.

This section compared the following three groups of customers:

1. All 'married-type' couples in 2002 and all 'married-type' couples in 2003
2. Postal versus email respondents in 2003
3. Common customers to both the 2002 and the 2003 surveys

Firstly, we tested the SEM from 2002 using the data from the 2003 questionnaire, for both the Conversion model and the extended model. We then calculated commitment levels for the 2003 couples using the factor weights and cut-off points from the 2002 analysis for consistency sake. We scored the 2003 customers in Enterprise Miner using the 'best' classification model from 2002, in order to validate this model and establish if our commitment measure is stable.

Independent t-tests and non-parametric Wilcoxon tests were used to compare the commitment scores for all useable 2002/2003 data. Commitment levels for the two surveys were compared as were life stage variables using chi-squared tests. Life stage and commitment levels were compared to establish whether these variables are significantly different for different levels of commitment. This was performed for 2002 and 2003 separately.

Secondly, the 2003 survey consisted of 58% postal and 42% emailed questionnaires, which we used to compare commitment levels by delivery type. We were also interested in comparing response rates and effective response rates of the two delivery methods. Dillman (1978) and Hansen (1980) have described methods of comparing response rates and response completeness. Response rates are typically conservatively computed by using Dillman's method of:

Response rate = total number of useable responses/(total sample size – undelivered)

Response completeness for closed-ended questions can be assessed using the mean number of unanswered questions (Hanson, 1980) and comparisons made.

We performed independent t-tests and non-parametric Wilcoxon tests on the commitment scores by delivery method to determine whether there was a significant association between the type of delivery method and commitment score. Commitment levels were tested for significance to delivery method using chi-squared tests. We compared delivery method and the significant variables from the 2002 commitment model to establish if there were any significant differences between methods of delivery for these variables. We also noted which method of delivery had the highest commitment level/score by comparing mean values. Finally, we compared commitment of common customers in 2002 and 2003. Paired t-tests for means were used; as were non-parametric Wilcoxon signed ranked tests for population differences to compare their commitment scores in 2002 versus 2003.

Assumptions for the Wilcoxon signed ranked tests are that the differences are continuous and symmetrical. The test statistic is calculated by obtaining the absolute differences and then ranking them. To each rank, attach a plus sign or a negative sign according to whether the difference is positive or negative, respectively. Any differences which are zero are discarded. Sum the signed ranks and denote as T. We then use a standardised test statistic as follows:

$$Z^* = \frac{T - 0}{\sigma\{T\}}$$

Where:

$$\sigma\{T\} \qquad = \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

when the population median difference is 0, z* follows approximately a standard normal distribution. The null hypothesis for this test is the population median difference is zero, while the alternative hypothesis can be one or two tailed (Neter et al, 1988).

Commitment levels were also compared for the common customers using a sign test which will allow us to test only those customers who either decreased or increased commitment level over the year, those with no change could be ignored for this test. This test is called the Cochran test and is detailed below.

H0: No difference between the two years
H1: there is a difference between the two years

Observation frequency: n_1 = positive change, n_2 = negative change

Expected frequency under H0: $\frac{n_1 + n_2}{2}$ for both positive and negative change in commitment.

Statistic: $$\chi_1^2 = \frac{\left(n_1 - \frac{n_1 + n_2}{2}\right)^2 + \left(n_2 - \frac{n_1 + n_2}{2}\right)^2}{\frac{n_1 + n_2}{2}}$$

We performed paired t-tests on the significant continuous predictor variables from the 2002 regression model for commitment. We wanted to find if there were any variables, which were significantly different for these common customers over the year.

We also tested whether there was a significant decrease in commitment for those customers who decreased their commitment level over the year. We did this by using a one-sample t-test. Finally, for those customers who did decrease their commitment we used the Wilcoxon tests to find if there were any variables which were significantly different in distribution from customers who did not decrease their commitment. We were particularly interested in those customers with decreased commitment, as this is likely to be the segment that the financial institution would be most concerned about.

Chapter 3 Results

3.1 Introduction

This chapter describes the derivation of commitment level and a model for these levels based on transactional and demographic data. A comprehensive analysis is then performed for the 2002 and 2003 data.

The stages used to develop a discrete ordinal commitment scale were: (see figure 2.1)

1. Code the responses to the relevant commitment questions from the questionnaire to create component scores to be used in structural equation modelling.
2. Test Hofmeyr and Rice's 'Conversion' model for commitment and an extended commitment model to establish standardised regression weights.
3. use the standardised regression weights from SEM to produce logical commitment levels in a rather ad hoc manner.
4. Factor analysis to create a continuous commitment score
5. Compare the factor scores and preliminary commitment levels to confirm the reliability of the boundaries of the ad hoc method using the SEM weights.

These stages are discussed in section 3.2. In section 3.3 we discuss creating a model to predict commitment using demographic and transactional data. Data mining techniques were used for this modelling. In section 3.4 we validate the 2002 models using the 2003 data. We also compare the commitment levels for the 2002 and 2003 survey, the email versus postal method of delivery and the commitment levels of the common customers in the two surveys.

3.2 Commitment Level Derivation from the Questionnaires

Stage one of this research established a commitment level for each couple using only their responses to questions from the 2002 survey.

3.2.1 Preparing the questionnaire responses

Four component scores are included in the Conversion model, namely satisfaction with brand, ambivalence, perception of alternatives and importance of brand choice. The extended model also included inertia and strength of relationship. Each component score in the models was created in a slightly different manner as described below.

Satisfaction component scores

Satisfaction with brand was assessed using two questions, which were measured on a 5-point scale ranging from (5) excellent to (1) poor. The first question required customers to rate the overall quality of service provided by their personal financial provider. The second asked customers to rate their overall satisfaction with the financial institution.

The mean of the two questions was calculated and allocated to the satisfaction component. If the mean was greater than 4 then this component code was 3 (low score). If the mean was greater than or equal to 2 and less than or equal to 4 then the component code for satisfaction was 2 (medium). Finally, if the mean was less than 2 then the component code was 1 (high).

Inertia component scores

The inertia question contained three statements, all measured on a five-point agree/disagree scale, ranging from (5) 'strongly agree' to (1) 'strongly disagree'. The statements used were 'Changing financial institutions is a real hassle', 'I have thought recently about changing financial institutions' and 'I like to look around for the best interest rates'. This component was calculated in the same way as satisfaction but with the exception that the scales of the two negatively worded questions were reversed so that the interpretation of the scales was consistent.

Strength of relationship component scores

Strength of relationship consisted of the question 'Who would you consider to be your main Financial institution?' We allocated a score of one if 'the financial institution' was chosen, otherwise the score allocated was zero. We also used 'Would you recommend the financial institution in the future to friends or relatives for their financial needs?' this was a 5 point scale ranging from (5) definitely would to (1) definitely would not. The mean of the two questions was calculated. If the mean was less than 2 then the component code for strength of relationship was 1 (low), if the mean was 2 then the component code was 2 (medium) and if the mean was greater than 2 then the component code was 3 (high).

Importance of brand component scores

There was no specific question to assess the importance of brand to the customer, so a surrogate question was used: 'All financial institutions are the same?' This was a five-point agree/disagree scale, ranging from (5) 'strongly agree' to (1) 'strongly disagree'. Presuming that a customer who strongly agreed does not find the choice of brand important. We allocated scores as follows: if the response was 1 or 2 then the component code for strength of relationship was 1 (low), if the response was 3 then the component code was 2 (medium) and if the response was 4 or 5 then the component code was 3 (high).

Ambivalence component scores

Finally, the ambivalence question consisted of three options designed to identify how undecided the customer feels about continuing to use the financial institution as a provider of financial services. This component also consisted of one question on a three-point scale. As the scale was in reverse order of the previous scales, the raw responses were reversed to create this component code. If the response was 1 then the component code was 3 (high score); a response of 2 was coded 2 (medium score) and a response of 3 had a code of 1 (low score).

Perception of alternatives component scores

The perception of alternatives question was a simple ten-point assessment scale for six New Zealand registered financial institutions including the financial institution. Response categories ranged from (10) 'perfect in every way' to (1) 'completely unsatisfactory'. This component was quite different from the others. The set of questions used were designed to establish how many other financial institutions, from a sample of five major financial institutions, the customer thought were as good as or better than the financial institution. If a customer thought that 4 or 5 financial institutions were better than or equal to the financial institution then the component code was 1 (low score). If they thought 2 or 3 were better or equal then the component code was 2 (medium score) and finally if 0 or 1 were better to or equal then the code was 3 (high score).

The above six component scores could be used to create a commitment scale. Cronbach's Alpha was used to test the reliability of this scale. The Cronbach's Alpha test for these components gave a standardized coefficient alpha of 0.6384. Although values of above 0.7 are typical of reliable scales, Santos (1999) notes that values below this are often used in literature. As this is exploratory research an alpha value of 0.6384 suggests a reasonable correlation between the six components, indicating that this questionnaire is a reliable source of commitment information.

3.2.2 Structural Equation modelling

The component scores were then used as inputs for the observed variables in the structural equation models (SEM). SEM allowed us to identify which of the two theoretical models fitted the data more closely. The components and weights for the 'best' model were then used to create a commitment level for each respondent.

The starting point was to test the 'Conversion' model with its four components, namely importance of brand choice, perception of alternative, satisfaction with brand and ambivalence. Next we tested an extended model, which consisted of the four components from the Conversion model with the addition of strength of relationship (Barnes, 1997) and inertia (Papatla and Krishnamurthi, 1992).

(a) Conversion Model

The path diagram used for the recursive structural equation model for the 'Conversion model' can be seen in figure 3.1

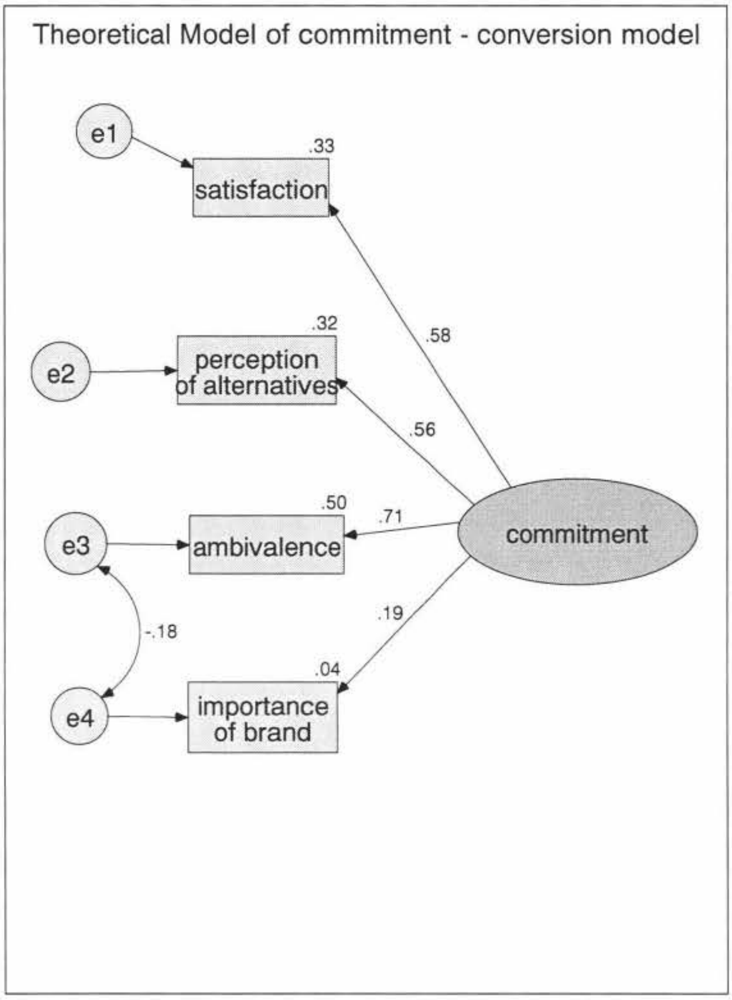

Figure 3.1 Theoretical model using the Conversion model components (2002)

The rectangles on the path diagram in figure 3.1 are the observed component scores from the questionnaire, while the oval is an unobserved variable. The standardized critical ratios for the weights along the arrows from the commitment variables show that all of the components were significant with the exception of 'importance of brand'. The figures at the top right hand corner of the rectangles show the proportion of variance of the component, which is explained by the latent variable commitment. The poor performance of the 'importance of brand' component could be attributed to the fact that the questionnaire did not include a specific question for this component. We used a substitute question for this component, which was not specific enough to fully describe the nature of the component. The only question available was a 5-point agree/disagree scale 'Are all financial institutions the same?'

(b) Extended Model

The model is shown in figure 3.2.

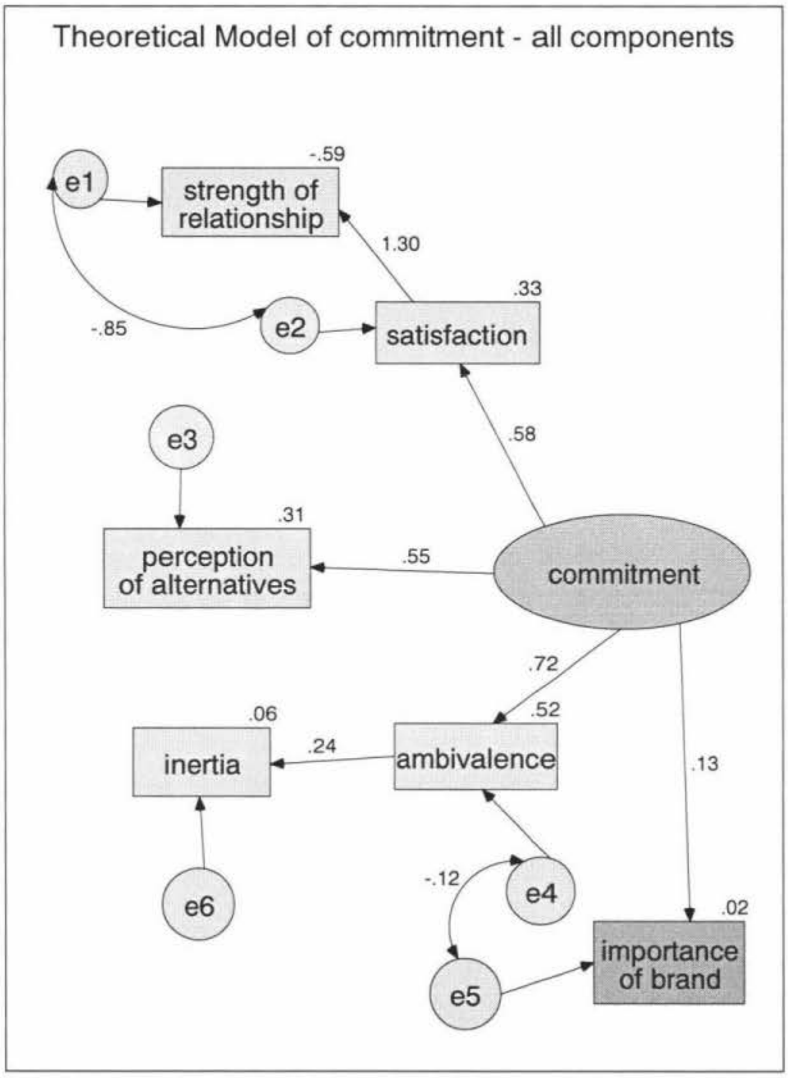

Figure 3.2 Theoretical Extended model using all six components (2002)

We were justified in allowing strength of relationship to be correlated with satisfaction, as one of the variables included in the strength of relationship component asked whether the customer would recommend the financial institution to friends or family. We would generally expect a customer to be satisfied with the financial institution before they would be willing to recommend it to family and friends. This is especially true in financial institutions as trying a new financial institution is a more involved process compared to, say, trying a new brand of washing powder on the recommendation of someone close.

The correlation between ambivalence and importance of brand can also be justified. Ambivalence is defined as having 'mixed feelings towards something' (Oxford Dictionary, 1991) and the importance of brand component consisted of the question, 'Are all financial institutions the same?' If a customer states that they strongly agree that all financial institutions are the same then we would expect them to be ambivalent in their choice of financial institution. Those who strongly disagree that all financial institutions are the same would probably be less ambivalent about their choice of financial institutions.

The values above the observed variables in figure 3.2 are the squared multiple correlations. This is the proportion of the variable's variance accounted for by its predictors. Importance of brand only had 2% of its variance accounted for, while inertia only had 6% of its variation accounted for in the model. These two components are weak in this analysis, probably as a result of inadequate questions on the survey. The figures on the arrows represent the standardised weights for the components and again importance of brand and inertia are not significant.

3.2.3 Measures of fit for structural equation models

When using structural equation modelling the default method of estimation is maximum likelihood. This method produces estimates with very desirable properties. Maximum likelihood estimation (MLE) is an alternative to ordinary least squares used in multiple regression. It is a procedure that iteratively improves estimates to minimise a specified fit function, in other words minimizing the deviance. MLE is efficient and unbiased when the assumption of multivariate normality is met.

One method to diagnose the presence of distributional problems is to use bootstrap simulations. Firstly the SEM is run using the maximum likelihood estimation method and then the same model is re-run using 1000 bootstrap simulations. Bootstrapping is a form of resampling in which the original data are repeatedly sampled with replacement. Parameter estimates and standard errors are based on empirical observations. If there are significant differences between the standard errors of the maximum likelihood method and the bootstrapping then this would indicate that there are distribution concerns suggesting that alternative methods of estimation such as asymptotically distribution-free (ADF) estimation should be used. Both of the theory-based models were assessed in this way using AMOS.

(a) Conversion model

The standard errors for each component were compared using maximum likelihood estimates and bootstrap methods.

Table 3.1 Maximum Likelihood Estimates – Conversion model

Regression Weights:	S.E.
satis_sc <------ commit	
percep of alt<-- commit	0.066
ambiv_sc <------ commit	0.072
import of brand<-commit	0.054
Variances:	**S.E.**
commit	0.009
e2	0.012
e3	0.010
e4	0.015
e1	0.009

All of the bootstrap samples could successfully be computed and were included in the estimation process.

Table 3.2 Bootstrap Standard Errors – Conversion model

Regression Weights:	S.E.	S.E. S.E.	Bias	S.E. Bias
satis_sc <------ commit	0.000	0.000	0.000	0.000
percep of alt <-- commit	0.071	0.002	0.002	0.002
ambiv_sc <------ commit	0.072	0.002	0.002	0.002
import of brand <-commit	0.057	0.001	-0.002	0.002

Variances:	S.E.	S.E. S.E.	Bias	S.E. Bias
commit	0.009	0.000	0.000	0.000
e2	0.014	0.000	-0.000	0.000
e3	0.011	0.000	-0.001	0.000
e4	0.015	0.000	-0.001	0.000
e1	0.008	0.000	-0.000	0.000

As can been seen from the above tables the standard errors for the regression weights for the maximum likelihood and the bootstrapping were almost identical with no significant differences. S.E. S.E. in table 3.2 is the approximate standard error of the bootstrap standard error estimate. All estimates are small for this column, which is as it should be. Bias in table 3.2 is the difference between the bootstrap mean and the original estimate. This is also small which is good. S.E. Bias is the standard error of the bias, which again is very small.

These results suggest that the data is approximately multivariate normal in distribution so we can use the maximum likelihood method of estimation for our analyses.

(b) Extended model

The same procedure for testing multivariate normality was used for the extended model.

Table 3.3 Maximum Likelihood Estimates – Extended model

Regression Weights:	S.E.
ambiv_sc <------ commit	0.049
satis_sc <------ commit	0.049
str_rel_ <---- satis_sc	0.066
inert_sc <---- ambiv_sc	0.010
percep of alt <-- commit	
import of brand <-commit	0.048

Variances:	S.E.
commit	0.011
e4	0.008
e2	0.008
e6	0.002
e3	0.010
e1	0.047
e5	0.015

Table 3.4 Bootstrap Standard Errors – Extended model

Regression Weights:	S.E.	S.E. S.E.	Bias	S.E. Bias
ambiv_sc <------ commit	0.055	0.001	0.002	0.002
satis_sc <------ commit	0.051	0.001	0.002	0.002
str_rel_ <---- satis_sc	0.067	0.001	0.001	0.002
inert_sc <---- ambiv_sc	0.012	0.000	-0.000	0.000
percep of alt <-- commit	0.000	0.000	0.000	0.000
import of brand <-commit	0.049	0.001	-0.001	0.002

Variances:	S.E.	S.E. S.E.	Bias	S.E. Bias
commit	0.011	0.000	0.000	0.000
e4	0.009	0.000	-0.001	0.000
e2	0.008	0.000	-0.000	0.000
e6	0.005	0.000	-0.000	0.000
e3	0.013	0.000	-0.000	0.000
e1	0.049	0.001	0.002	0.002
e5	0.015	0.000	-0.001	0.000

As before it seems that the assumption of multivariate normality is supported.

Assuming normality of the variables we can check the significance of the regression weights for our models by looking at the critical ratios for the unstandardised results. If the ratios are greater than 2, in absolute value, then the weights are significantly different from zero.

(a) Conversion model

Table 3.5 Critical Ratios for the Conversion model (Maximum likelihood estimates)

```
Regression Weights:                                C.R.
-------------------                              --------

             satis_sc <------------commit         17.641
             percep of alt <-------commit
             ambiv_sc <------------commit         16.108
             import of brand <-----commit          6.259
```

(b) Extended model

Table 3.6 Critical Ratios for the extended model (Maximum likelihood estimates)

```
Regression Weights:                                C.R.
-------------------                              -------
             ambiv_sc <----------- commit         22.363
             satis_sc <----------- commit         17.962
             str_rel_ <----------- satis_sc       23.171
             inert_sc <----------- ambiv_sc       12.441
             percept of alt<-------commit
             import of brand <---- commit          5.458
```

Tables 3.5 and 3.6 show that all components have a critical ratio greater then 2. Perception of alternatives has no critical ratio as it was assigned a value of 1 for its regression weight in the model building stage as it was thought to be the most important component for commitment. It would be difficult for a customer to be committed to the financial institution if they believe other financial institutions are better. The model cannot therefore test the significance of the relationship for this variable.

When assessing the fit of structural equation models, we checked only the measures shown in Table 3.7. In this research we cannot use the chi-squared statistic as it is not reliable for sample sizes greater than 200 (Babakus, 1987). The sample size was 2453 for the 2002 survey. The overall assessment of the two models can be seen in the table below; these measures are described in detail in the methodology chapter. The preferable ranges of values are taken from Hair et al (1998).

Table 3.7 SEM evaluation for 2002

	Conversion model (n = 2453)	Extended model (n = 2453)	Preferable acceptable range of values
Absolute fit measures:			
GFI	0.997	0.994	>0.95
RMSEA	0.073	0.046	<0.08
PCLOSE	0.102	0.674	>0.10
MECVI	0.013	0.034	lower = better
Incremental fit measures:			
AGFI	0.972	0.982	>0.90
CFI	0.987	0.999	>0.90
NFI	0.986	0.999	>0.90
TLI	0.923	0.998	>0.90

This table shows for the absolute fit measures that the extended model has a better fit. The higher the GFI the better the fit, and we require RMSEA to be preferably less than 0.05 and certainly less than 0.08. It can be seen that the extended model has a RMSEA less then 0.05 while the Conversion model is close to 0.08. The PCLOSE is not significant for the extended model which means that we cannot reject the null hypothesis that the RMSEA is no greater than 0.05. Finally for this group of measures we compare the MECVI, the lower the value the simpler the model and the better the fit. This measure shows that the Conversion model is better, but this is probably due to the simplicity of that model.

Next we compare the incremental fit measures. All of these measures for the extended model easily exceed 0.9. All these measures exceed 0.9 for the Conversion model, but the extended model has higher values for all of these measures.

A final measure to determine overall model fit is the value of standardised residuals representing the differences between observed and estimated covariance matrices. With 95% confidence the absolute values for standardised residuals should not exceed 2.58 (Baumgartner and Homburg, 1996). Both models had less than 5% of absolute standardised residuals above the 2.58 level for both models confirming a good fit.

Given the adequacy of the overall goodness-of-fit indices, no respecifications of the extended model were needed and it was concluded that the 2002 data fits the extended model better than the Conversion model.

3.2.4 Logical Commitment level from SEM

Using the standardised weights from the structural equation analysis as a guide we manually allocated commitment levels to customers. This provided a calibration tool for the final commitment scores calculated using factor analysis. As seen in section 3.2.3 the data fit for the extended model was better than the data fit for the Conversion model, so weights from the extended model were used at this stage. Because of the low significance in the structural equation modelling of 'importance of brand', we ignored this component at this stage.

Using the component codes established in section 3.2.1 and the SEM standardised weights (shown in figure 3.2) we formed component scores. This method allowed more emphasis to be given to those components with higher weights when developing the commitment levels. This was done in a rather ad hoc manner as indicated below.

Firstly the ambivalence/inertia component was determined as shown in table 3.8. The idea here was that for an inertia component code of 1 and an ambivalence component code of 1 then the overall ambivalence component score was 1 and so on. Ambivalence is better explained than inertia in the SEM so the ambivalence value tends to dominate the final score.

Table 3.8 Ambivalence/inertia Component Scores

INERTIA	AMBIVALENCE	SCORE
1	1	1
2	1	1
1	2	2
2	2	2
3	1	2
3	2	3
1	3	3
2	3	4
3	3	4

Table 3.9 shows the satisfaction/strength of relationship component scored using the same procedure. Strength of relationship tended to be dominated the final score because it had a higher standardised weight in the SEM.

Table 3.9 Satisfaction/strength of relationship component scores

STRENGTH OF RELATIONSHIP	SATISFACTION	SCORE
1	1	1
1	2	1
2	1	1
1	3	2
2	2	2
3	1	3
2	3	3
3	2	3
3	3	4

Finally the perception of alternatives was created in table 3.10.

Table 3.10 Perception of alternatives component scores

PERCEPTION OF ALTERNATIVES	SCORE
1	1
2	2
3	4

The reason a 'perception of alternatives' value of 3 was scored as 4 for the overall component, in table 3.10, was because the original question was a 3-point scale and it was believed that these customers would be highly committed. While those with a perception of alternatives score of 2 would be medium low committed as they rated 2 or 3 of the five alternative financial institutions being equal to or higher than THE financial institution, even though the financial institution was their main financial institution.

Once the component scores were calculated as indicated in table 3.8 – 3.10 we summed them to create preliminary commitment levels for each couple. If the sum of the components was 3, 4, or 5 then the ad hoc commitment level for that customer was low. If the sum of the components was 6, 7 or 8 then commitment level for that customer was medium low. An overall component score of 9 or 10 was classified as a commitment level of medium high and a component score of 11 were classified as having a high commitment level. Four commitment levels were used as initial work using three levels (low, medium and high) produced a large 'medium' category which dominated when modelling using operational/demographic data as it represented 61.4% of the total number of customers we were analysing. This is summarised in table 3.11 below:

Table 3.11 Sum of scores and ad hoc commitment levels

Sum of scores	**3-5**	**6-8**	**9-10**	**≥ 11**
Ad hoc commitment level	Low	Medium low	Medium high	High

3.2.5 Continuous commitment score from factor analysis

The next step was to create factor scores for the couples using the component codes from section 3.2.2 as the input variables. Principal component factor analysis was performed and one factor was retained.

Table 3.12 Eigenvalues for the questionnaire responses

	Eigenvalue	Difference	Proportion	Cumulative
1	1.66661864	1.50821559	1.2397	1.2397
2	0.15840305	0.20646534	0.1178	1.3576
3	-.04806229	0.04727506	-0.0358	1.3218
4	-.09533735	0.05053951	-0.0709	1.2509
5	-.14587686	0.04553196	-0.1085	1.1424
6	-.19140882		-0.1424	1.0000

Table 3.13 Factor pattern for the questionnaire

The FACTOR Procedure
Initial Factor Method: Iterated Principal Factor Analysis

Factor Pattern

	Factor1
satis_code	0.57729
percep_alt_code	0.54297
ambiv_code	0.73517
import_b_code	0.09743
str_rel_code	0.74294
inert_code	0.24464

Table 3.13 shows the most important variables were strength of relationship and ambivalence. This factor can be described as a contrast between satisfaction, perception of alternatives, ambivalence and strength of relationship versus importance of brand and inertia. The order of importance of the components in the factor analysis was the same as the order displayed by the standardised weights in structural equation modelling for the extended model.

Table 3.14 Communality Estimates

Factor1
1.7898560

Final Communality Estimates: Total = 1.789856

satis	percep_of alternatives	ambiv	import_ b	str_rel
0.33326492	0.29481288	0.54047086	0.00949306	0.55196539

```
    inertia

  0.05984893
```

The final communalities shown in table 3.14 show that the single factor only explains 29.82% of variability in the variables. One factor was retained as further analyses showed that additional factors did not significantly improve the communalities. This factor was used to compare with the standardised regression weights from Structural Equation Modeling in order to confirm boundaries for the four commitment levels. It was decided that as more factors did not improve the model, for simplicity sake we would continue with one factor, particularly as its function was only to confirm the reliability of the ad hoc method of allocating commitment levels.

3.2.6 Final commitment level: Comparison of ad hoc commitment levels and factor scores

The last stage in developing final commitment levels was to compare the SEM commitment levels with the rounded factor scores. The aim in this stage is to use the standardised weights to allocate customers to one of four commitment levels. The factor scores allowed us to check whether the SEM commitment levels were consistent, while boundaries were established using the SEM standardised weights. A one factor model was used as the order of importance of the 6 components in the factor pattern was the same as the order of importance of the SEM weights. Having compared the SEM commitment levels with the factor scores we found that there was only a slight overlap between the two methods at the boundaries, this gave us confidence that the logic we had applied in forming levels of commitment was correct. We concluded that the reliability of the ad hoc method boundaries was confirmed by the agreement with the factor scores. We now had a method of allocating a final commitment level to each couple.

The final commitment level frequencies for the 2002 sample are set out in table 3.15 below.

Table 3.15 Proportion of final commitment levels in the total 2002 sample (n = 2453)

FINAL COMMITMENT LEVEL	FREQUENCY	PERCENTAGE (%)
Low	298	12.15%
Medium Low	701	28.58%
Medium High	805	32.82%
High	649	26.46%

For the sake of consistency the commitment levels from the 2003 survey were created using exactly the same factor weighting and cut-off points as were used in 2002, as shown in table 3.16 below.

Table 3.16 Proportion of final commitment levels in 2003 sample (n = 852)

FINAL COMMITMENT LEVEL	FREQUENCY	PERCENT
Low	77	9.04%
Medium Low	241	28.29%
Medium High	280	32.86%
High	254	29.81%

This section detailed how the discrete ordinal commitment scale was created using a combination of structural equation weights and factor analysis scores. We found that the data did not fit the Conversion model (Hofmeyr and Rice, 2000) as well as the extended model. The data in 2002 produced an acceptable fit for our Extended model with all but two of the goodness-of-fit measures used to assess the SEM showing an improvement relative to the Conversion model. We created four levels of commitment, namely low, medium low, medium high and high.

The next goal was to derive a model for predicting commitment level from existing secondary data.

3.3 Commitment prediction modelling from operational and demographic data

3.3.1 Introduction

This section of the research aimed to predict commitment level, using customer data held by the financial institution. The variables used for all the analyses can be seen in appendix 2. SAS Enterprise Miner 4.1 was used for modeling, specifically using logistic regression, classification trees and neural networks. Several variations of the target variable were used: all four commitment levels (low, medium low, medium high, high) formed in section 3.2, low versus the 'rest', high versus the 'rest' and finally 'low and medium low' versus 'medium high and high'. This last target variable produced the best results and was therefore adopted.

A substantial part of this research was spent collecting data from the financial institution's data warehouse and forming derived variables for use in this modeling stage. To extract maximum power and information from the data, variables were formed using ratios, differences, counts and amounts over various time frames. The methodology chapter gives more details about these variables.

Lloyds TSB found that 5% - 10% of all of its customers change their commitment levels every quarter (Hofmeyr and Rice, 2000). This suggests that benefits and penalties should be based on a one-quarter time horizon. However, it also suggests that a commitment survey that is conducted only once a year is not frequent enough. To repeat the survey every quarter would not be appropriate because of the time required to collect and analyse responses. Customers are selected in May, the questionnaire is delivered in June, the cut-off time is the end of July, so responses are not available until the August for analysis. So overall the whole process takes at least 3-4 months, making a quarterly report difficult. This makes it imperative that the financial institution have an alternative method for predicting commitment on a more regular basis, hence the need for this modelling exercise.

3.3.2 All four commitment levels for target

Firstly a four commitment level target variable was used in backward, stepwise and forward logistic regression, but the results were very poor. The partition of the data was 60% training, 20% test data and 20% validation data. The better model (backwards logistic regression) achieved an overall hit ratio of only 41.12% of which 5.88% of low; 39.64% of medium low, 56.52% of medium high and 39.77% of high commitment customers were correctly predicted. No improvement was found using either classification trees or neural networks. In fact the classification trees were unable to create a tree, simply grouping all of the customers into one large node. The poor accuracy for low commitment customers was particularly disappointing.

3.3.3 Binary target variables

We also tried a variety of binary variables for targets. A binary target of low versus all other commitment levels gave poor results. High versus the other commitment levels also had no success. The final version of the target variable was to combine the low and medium low into one group and then medium high and high combined in a second group. This target variable proved to be more successful, with backward logistic regression producing the 'best' regression model.

The profit matrix used for this binary target allowed different misclassification errors to incur different penalties (Elkan, 2001). We chose a profit matrix rather than a cost matrix so that we could work from a common baseline. A baseline is described by Elkan as '... the state of the agent before it takes a decision regarding an example. After the agent has made the decision, if it is better off, its benefit is positive otherwise the benefit is negative.' He goes on to say that 'An opportunity cost is a foregone benefit i.e. a missed opportunity rather than an actual penalty.'

The matrix used in this research can be seen in table 3.17 below.

Table 3.17 Profit matrix for backward logistical regression

	PREDICTED LOW/MED LOW	**PREDICTED HIGH/MEDIUM HIGH**
ACTUAL LOW/MEDIUM LOW	\$500	-\$500
ACTUAL HIGH/MEDIUM HIGH	-\$10	\$10

The figures were chosen for this matrix on the following grounds. Predicting high/medium high commitment into low/medium low commitment causes the unnecessary cost of an interaction (to increase commitment level) and may annoy the customer, as they are already committed. For this error a penalty of \$10 was awarded.

Predicting a low/medium low committed customer into the high/medium high commitment segment could mean the loss of that customer, as they are not then contacted in an attempt to increase their commitment level. The average revenue made from a customer with a home loan is \$2000 p.a. Based on a quarterly figure, as suggested by the Lloyds TSB finding this would be a lost opportunity of \$500 for those with a home loan. For this error a penalty of \$500 was therefore awarded.

Predicting a low/medium low committed customer into the low/medium low segment has a non-trivial benefit as this may prevent the defection of the customer, as immediate action can be taken to increase their commitment. The benefit is potentially the continuation of the $500 per quarter of the home loan for the foreseeable future. For this correct decision a benefit of $500 was awarded.

Predicting high/medium high committed customer into high/medium high segment has a benefit of $10, as we do not need to contact this customer and therefore save on the cost of contacting this customer in the short-term (quarter). Of course we cannot ignore this customer indefinitely as they may well become less committed in the future but at this point in time they can be left alone.

This target variable produced a viable model from a business point of view, as the financial institution would still be interested in the low/medium low grouping for commitment enhancing interaction. The hit ratio of the backward logistic regression model was 67.66% with 47.15% of low/medium low correctly classified and 81.77% of medium high/high correctly classified, see Table 3.18.

Table 3.18 Classification of customers for the backward regression model

1. frequency 2. percentage correctly classified	**To: high/medium high**	**To: low/medium low**	**TOTAL**
From: high/medium high	1189 81.77%	265 18.23%	1454
From: low/medium low	528 52.85%	471 47.15%	999
TOTAL	1717	736	**2453**

To assess whether this model is better than what would be achieved by chance, Hair et al (1998) recommend comparing the hit ratio to the proportional chance-based criteria. The hit ratio is the percentage of customers correctly classified by the model. The proportional chance criteria, used when group sizes are unequal, is calculated as follows:

$$C_{PRO} = p^2 + \left(1 - p^2\right)$$

where p = proportion of customers in low/medium low

1 – p = proportion of customers in medium high/high

Having calculated the proportion chance criterion, Hair (1998) goes on to say the classification accuracy of the model to be assessed should be at least one-fourth greater than achieved by chance.

For this model the proportion chance criterion is as follows:

$$p = \frac{999}{2453} = 0.407 \text{ therefore } 1 - p = 0.593$$

$$C_{PRO} = 0.407^2 + 0.593^2 = 0.517$$

This chance criterion would therefore produce the following confusion matrix (see table 3.19).

Table 3.19 By chance confusion matrix for 2002

	To: high/medium high	**To: low/medium low**	**TOTAL**
From: high/medium high	864	590	1454
From: low/medium low	592	407	999
TOTAL	1456	997	**2453**

Using this by chance confusion matrix and the profit matrix previously mentioned (table 3.17) we worked out the profit if no model was used.

Profit = (864 x \$10) + (590 x -\$10) + (592 x -\$500) + (407 x \$500)
= -\$91,260

Acceptable classification accuracy would therefore be greater than 1.25 x 0.517 = 0.6463 (64.63%). As the hit ratio for the model was 67.66% we can therefore conclude that the predictions are better than by chance.

If we work out the profit/loss using this model we have:

Profit = (1189 x \$10) + (265 x -\$10) + (528 x -\$500) + (470 x \$500)
= -\$1,760

Clearly, using the model has the potential to avoid loss of $89,500. As well as having an acceptable hit ratio the significant variables identified by the backward logistic regression made sense and included some of the life stage variables. We also checked the residuals for this model, which appeared to be acceptable, see appendix 5 for significant variables, statistics and residuals. The regression plots of residuals versus the predicted target variable are typical for binary response data, since there are only two possible outcomes and they are predicted by a logistic curve. Note that the plots for the frequency of the residuals looked like dual normal distributions as expected for a binary target variable.

Classification trees for this target variable were unable to distinguish between the two commitment levels and no tree was produced. Two approaches were used (i) all variables were used as inputs (ii) only those variables, which were significant in the backward logistic regression were used as inputs. Both methods were unable to find any differences amongst the customers and simply grouped everyone together. Finally a neural network with two hidden nodes was tested, using only those input variables which were significant in the backward logistic regression. There were insufficient observations to have three hidden nodes, even when the data partition allocated 80% of the observations to the training node.

The test and validation misclassification rates for the neural network with two hidden nodes appeared to be slightly better than the backward logistic regression as shown in table 3.20

Table 3.20 Comparing the backward logistic regression and a neural network with two hidden nodes.

MODEL	root ase	valid: root ase	test: root ase	misclass rate	valid: misclass	test: misclass
Backward regression	0.4271	0.5446	0.5330	0.2774	0.4415	0.4352
Neural network 2 hidden nodes	0.4926	0.4888	0.4921	0.4142	0.3928	0.4120

However, the cumulative lift chart (figure 3.3) shows that the models are almost identical with in fact the backward regression model slightly better from the 40th decile to the 85th. Note that the neural network with two nodes is hidden beneath the other models in the chart.

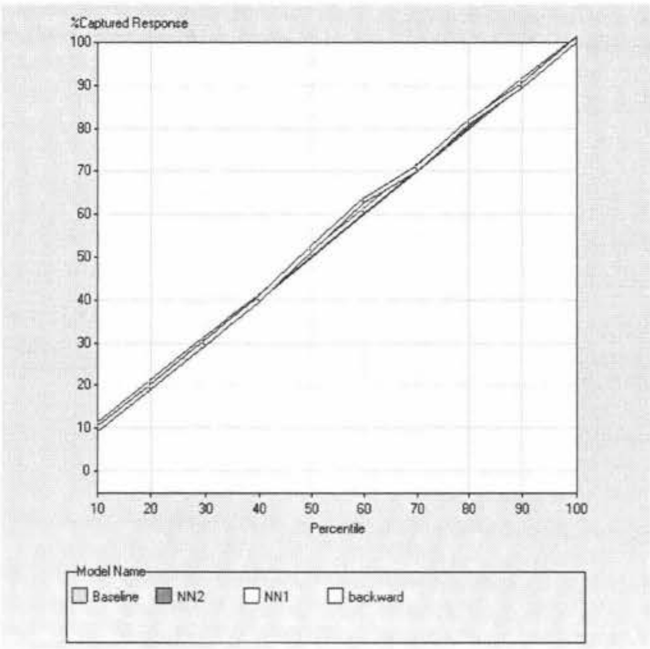

Figure 3.3 Cumulative Lift chart

When the two nodes were examined there appeared to be no distinction between the two commitment levels. The scatter plot in figure 3.4 shows that neither hidden node can distinguish between low/medium low and high/medium high commitment (H11 = hidden node 1 and H12 = hidden node 2). The red dots are the low commitment customers while the blue are high commitment customers.

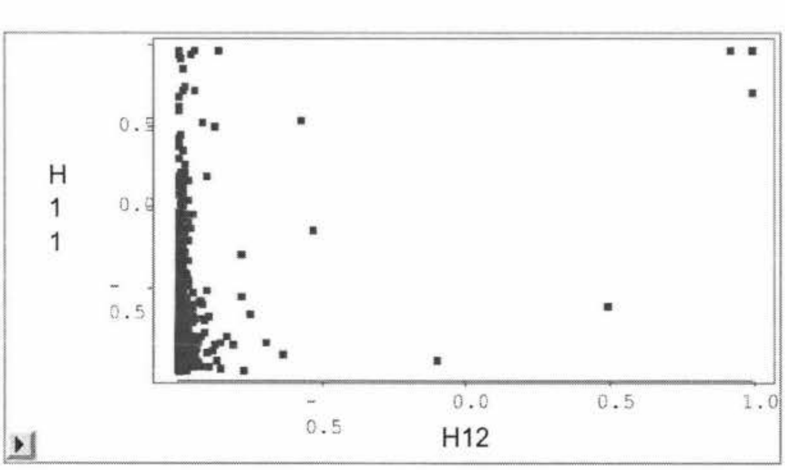

Figure 3.4 Scatter plot for the two hidden nodes by commitment level

The box plots in figure 3.5 of the hidden nodes versus the four commitment levels gave similar results.

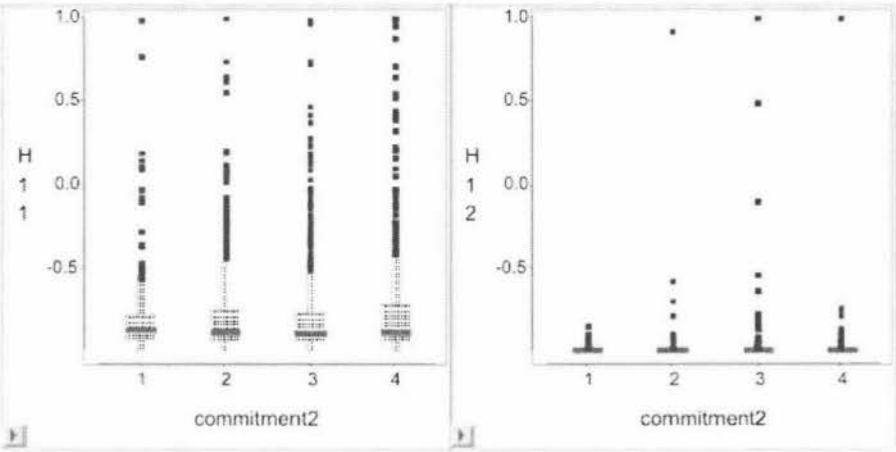

Figure 3.5 Box plots for hidden nodes and the four commitment levels

These box plots show that the first hidden node (H11) does not distinguish between the four commitment levels (all are approximately the same). The same is true for the second hidden node (H12). Although the distributions are different between the hidden nodes, they are the same within.

Interpretation of the hidden nodes in terms of the input variable also proved impossible. We concluded therefore that neural networks were of little use for predicting this target variable, and that the simpler logistic regression approach should be used instead.

3.3.4 Summarising the findings of the backward logistic regression model

The estimated parameters and t-values for this backward model can be seen in appendix 6.

The significant variables identified by backward logistic regression were interpreted as follows; these interpretations are based on the signs of the coefficients. Those couples whose maximum age was under 35 years were more committed than the other age groups, whether or not they had children. Middle and mature customers (maximum age of couple 36 years to 65 years) were less committed, than those over 65 years old.

Surprisingly customers who opened accounts in the three months prior to the survey were less committed. But, as expected customers who closed accounts in the three months prior to the survey were less committed than those who did not close accounts.

Customers who had been in contact with the financial institution in the six months prior to the survey (eg. enquires or financial appraisals whether customer or financial institution initiated) were more committed than those who had no contact. If a customer had made a complaint in the previous six months then they were less committed. Couples who had 2 or 3 other relationships with the financial institution were more committed than those with a maximum of one other relationship. 'Other relationships' are joint accounts with people other than their partner in the 'married-type' relationship, so their relationship with the financial institution is more complicated and this may be seen as a barrier to exit for the customer.

Those customers who have had a long-term term deposit mature in the three months prior to the survey or about to mature three months after the survey were less committed. These customers now had the opportunity to consider other financial institutions without facing a penalty on their term deposit, so their commitment to the financial institution could potentially fall. Customers who had four term lending products were more committed than those with only one or two term lending products. This is, of course, as one would expect.

Customers with at least one credit card were more committed. As the minimum age of the couple increases then their commitment level increases. The majority of couples (75%) had an age gap of less than 5 years and 95% of all couples had an age gap of 10 years or less, so the minimum age of the couple is probably a surrogate variable for the age of both partners. Customers with higher school qualifications or vocational qualifications are more committed. The fewer long-term savings products a customer has, the lower the commitment. A decrease in the maximum number of transactions through their transaction account in the previous quarter is related to lower commitment. This could be a sign that the customer is starting to take his business elsewhere. Another variable similar to this is an increase in debit transactions at the time of the survey. This is also associated with lower commitment. The customer could be taking his money from the financial institution and transferring it to another financial institution – a possible sign of imminent defection.

Although it was difficult to understand the meaning of the following variables in relation to commitment level, it is really this behaviour in conjunction with the previous behaviour, which is important. Those customers who did not reduce their automatic payments or direct debits in the three months prior to the survey were less committed than those who did. Those customers whose credit card credit balance increased in the last month before the survey were less committed than those whose credit balance did not increase. Those customers, whose eftpos usage increased in the last month before the survey, were less committed than those whose usage did not increase. An increase in credit card credit balance is related to lower commitment. Finally, an increase in the debit balance of a term loan at the time of the survey indicated lower commitment.

This section has established that commitment level can be predicted with some success from transactional and demographic data only. The significant variables mostly made sense and included the life stage variables. We have found that the 'best' model for predicting commitment from transactional/demographic data held by the financial institution was produced using backward logistic regression using a binary target variable of 'low and medium low' versus 'medium high and high' for commitment. Classification trees where unable to produce a tree while neural networks could not produce interpretable results in terms of the two hidden nodes.

3.4 Comparative analysis for the 2002 and 2003 questionnaires

3.4.1 Introduction

This section focuses on:

- Firstly, checking whether the 2003 data fit the extended SEM model from 2002. Allocated commitment levels to the 2003 customers and validated the 2002 backward regression model by scoring the customers in Enterprise Miner.

- Comparison of commitment scores/levels of all 2002 and 2003 respondents who qualified for the 'married-type' status. This was done using independent t-tests, non-parametric Wilcoxon tests and chi-squared tests.

- Comparison of commitment scores for email and postal 'married-type' 2003 respondents. This was done using independent t-tests, non-parametric Wilcoxon tests and chi-squared tests.

- Checking whether the sample of 223 common customers used in both the 2002 and the 2003 analysis was representative of the original 2453 customers used in the 2002 analysis. We want to use this sample to test for the significance of commitment change. This result can only be extrapolated to the research population if the sample is representative of this population. This was established by comparing the means of the significant variables from the 'best' backward regression commitment model, as we wanted them to be representative in respect to commitment.

- For the common 2002/ 2003 'married-type' couples, comparisons were made for commitment scores /levels and decrease in commitment levels. Two sample paired t-tests were used as well as the non parametric Wilcoxon paired rank tests and the Cochran test.

3.4.2 Validation of the 2002 models using 2003 data

This section checks whether the 2003 data validates the 2002 SEM and backward logistical regression models. Table 3.21 shows the break down of responders in 2003 and how many were used in this analysis.

Table 3.21 Break down of responders in 2003

TYPE OF RESPONDER	FREQUENCY
All responders	3365
All joint accounts	1864
'Married-type' couples	1776
Useable 'married-type' couples*	852

(* usable = answered all relevant commitment questions)

We checked which of the two theoretical models used in the 2002 structural equation modelling the 2003 data supports.

Firstly, we checked the Conversion model using the exact same models as for the 2002 respondents.

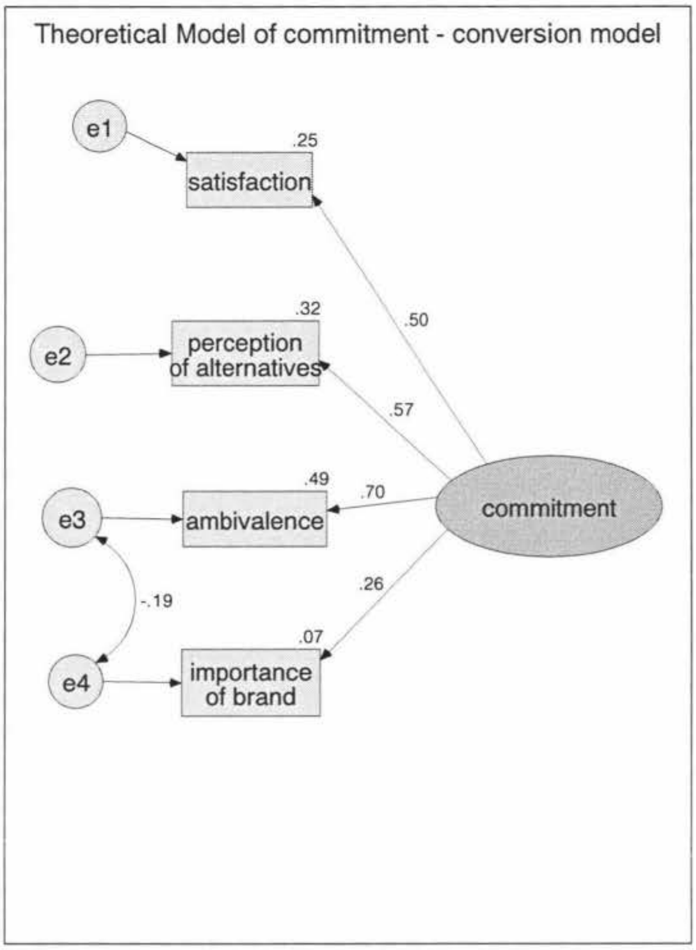

Figure 3.6 Theoretical model using the Conversion model (2003)

Figure 3.6 produced very similar standardised weights to the 2002 Conversion model diagram (figure 3.1). However, perception of alternatives has a larger weight than satisfaction, which is the opposite of the 2002 result.

We then checked the fit of the 2003 data to the extended model.

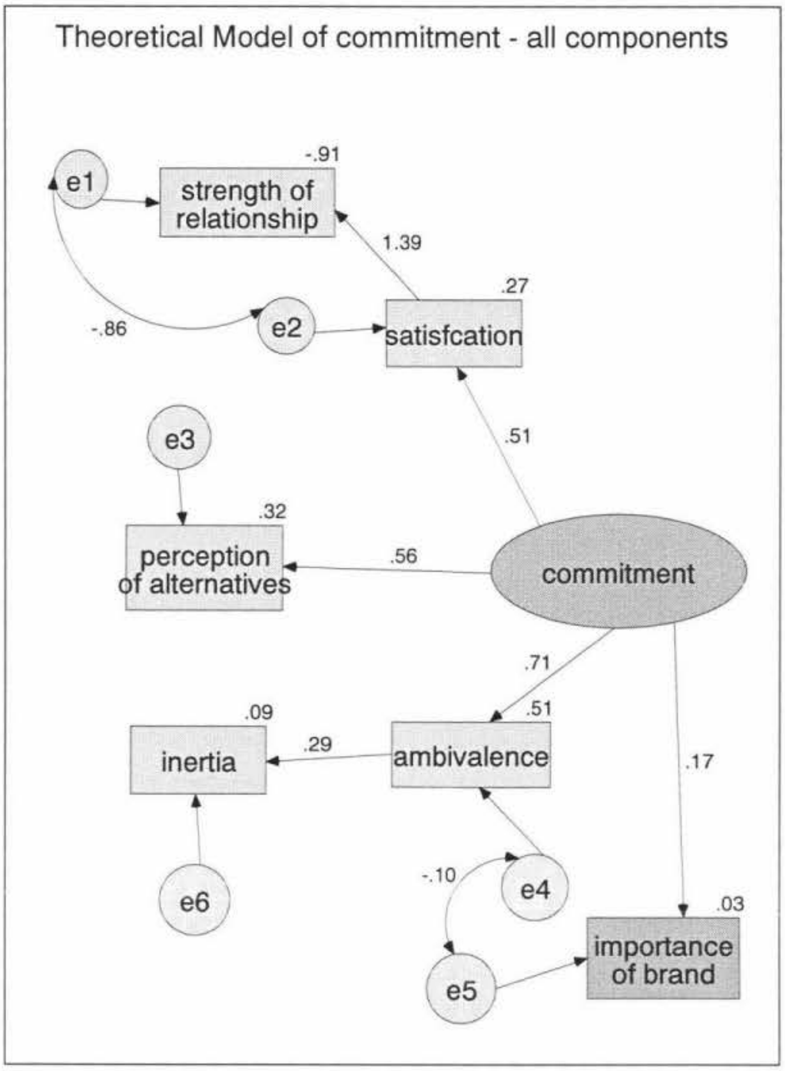

Figure 3.7 Theoretical model using extended model components (2003)

Figure 3.7 shows the extended model for the 2003 data. Again the weights are similar to those for the 2002 model (figure 3.2) but perception of alternatives has a higher weight than satisfaction.

The models are compared in table 3.22 using the same measures used for the 2002 data.

Table 3.22 SEM evaluation for 2003

	Conversion model (n = 852)	Extended model (n = 852)	Preferable acceptable range of values
Absolute fit measures:			
GFI	0.995	0.990	>0.95
RMSEA	0.093	0.056	<0.08
PCLOSE	0.074	0.289	>0.10
MECVI	0.031	0.064	lower = better
Incremental fit measures:			
AGFI	0.951	0.970	>0.90
CFI	0.976	0.975	>0.90
NFI	0.974	0.966	>0.90
TLI	0.858	0.946	>0.90

These measures show that the 2003 data fit the extended model better than the Conversion model, particularly in terms of the RMSEA and TLI. The RMSEA is above 0.08 in the case of the Conversion model suggesting a barely acceptable fit. The RMSEA of nearly 0.05 for the extended model indicates a good fit.

We then determined commitment levels for the 2003 customers using the same weights and cut-off points as in 2002 for the sake of consistency. The commitment level distribution of these 852 usable 'married-type' couples was shown earlier in Table 3.15.

To validate the 2002 classification model for commitment we scored the 2003 customers. This entails running the 2003 customers and their variables through the regression model rules predicting which commitment level these new customers fall into. Table 3.23 shows the confusion matrix for this procedure.

Table 3.23 Confusion matrix for all 'married-type' useable respondents in 2003

frequency percent row pct col pct	**Into medium high/high**	**Into low/ medium low**	**Total**
From medium high/high	505 59.27 71.03 85.59	85 9.98 60.28 14.41	590 69.25
From low/ medium low	206 24.18 28.97 78.63	56 6.57 39.72 21.37	262 30.75
Total	711 83.45	141 16.55	852 100.00

This has a hit ratio of 65.85%, for this model the proportion chance criterion is as follows:

p = 0.308 therefore 1 – p = 0.692

$$C_{PRO} = 0.308^2 + 0.692^2 = 0.5737$$

Table 3.24 By chance confusion matrix for 2003

	To: high/medium high	**To: low/medium low**	**TOTAL**
From: high/medium high	408	182	590
From: low/medium low	182	80	262
TOTAL	590	262	**852**

Using this by chance confusion matrix and the profit matrix previously mentioned we worked out the profit if no model was used.

Profit = (408 x \$10) + (182 x -\$10) + (182 x -\$500) + (80 x \$500)

= -\$48,740

Acceptable accuracy for this model (Hair, 1998) would be 1.25 x 0.5737 = 0.717 (71.7%). Clearly, the hit ratio is smaller and so we have to conclude that the model for the 2003 data is not much better than what would be achieved by chance.

If we work out the profit/loss using this model we have:

Profit = (505 x \$10) + (85 x -\$10) + (206 x -\$500) + (56 x \$500)
= -\$70,800

Using the model increases the potential loss by \$22,060 compared to what could be expected by chance for these customers.

As well as an unacceptable hit ratio according to Hair's criterion and a potential loss far greater than by chance, the proportion of low/medium low correctly classified was only 21.37%. While the proportion of medium high/ high correctly classified is 85.59%. Of the 711 customers classified into the medium high/high level 71.03% are correctly classified. Of the 141 customers classified into the low/medium low group 39.72% are correctly classified.

3.4.3 Comparison of commitment for all 2002 and 2003 'married-type' couples

This comparison was between all those responders who were 'married-type' couples from the 2002 and the 2003 surveys.

Tests were then performed to compare the commitment of all the 2002 and 2003 'married-type' customers, as shown in Table 3.25

Table 3.25 Tests on all 2002 and 2003 'married-type' customers

COMPARISON	TEST
Commitment score 2002 and commitment score 2003	Independent t-tests and Wilcoxon tests
Commitment level 2002 and commitment level 2003	Chi-square tests
Life stage variables plus variables of interest and commitment level in 2002	Chi-square tests
Life stage variables and commitment level in 2003	Chi-square tests

The first comparison we were interested in was to determine if commitment scores were different between the two samples. The means of the commitment scores for each year can be seen in Table 3.26

Table 3.26 Sample statistics for commitment scores by year

	N	MEAN	STANDARD DEVIATION
2002	2453	0.00587	0.8854
2003	852	0.109155	0.8378

Simply observing the mean commitment scores for 2002 and 2003 in table 3.26 seems to imply that the 2003 group of customers have higher commitment levels. Maybe this can be partly attributed to a major advertising campaign by the financial institution in 2003 and very little advertising in 2002.

To formally assess the differences in commitment factor scores we performed independent samples t-test of the means. We are justified in using this test as the samples are large (greater than 40) and the t procedure can be used even on clearly skewed distributions (Moore and McCabe, Second Edition, 1997).

The p-value for the test was 0.003 which is significant so we can reject the H0 and conclude that the mean commitment scores for 2002 and 2003 are not the same and there is a significant difference between mean commitment scores for all 'married-

type' couples in 2002 and 2003. A non-parametric Wilcoxon test for this same data was also significant with a p-value of 0.0026 for the two-sided test.
We then wanted to determine whether commitment levels were also significantly different for 2002 and 2003 so we performed a chi-squared test of association for the two commitment levels (low/medium low and medium high/high). This test was significant (p-value = 0.000); therefore there is evidence that there is an association between commitment level and year. Table 3.27 shows the cross tabulation for this test.

Table 3.27 Cross-tabulation for commitment level versus year

	Low/medium low (2003)	**Medium high/high (2003)**	**TOTAL**
Low/medium low (2002)	**48**	**38**	**86**
High/medium high (2002)	**25**	**112**	**137**
TOTAL	**73**	**150**	**223**

If our samples are indeed randomly and independently chosen we could be confident then that there has been a significant increase in commitment from 2002 to 2003 for 'married-type' couples.

The life stage variables had 7 categories which are shown in table 3.28.

Table 3.28 Life stage variable categories

LIFE STAGE	DESCRIPTION
1	Maximum age of couple less than 35 No kids
2	Maximum age of couple less than 35 Kids
3	Maximum age of couple 36 to 50 No kids
4	Maximum age of couple 36 to 50 Kids
5	Maximum age of couple 51 to 65 No kids
6	Maximum age of couple 51 to 65 Kids
7	Maximum age of couple 65+

In order to determine which of the life stage groups were different with respect to commitment levels, we produced the following box plots.

First we consider the 2002 customers, the box plots for commitment level and life stage are shown in figure 3.8.

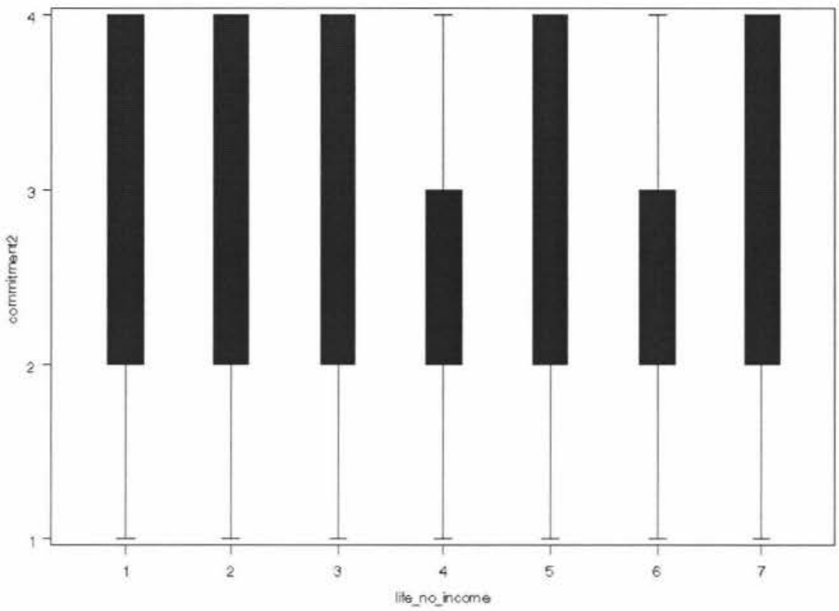

Figure 3.8 Box plots for commitment versus life stage (2002)

From Figure 3.8 it can clearly be seen that life stages 4 and 6 are different from the others. These two life stages do not have many customers with high commitment. As the life stage variable is a combination of age group and a kids indicator, we looked at both of these separately to establish which was more influential in terms of commitment. Figure 3.9 shows age and commitment level, in this case all age groups had the same distribution of commitment levels.

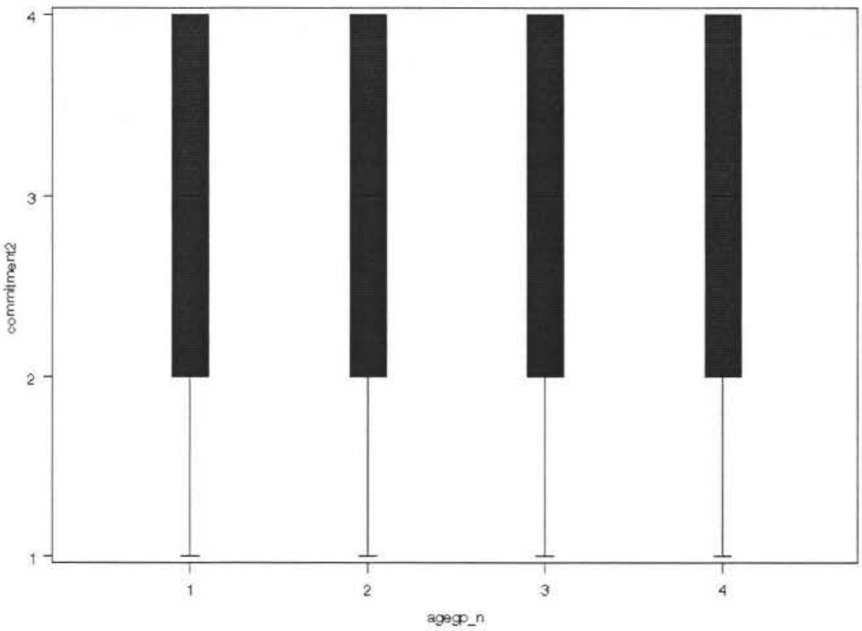

Figure 3.9 Box plots for commitment versus age group (2002)

The next figure shows commitment level and the kids indicator where '0' represents couples who probably do not have children and '1' represents couples who probably do have children.

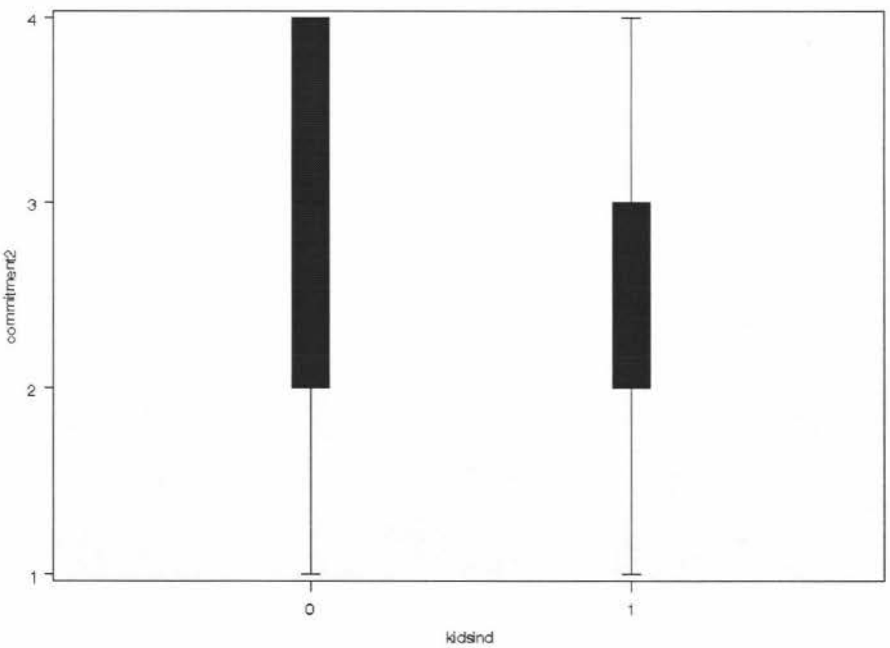

Figure 3.10 Box plots for commitment versus kids indicator (2002)

Whether the couples had children or not did make a difference as can be seen in figure 3.10. Customers with children had fewer highly committed customers.

When comparing the commitment level and life stage for the 2003 'married-type' couples we again see a difference between the seven life stages. Figure 3.11 shows the 2003 customer's commitment levels and life stage.

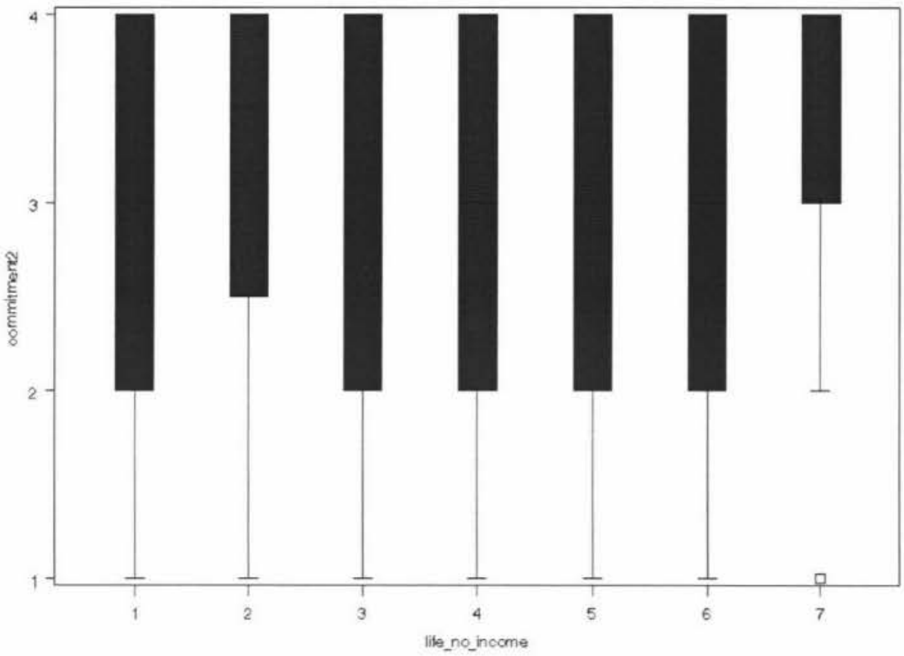

Figure 3.11 Box plots for commitment versus life stage (2003)

Life stage 2 had very few low and medium low customers compared to the other life stages. Life stage 7 was clearly different with no low committed customers and the majority high.

When we checked age and kids separately, this time age was the variable which showed some differences (see figure 3.12).

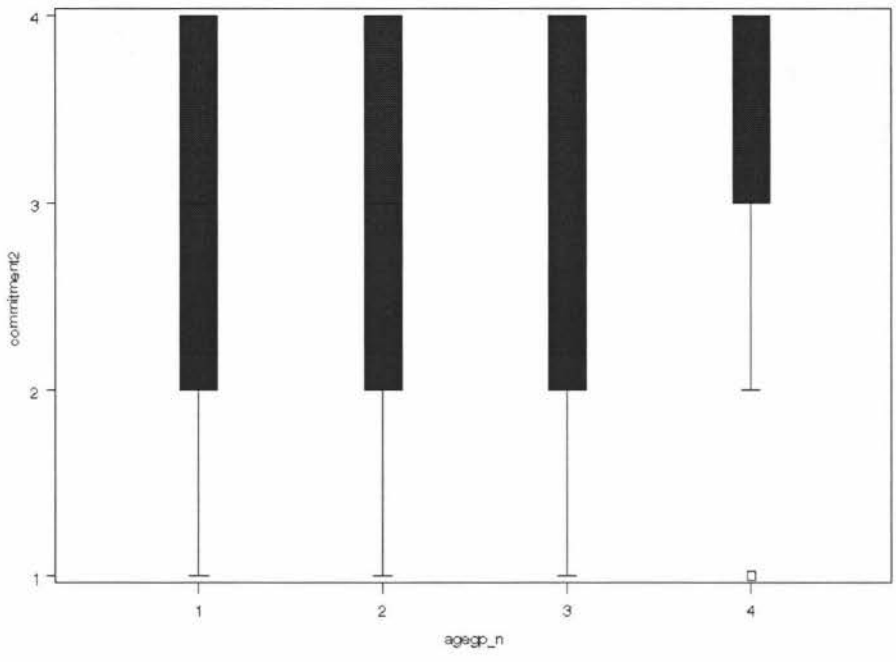

Figure 3.12 Box plots for commitment versus age group (2003)

Clearly the set of box plots in figure 3.12 show that age group 4 has mainly high committed customers. The kids variables had the same distribution of commitment scores for both categories (see figure 3.13) in 2003.

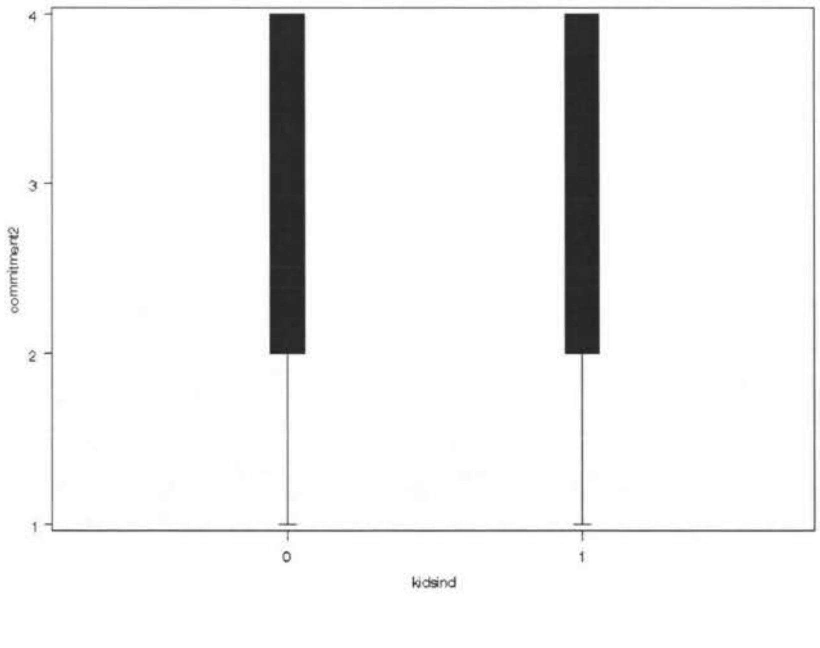

Figure 3.13 Box plots for commitment versus kids indicator (2003)

We can therefore conclude that there appears to be definite differences between commitment levels for some of the life stages. The differences appear to differ for 2002 and 2003. It seems that the two years are quite different, which suggests that regular surveys are required to track these changes.

A more formal analysis of commitment level and variables of interest was then conducted as initial evidence strongly suggested that there were differences.

Of interest to this research was the relationship between the life stage of the customer and commitment level. A one-way ANOVA was performed to assess whether commitment level is affected by the different life stages. The objective of the ANOVA is to decide whether the means of the life stages are identical.

H0: $\mu 1 = \mu 2 = \mu 3 = \mu 4 = \mu 5 = \mu 6 = \mu 7$

H1: at least two of the μ's are different

Table 3.29 shows the summary information on the commitment level and commitment scores in 2002.

Table 3.29 Summary information for commitment levels and scores (2002)

LIFE STAGE	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Sample size	209	138	650	612	496	207	141
Mean (level)*	2.69	2.63	2.76	2.67	2.79	2.71	2.94
Std dev (level)	0.99	1.00	0.96	1.00	0.99	0.97	0.92
Mean (score)**	-.02	-0.13	0.03	-0.06	0.03	0.02	0.19
Std dev (score)	0.87	0.91	0.87	0.93	0.90	0.85	0.72

(* measure on the 1-4 discrete ordinal scale)

(** factor scores)

The summary information was used to check assumptions for the ANOVA. The assumption of equal standard deviations for each life stage seems reasonable. The largest of the sample standard deviations is not greater than twice the smallest one (Devore and Peck, 2001).

We performed a number of chi-square tests for commitment level in 2002 and some variables of interest (eg revenue six months later, products six months later, life stage, income and defection six months later). We used defection six months after the questionnaire as it takes a while for a customer to close down all of their products, causing a delay before low commitment translates into defection.

Using the same two groups of commitment level from the modelling, namely low/medium low and medium high/high, the following results were obtained in Table 3.30

Table 3.30 Chi-squared tests for commitment (2002)

VARIABLE	P-VALUES	SIGNIFICANT
Life stage (no income)	0.01	Significant at 1%
Age group	0.01	Significant at 1%
Kids indicator	0.02	Significant at 5%
Income group	0.20	Not significant
Life stage (includes income groups)	0.10	Not significant
Revenue 6 months later	0.01	Significant at 1%
Defection 6 months later	0.25	Not Significant
Dropped a home loan 6 months later	0.75	Not Significant
Change in long term savings 6 months later	0.025	Significant at 5%
Change in insurance products 6 months later	0.50	Not significant

Revenue consisted of 4 groups – less than zero revenue, \$1- \$2999, \$3000-\$5999 and over \$6000. Defection consisted of 2 groups – no defection and partial plus total defection. We grouped partial (closing some accounts) with total (closed all accounts) as out of 2453 customers there were only 3 total defections 6 months later and there were 390 partial defections.

Hofmeyr and Rice found that customers who are less committed have a higher likelihood of defection. This likelihood may not translate into direct actual defection but is a factor. The chi-squared test found that commitment level and defection six months later were not significantly related for personal customers. This may because there could be a longer delay in low commitment and defection. Perhaps the relationship is more complicated for financial institutioning due to barriers to exit.

Colgate, Nguyen and Lee (2003) found in a recent New Zealand study, seven barriers to exit for service providers, namely confidence, inertia, alternatives, being locked in, social bonds, service recovery and emotional bonds.

Another explanation for the lack of significance between commitment and defection could be due to a moderating factor identified by Hofmeyr and Rice (2000) that the bigger brands retain their uncommitted customers more easily than smaller brands. This research appears to confirm that defection is not directly associated with commitment level for large well-known brands.

Although commitment level and defection was not significant, customers with low commitment had a much higher than expected defection than the other commitment levels, eg we would expect only 48 of the low committed customers to have defected (partial/total) from the chi square test, however there were 63 actual defections.

Hofmeyr and Rice (2000) also believe that the more committed a customer, the more likely they are to buy more products in the future. In terms of product take-up only home loans, long term savings and insurance products were tested for significant changes over the two years as all other product categories (eg transaction accounts, credit cards, term loans, short term savings, etc) were unchanged over the six month period.

Dowling and Uncles (1997) found that the more loyal a customer was the more revenue they generated. We found evidence that the financial institution's personal customers supported this theory, as a chi-squared test of association found that revenue six months after the questionnaire and commitment level were significant at 1%. Also Lloyds TSB found that the inclusion of commitment in the database correlated with current and future profitability (Hofmeyr and Rice, 2000).

So we can conclude that commitment level affects long term savings six months later; that commitment is not directly related to defection, but is probably a useful input variable for defection modelling; and finally that commitment is related to future revenue.

We also performed chi-squared tests the commitment levels (2 levels) of the 2003 customers and the life stage variables.

Table 3.31 Chi-squared tests for commitment versus life stage variables (2003)

VARIABLE	P-VALUE	SIGNIFICANT
Life stage (no income)	0.01	Significant at 1% level
Age group	0.02	Significant at 5% level
Kids indicator	0.05	Significant at 5% level
Income group	>0.25	Not significant

The 2003 customers gave very similar results for life stage variables. The next table shows the means of the commitment scores by life stage for 2002 and 2003

Table 3.32 Commitment scores by life stage for 2002 and 2003

LIFE STAGE	N (2002)	MEAN (2002)	STD DEV (2002)	N (2003)	MEAN (2003)	STD DEV (2003)
1	209	-0.022	0.869	58	-0.0052	0.9858
2	138	-0.1254	0.905	20	0.1800	0.8924
3	650	0.0342	0.872	230	0.1196	0.8570
4	612	-0.0556	0.925	213	-0.0286	0.8509
5	496	0.0323	0.902	185	0.1746	0.8104
6	207	0.0367	0.851	81	0.1049	0.7658
7	141	0.1929	0.724	65	0.4231	0.6274

Looking at mean commitment scores by life stages in table 3.32 we found that the highest mean by far is life stage 7 eg couples whose maximum age is 65+ for both years. The lowest score in 2002 was for life stage group 2 which represents couples whose maximum age is between 18-35 who have children and in 2003 was life stage 4 which represents couples whose maximum age is 36-50, who have children. If we compare the means for 2002 and 2003 we can see that without exception the commitment scores in 2003 are higher than in 2002. The box plots shown earlier also confirm that overall commitment levels were higher in 2003 for each life stage.

In summarising this section we found that there was a significant difference in commitment level between all 2002 and 2003 'married-type' couples. We also found that the most important associations with commitment level in 2002 were life stage, age group and revenue 6 months later. For 2003, the most important association with commitment level was life stage.

We then needed to check whether the differences between commitment was due to a change in commitment level from 2002 to 2003 for common customers or simply a change due to a different sample of customers. But first we needed to assess the affect of email for part of the 2003 survey.

3.4.4 Comparison of email and postal 'married-type' respondents from the 2003 survey.

This section compares commitment for all 'married-type' couples by delivery method of the questionnaire in 2003, namely postal or email.

There were 5198 postal questionnaires sent out in total for 2003 and 3792 sent by email. The response rate for postal was 33.42% while the response rate for email was 42.77% for all customers in the 2003 survey. However, email respondents were less likely to respond comprehensively as a very high proportion (approximately 69%) of email responders did not respond to all of the commitment questions, while only 3% to 4% of the postal responders did not answer these questions. The question that caused problems was 'How much do you agree or disagree with each of the following statements?' The statements relevant to this research which had poor completion rates were, 'Changing financial institutions is a real hassle', 'I have recently thought about changing financial institutions', 'All financial institutions are the same' and 'I like to look around for the best interest rates'. This meant that the actual response rate of usable questionnaires was lower for email responders (see methodology chapter Table 2.5). From the responders we selected those customers who were 'married-type' couples and compared commitment levels and scores for the method of delivery.

We used Hanson's (1980) method of response completeness where closed-ended questions can be assessed using the mean of unanswered questions by, in this case method of delivery. We calculated this for all closed-ended questions on the questionnaire not only the commitment ones, to get an overall view of response completeness by method. Out of a total of 29 questions the mean for postal was 1.2185 questions unanswered and for email was 4.7438 unanswered. This clearly shows that email responders answered 3.5 fewer questions on average, than postal responders.

The break down of commitment level by type of delivery is shown in Table 3.33.

Table 3.33 Commitment levels for 2003 responders by method of delivery

COMMITMENT LEVEL	POSTAL	EMAIL	TOTAL
Low	43	34	77
Medium Low	152	89	241
Medium High	197	83	280
High	181	73	254
TOTAL	573	279	852

The tests performed for this section are shown in table 3.34

Table 3.34 Tests for method of delivery

COMPARISON	TEST
Commitment scores by method of delivery	Independent t-test and Wilcoxon tests
Commitment levels by method of delivery	Chi-square tests
Compare delivery method and life stage variables for both groups	Chi-square

We firstly performed independent sample t-tests on the means of the commitment scores. As the samples were large we were able to continued to use the t-tests even if the distributions were skewed.

The t-test was significant ($p = 0.007$) for the commitment score by method of delivery. To be sure of this result we also performed the non-parametric Wilcoxon test for commitment scores. This was significant with p-value of 0.0197 so we can be confident that there is a significant difference between the commitment scores of 'married-type' couples respondents by postal questionnaire and emailed questionnaire. The sample statistics for the commitment scores are shown in table 3.35.

Table 3.35 Sample statistics for commitment scores by method of delivery

METHOD OF DELIVERY	N	MEAN	STANDARD DEVIATION
Postal	573	0.1653	0.8009
Email	279	-0.0061	0.8994

We can see from table 3.35 that the postal questionnaire has a higher mean commitment score.

A chi-squared test of association for the commitment levels gave significant results with a p-value of 0.025. We can therefore conclude that there is a significant difference between type of delivery and commitment level.

Table 3.36 shows the results of chi-squared tests for the life stage variables for customers by type of delivery.

Table 3.36 Life stage variables by delivery type

VARIABLE	P-VALUE	SIGNFICANT
Life stage (no income)	0.25	Not significant
Age group	0.25	Not significant
Kids indicator	0.90	Not significant
Income group	0.025	Significant at 5%

The results of these chi-squared tests above are not really surprising, as we would expect income to be correlated with having access to email.

To summarise this section, we found that method of delivery is associated with commitment level. It appears that the postal questionnaire respondents had the higher commitment scores. We also found that the email questionnaire had, on average, 3.5 more questions unanswered than the postal questionnaire. This indicates that the results for 2003 may have been biased by the questionnaire delivery method. This suggests that an analysis of common 2002/2003 data is required.

3.4.5 Checking whether the common customers are representative

In this section we establish whether the sample of 223 common customers from the 2002 and 2003 surveys were a representative sample of the original 2453 customers analysed in 2002.

To establish if the 223 customers common to both surveys were representative of the initial 2002 population (n = 2453) we firstly performed independent t-tests on the significant variables for predicting commitment. We compared the 2002 variables for all the 2453 customers with the 2002 variables for the 223 common customers to both surveys. If the sample were representative we would expect there to be very little significant difference between the variables for the population and the sample (see table 3.37).

Table 3.37 Independent t-tests for significant commitment variables

VARIABLE	T-VALUES	SIGNIFICANT
Change in eftpos	1.454	Not significant
Home loan scheme	3.56	Significant
Number of credit cards	0.845	Not significant
Number of long term savings products	0.963	Not significant
Number of term loans	0.309	Not significant
Number of accounts closed	0.243	Not significant
Number of accounts opened	0.324	Not significant
Reduced automatic payment or direct debits?	0.744	Not significant
Long-term term deposits maturing	0.499	Not significant
Difference between credit card credit average balance for May and June	0.505	Not significant
Number of Contacts	0.139	Not significant
Original number in managed group	0.00566	Not significant
Number of joint accounts with customers other than their partner	0.742	Not significant
Minimum age of couple	1.945	Almost significant
Age group of couple	1.212	Not significant
Life stage (no income)	1.047	Not significant
Number of complaints	0.606	Not significant

To be significant the t-values for a two-tailed test at the 5% level need to be greater then 1.96. As we can see from Table 3.37 none of the variables are significant with the exception of home loan scheme. This leads us to conclude that the sample is representative of the population. We can therefore be confident that this sample is representative of the population and any results found can be extrapolated to the whole research population.

3.4.6 Comparison of common 2002 and 2003 'married-type' couples

This section compares the 223 common 'married-type' couples at the time of the 2002 and the 2003 surveys.

The tests used in this section are shown in Table 3.38

Table 3.38 Tests for common customers to both surveys

COMPARISON	TEST
Change in commitment scores of common customers	Paired t-tests and Wilcoxon Signed Rank tests
Change in commitment levels of common customers	Cochran test
Significant variables by decreased/not decreased commitment in 2003	Wilcoxon signed rank tests

We firstly establish whether there had been significant changes in the commitment scores and commitment levels for these customers over the year. Paired t-tests were performed on the commitment scores of the common customers as can be seen in table 3.39.

Table 3.39 Sample statistics for commitment scores of the common customers

YEAR	N	MEAN	STANDARD DEVIATION
2002	223	0.1215	0.7695
2003	223	0.1964	0.7695

The two sample paired t-tests were not significant, $p = 0.1957$ for the commitment levels between the two groups. We conclude that there is no significant difference between the commitment scores of the common 2002 and 2003 'married-type' couples. We also performed the non-parametric Wilcoxon paired test on the scores, which again was not significant (p = 0.2029).

Commitment levels were tested using a Cochran test also was not significant with a p-value was greater than 0.25. We can therefore conclude that there is no significant difference between the common customers commitment in 2002 and 2003.

Table 3.40 shows the movement of customers' commitment level (4 levels) in 2003 compared to 2002.

Table 3.40 Movement in common customer commitment levels (4 levels) in 2003.

DECREASED LEVEL	NO CHANGE	INCREASED LEVEL
24.66%	46.64%	28.70%

Even though the paired t-tests and Wilcoxon rank test show that the overall change in commitment levels and scores is not significant, over 50% of customers changed commitment level. It is also of interest that approximately 25% of the common customers decreased their commitment level. This is certainly significantly different from zero customers decreasing their commitment. As this is probably the most interesting group to the financial institution, as they would want to be able to identify those customers whose commitment level is decreasing, we checked whether any transactional or demographic variables have a significant relationship with change in commitment level.

We performed a Wilcoxon signed ranked test on the variables comparing those customers who decreased or did not decrease commitment level. Table 3.41 shows the significant variables between those who decreased commitment in 2003 and those who did not.

Table 3.41 Significant variables from Wilcoxon rank tests for those customers who decreased commitment in 2003.

VARIABLE	P-VALUE
Long term term deposit maturing in 2003	0.0502
Contact in Jan to May 2003	0.0054
High school or less qualification	0.0153
Higher school qualifications include vocational study	0.0248
Other in 2002	0.0005
Contact in April to June 2003	0.0100
Term lending 2003	0.0064
Number of accounts closed 2002	0.0281
Credit card 2002	0.0073
Term lending 2002	0.0043

The most significant difference between those customers who decreased commitment level and those who did not was the number of term loans in 2002 and 2003, contact between January and May 2003 and the number of credit cards in 2002. These appear to be the most reliable predictors of decreased commitment.

To summarise, this section found that there is no significant difference between commitment levels/scores in 2002 and 2003. In addition we found that some transactional variables could be used to identify a decrease in commitment.

3.4.7 Summary of comparative results

We can summarise the results section by checking the tables of tests and their results again.

Table 3.42 Test results for all 2002 and 2003 'married-type' customers

COMPARISON	TEST	P-VALUES
Commitment score 2002 and commitment score 2003	Independent t-tests Wilcoxon tests	0.003 0.0026
Commitment level 2002 and commitment level 2003	Chi-square tests	0.05
Life stage variables plus some variables of interest and commitment level in 2002	Chi-square tests	0.01 – life stage, age group and revenue.
Life stage variables and commitment level in 2003	Chi-square tests	0.01 – life stage 0.02 - age group

There was a significant difference in commitment levels for all 2002 and all 2003 'married-type' responders. Life stage, age group and revenue were significantly related to commitment level in 2002. Life stage and age group were also significantly related to commitment level in 2003 but we were unable to check revenue, as the figures for 6 months ahead were not available.

Table 3.43 Test results for method of delivery

COMPARISON	TEST	P-VALUES
Commitment scores by method of delivery	Independent t-test Wilcoxon tests	0.007 0.0197
Commitment levels by method of delivery	Chi-square tests	0.025
Compare method of delivery and life stage variables	Chi square tests	0.025 – income group

In this section we found that method of delivery is associated with commitment level. It appeared that postal respondents had a higher commitment level than email respondents. There was evidence that income group was associated with method of delivery which is not surprising.

Table 3.44 Test results for common customers

COMPARISON	TEST	P-VALUES
Change in commitment scores of common customers	Paired t-tests Wilcoxon Signed Rank tests	0.1957 0.2029
Change in commitment levels of common customers	Sign test – possibly Cochran test	>0.25
Significant variables by decreased/not decreased commitment in 2003	Wilcoxon	0.0043 – term lending 2002 0.0054 - contact 2003 0.0064 – term lending 2003 0.0073 – credit cards 2002

Finally, there was no evidence of a significant change in commitment levels for the common customers responding to both surveys. Variables, which were significantly associated with customers who decreased their commitment level, were the number of term lending products in 2002 and 2003, the number of credit cards in 2002 and also contacts made between January and May 2003.

Chapter 4 Conclusions and recommendations

The main aim of this research was to establish whether commitment levels for the financial institution's personal customers were significantly different for the various life stages.

4.1 Findings

In this section we conclude our findings for the research questions noted in the introduction.

4.1.1 Is the Conversion model relevant for this data?

A popular model for measuring commitment is the 'Conversion' model developed by Hofmeyr and Rice, in this research we tested whether the Conversion model was relevant for this data. Structural equation modelling was used to assess whether the data from the surveys fit the Conversion model. Several goodness-of-fit measures were checked to assess the fit of the data and we concluded that the data did not fit the Conversion model well, especially for the 2003 data. In particular the RMSEA was 0.073 in 2002 and 0.093 in 2003. We therefore concluded that the Conversion model was not adequate enough model for commitment for our data.

4.1.2 Can we improve on the Conversion model?

We then tried to improve on the Conversion model by introducing two extra components, namely strength of relationship and inertia. All of the goodness-of-fit measures were acceptable for the structural equation model. In particular the RMSEA was 0.046 in 2002 and 0.056 in 2003, which are both acceptable. We could therefore conclude that although the extended model was only marginally better than the Conversion model in 2002, it was far better for 2003. The extended model was therefore adopted as the model used to form commitment levels. Strength of relationship in particular contributed to the improvement in the SEM measurements, as the standardised weight from the SEM was particularly high.

4.1.3 Formation of commitment levels

An ordinal discrete commitment scale was developed using a combination of regression weights from the SEM and factor scores from factor analysis, which used the six components of the extended model. Commitment levels were established for both the 2002 and 2003 customers using the weights and cut off points from the 2002 analysis for the sake of consistency. Four commitment levels were created, namely low, medium low, medium high and high.

The distribution of commitment levels from the 2002 survey were 12.15% low, 28.58% medium low, 32.82% medium high and 26.46% high commitment. For 2003 there were 9.04% low, 28.29% medium low, 32.86% medium high and 29.81% high commitment. We can see that there were fewer low committed customers in 2003, which may be a result of an intense advertising campaign in 2003. The mean commitment scores for the two years showed that 2003 had a higher mean commitment, while independent t-tests found that there was a significant difference between the commitment levels for 2002 and 2003.

4.1.4 Can we predict commitment levels from transactional/demographic data?

Data mining techniques were used to build a model to predict commitment levels using only the data held by the financial institution.

The 'best' model in 2002 was backward logistic regression using a binary target variable of low/medium low versus medium high/high commitment. This model had a hit ratio of 67.66%; correctly classifying 47.15% of low/medium low and 81.77% of medium high/high committed customers. Using the model was found to be better than chance which would have correctly classified 51.70% of all customers.

The 2003 survey was then used to validate this model. For this data the hit ratio of 65.85% was not sufficiently better than by chance (Hair, 1998). Using the model would incur a potential loss of $22,060 more than by chance with only 21.37% of the low/medium low committed customers correctly classified. We conclude that the 2003 data did not validate the model; the model for one year is not valid for another year.

4.1.5 Were life stage/demographic variables significant in predicting commitment levels?

The American Customer Satisfaction Index (Bryant and Cha, 1996) found that demographics affect satisfaction rating. As satisfaction is a component of commitment (Hofmeyr and Rice, 2000), this research aimed to assess the validity of these findings for the Personal customers of the financial institution.

The backward logistic regression model identified life stage and age in particular as being the most important variables for predicting commitment. Those couples whose maximum age was under 35 years, with or without children, were more committed than other life stages. While middle and mature customers (maximum age of couple 36 years to 65 years) were less committed than those over 65 years.

Other variables significant in predicting commitment were as follows.

a) home loan scheme
b) new accounts in the last three months
c) closed accounts in the last three months
d) contact with the financial institution in the last six months
e) a complaint in the last six months
f) number of joint accounts
g) long-term term deposit maturing
h) number of term lending products
i) having at least one credit card
j) minimum age of the couple
k) having higher school qualifications or vocational qualifications
l) number of long term savings products
m) decrease in the maximum number of transactions through their transaction account

n) increase in debit transactions through their transaction account
o) reduction of automatic payments or direct debits in the last three months
p) increase in credit card credit balance
q) increase in eftpos usage in previous month
r) increase in the debit balance of a term loan

Independent t-tests, Wilcoxon non-parametric tests and chi-squared tests for commitment scores for both 2002 and 2003 provided evidence that life stage, age group and whether they had children, are all significant predictors for commitment. Box plots for these three variables identified some differences between the life stages and commitment. In 2002, the kids indicator is more influential in terms of commitment, while in 2003 age is more influential. As the two years gave quite different results, this suggests that regular surveys are required to track these changes.

We can therefore conclude that this research goes some way towards supporting the work by Bryant and Cha, as demographics were associated with commitment, which in turn implies association with satisfaction levels.

4.1.6. Determine whether theories regarding commitment and product uptake, defection and profitability apply to the financial institution's personal customers.

We checked Hofmeyr and Rice's theory that the more committed a customer is to a brand the more products from that brand he will have in the future. We found that the only product to be significantly associated ($p = 0.025$) with commitment was long-term savings six months after the survey. The only other product categories to change after six months were home loans and insurance but they were not significantly associated with commitment.

Another of their theories tested was, as the level of commitment decreases the likelihood of defection increases. A chi-square test found that commitment level and defection six months later were not significantly ($p = 0.25$) related.

This research however did support the theory that commitment and revenue are associated (Werner, 2002; Dowling and Uncles, 1997).

We can therefore conclude that commitment level affects long-term savings six months later; that commitment is not directly related to defection, but is probably a useful input variable for defection modelling; and finally that commitment is related to future revenue.

4.1.7 Is type of delivery associated with commitment?

The 2003 questionnaire was delivered to customers using two methods, namely postal and email. We compared the email and postal questionnaires for all of the useable 'married-type' respondents. Kiesler and Sproull (1986) found that email is more accurate for opened ended questions but not as accurate for closed questions. This research appeared to confirm this theory; 69% of email respondents did not answer one particular closed ended question, while only 4% of postal respondents did not answer. The question required customers to respond on a five-point agree/disagree scale to the following statements: 'Changing financial institutions is a real hassle', 'I have recently thought about changing financial institutions', 'All financial institutions are the same' and 'I like to look around for the best interest rates'. This finding also agrees with Metha and Sivadas (1995) who believed that emails might be more difficult to complete. Although email had a better response rate than postal (42.77% compared to 33.42%, respectively) the effective rate was lower.

We can conclude that when comparing the two methods of delivery we found that there was a significant difference in commitment levels for the two methods, with postal questionnaires having the higher commitment levels. This means that the results for the 2003 survey were biased as a result of the use of email in this survey.

4.1.8 Was there a significant change in commitment levels of common customers between surveys? Is there any significant difference between those customers who decreased commitment in 2003 from those who did not?

We compared the common 'married-type' customers to both the 2002 and 2003 surveys. This allowed us to check commitment changes and whether life stage was associated with those who did and did not decrease commitment level. This research found that there was no significant change in commitment levels for the common customers. We found that the previous years number of credit cards, the number of term loans in both 2002 and 2003, and contacts in the 5 months prior to the questionnaire were all significant in identifying those customers who decreased commitment. Life stage however was not significantly associated with those customers who decreased commitment.

To summarise, this research found that the extended model was preferred. Commitment levels were established and modelling was able to predict these levels with some success in 2002, however the 2003 data did not validate the model. The significant variables identified by the 2002 model to predict commitment are summarised in table 4.1.

Table 4.1 Summary of significant variables for commitment prediction modelling

LESS COMMITTED CUSTOMERS	MORE COMMITTED CUSTOMERS
Age 36 years to 65 years	Under 35 years
Closed accounts in 3 months prior	Contact in previous 6 months
Opened accounts in 3 months prior	2 or 3 other joint accounts with someone other than the 'married-type' partner
Home loan scheme	4 term deposits
Complaint in previous 6 months	Having a credit card
Long-term term deposits matured	Higher school or vocational qualifications
Low minimum age of couple	
Few long term savings products	
Decrease in maximum transactions in second quarter	
Increase in debit transactions at time of questionnaire	

There was a significant difference between commitment in 2002 and 2003 for all 'married-type' customers with life stage being significantly associated with commitment in both 2002 and 2003.

Email had a poor effective response rate, while postal respondents had higher commitment levels. This indicates that the results for 2003 may have been biased by the questionnaire delivery method, suggesting that an analysis of common 2002/2003 data was required.

The 223 common customers were representative of the original 2453 customers used in the 2002 analysis so we were justified in studying this group in more detail. There was no significant difference found between commitment levels for the common customers from the 2002 and 2003 surveys. Of these customers 24.66% had a decreased level of commitment, 46.64% did not change and 28.70% had increased commitment. No life stage variables were associated with those customers who decreased commitment.

As the main aim was to establish whether life stage variables effect commitment, we can conclude that this research has shown that a definite association exists. We also identified which are the most useful classification variables for commitment.

4.2 Recommendations and future research

Although the results presented in this thesis represent a somewhat preliminary analysis of the data from this research on customer commitment and life stage, they do indicate that there is much to be learned about the exact relationship between life stage and commitment.

The results and implications are to some extent constrained by the research method employed. Though the tests of the models yield several results that are consistent with the hypothesis, the cross-sectional design used limits the ability to rule out alternative causal inferences. The inclusion of email, as a method of delivery in 2003, appeared to bias the results. The effective response rate for email respondents was particularly poor. It is therefore recommended that the email questionnaire be assessed, as there appears to be difficulty with one set of questions in particular (69% of email responders failed to answer this question compared to 4% of postal).

As far as the Structural Equation Models are concerned, the usual warnings about the interpretation of causal models apply. For example showing that a model is superior to another or seems to satisfy the hypothesis does not 'prove' causality as other (untested) models may fit the data equally well. The next stage for research may be to test whether further improvements can be made to the extended model with the addition of more components.

Finally, we assumed that there was no measurement error in the explanatory variables when using statistical analyses techniques such as logistic regression.

Data quality was an issue in this research; in particular the accuracy of the life stage variable was questionable. The kids indicator, used as part of the life stage variable, was an approximation based on credit card and eftpos spending on items which could be associated with children. The accuracy of this variable could not be established. Generalisations were also made for the educational variables, based purely on the mesh block a customer lives with in.

There are several limitations to take into consideration for these analyses. Firstly the method of collecting the information to create commitment levels was by use of a mailed/emailed questionnaire. As mentioned earlier the usable response rates for the emailed questionnaire were far lower than those for the mailed questionnaire. Also the questionnaire was addressed to couples with no way to identify which of the two had actually completed the questionnaire. This meant that we had to base the analysis at the couple level where the data collected from the questionnaire was most likely an individuals point of view. This would bias the results as attitudes are an individual feeling and two people, no matter how close, can and often do have totally different attitudes. As we wanted to be as accurate as possible in allocating commitment levels we only included those customers who filled in all of the relevant commitments questions which may not be representative of the survey population.

Secondly, there was a problem with the actual questions used for the different six components of the commitment modelling. Importance of brand in particular could not be directly measured as no specific question had been included in the questionnaire to measure this component. This components was therefore derived from the question 'All financial institutions are the same?' thus reducing the accuracy of the measurement.

The analysis used imputed variables for a number of the variables from the database, this was a limitation but a necessity. Urban area and educational level were imputed for 29.61% of the 2002 and 22.65% of the 2003 customers. These customers did not have up-to-date mesh blocks as they had changed address since the mesh block were allocated. The mesh blocks were used to assign urban area and also used in conjunction with the 2001 census to establish a probable educational level. There was also an issue with some of the data which although available was out-of-date, this affected variables which are likely to change over time such as marital status, income and so on.

Finally, we assumed that there was no measurement error in the explanatory variables when using statistical analyses techniques such as logistic regression. The sources of measurement error were imputation of missing variables, rounding of these imputed variables, a discrete measurement scale for commitment level, creation of a 'kids indicator' variable which was determined solely by amount of spending by the couple which could be directly related to children, using the age of the older partner to represent the 'age' of the couple, and the use of only two commitment groups, namely low/medium low and high/medium high, in the modeling process.

We would recommend that a wider study be conducted across the financial institutions other customer segments. Separate studies on the other personal segments in the financial institution are needed, as we cannot be confident that the same results would be attained for any two segments. Of particular interest for future research would be the prediction of customers whose commitment decreases.

An annual commitment survey is still recommended in order to detect any low committed customers who cannot be identified from transactional/demographic data. More modelling is also required, when a more reliable life stage variable is established, which may allow the formation of a model to predict successive commitment levels. It is also recommended that the survey be sent to individuals in the future, as commitment tends to be a personal psychological feeling. Using couples for this research may have limited the results, as individual's commitment can vary considerably even within a 'married-type' relationship.

Future research could involve the impact of personality traits on customer's commitment with their main financial institution. Combining attitudinal information with database information seems the obvious route to the future. This approach adds value to a customer database and provides the marketer with a clear focus as to where the marketing spend should be concentrated. Also being able to identify when major life events have occurred and predict when they are about to occur, may also be a useful supplement to life stage.

Appendices

The appendices have been removed from this thesis due to their confidential content.

Bibliography

Almquist, E. Heaton, C and Hall, N (2002) Making CRM make money *Marketing Management*, vol11, Issue 3

American Customer Satisfaction Index website, 2003
http://www.theasci.org/what_it_measures.htm

Andreasen, A.R. (1984) Life status changes and changes in consumer preferences and satisfaction. *Journal of Consumer Research*, Vol 11, December, 784-794.

Arbuckle, J.L. and Wothke, W. (1997) *AMOS 4.0 User's Guide*. United States of America: SmallWaters

Arndt, J. (1979). Family life cycle as a determinant of size and composition of household expenditures. *Advances in Consumer Research*, Vol.6, 128-132

Assael, H. (1987) *Consumer Behaviour and Marketing Action*. Boston: Kent.

Babakus, E; Ferguson C.E. & Joreskog, K.G. (1987) The sensitivity of Confirmatory Maximum Likelihood Factor Analaysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research* Vol 14 (May), 222-228

Bagozzi, R.P. and Y, Y. (1988) 'On the evaluation of structural equation models' *Journal of the Academy of marketing science*. Spring, Vol16 No1. p 74-94

Barnes, J.G. (1997) Closeness, strength, and satisfaction: Examining the nature of relationships between providers of financial services and their retail customers *Psychology and marketing* Dec Vol 14(8) p 765-790.

Baumgartner, H and Homburg, C (1996) Applications of structural equations modeling in marketing and consumer research: a review. *International Journal of Research in marketing* 13, 139-161

Beardon, W.O; Sharma, S. and Teel, J.E. (1982) Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research*, 19, 425-430

Bentler, P.M. (1990) Comparative Fit Indexes in Structural Models. *Psychological Bulletin* March 1, Vol 107, Issue 2

Berthold, M. and Hand, D.J. (1999) *Intelligent Data Analysis: An Introduction*. Springer Heidelberg

Beatty, S.E., Kahle, L.R. and Homer, P. (1988) The involvement-commitment model: Theory and implications. *Journal of Business Research*, vol 16, 149-167

Bentler, P.M. and Bonett, D.G. (1980) Significance Tests and Goodness of fit in the Analysis of Covariance Structures. *Psychological Bulletin,* vol 88, 588-606

Berry, M.J.A and Linoff, G. (1997) *Data Mining Techniques for Marketing, Sales and Customer Support.* Canada: Wiley

Bollen, K. A. and Long, J.S. Editors (1993) *Testing Structural Equation Models.* Newbury Park, California: Sage

Brieman, L., Friedman, J., Olshen, R. and Stone, C. (1984) *Classification and Regression Trees.* Wadsworth, Pacific Grove, CA

Browne, M.W. and Cudeck, R. (1989). Simple sample Cross-validation indices for covariance structures. *Multivariate Behavioral Research*, 24, 445-455.

Bryant, B.E. and Cha, J. (1996) Crossing the threshold. Some customers are harder to please than others, so analyze satisfaction scores carefully. *Marketing Research*, Winter, pp21-28

Bryman, A., Bytheway, B., Allatt, P. & Keil, T. (1987). *Rethinking the life cycle.* London: The MacMillian Press Ltd

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S. (2003). *Applied Mulitiple Regression/Correlation Analysis for the Behavioural Sciences.* Third Edition. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Callingham, M. and Baker, T. (2002). We know what they think, but do we know what they do? *International Journal of Market Research*, vol 44 Quarter 3

Council on Financial Competition (2000) Retail Economics revisited: Undoing the myths of mass-market profitability.

Colgate, M., Nguyen, V. and Lee. C, (2003) An analysis of why New Zealanders remain with their service providers. *University of Auckland Business Review*, Volume 5 No1

Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika*. 16, 297-334

Day, G.S. (1969) A two-dimensional concept of brand loyalty. *Journal of Advertising Research* 9 (September): 29-35

De Wulf, K. and Odekerken-Schroder, G. (2003) Assessing the impact of a retailers relationship efforts on consumers' attitudes and behaviour *Journal of Retailing and Consumer Services* 10, 95-108

Devore, J and Peck, R (2001) 'Statistics: The exploration and Analysis of Data'. Fourth Edition Duxbury, Pacific Grove, CA, USA

Dick, A. and Basu, K. (1994) Customer Loyalty: Toward an integrated conceptual framework. *Journal of Academy of Marketing Science*. **22** Spring, pp 99-113

Dillon, W.R, Madden, T.J. and Firtle, N.H. (1993) *Essentials of Marketing Research.* Homewood, IL: Richard D. Irwin, Inc.

Dillman, D.A. (1978) *Mail and telephone Surveys: the total design method.* New York: Wiley

Dowling, G.R. and Uncles, M.D. (1997) Do customer loyalty programs really work? *Sloan Management Review* Summer Vol 38 Issue 4 pp 71-82

Ehrenberg, A.S.C. (1988). *Repeat-buying.* Aske Publications Ltd, London.

Elkan, C. (2001) The foundations of cost-sensitive learning. *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI'01)* Seattle, Washington, USA. 4-10 August
Groth, R. (1998) *Data Mining: A Hands-on Approach for Business Professional* Prentice Hall, New Jersey

Gulliksen, H, and Tukey, J.W. (1958) Reliability for the law of comparative judgement. *Psychometrika*, 23, 95-110.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1998) *Multivariate Data Analysis.* New Jersey: Prentice-Hall

Hansen, R. A. (1980). A self-perseption interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior. *Journal of Marketing Research,* 17, 1, pp63-8

Hardie, B.G.S. and Fader, P.S. (1999). Applied probability models in market research. *10th Annual Advanced Research Techniques Forum*, June 13-16

Hartigan, J.A. (1975) *Clustering Algorithms.* New York, Wiley

Hofmeyr, J and Rice, J (1995) Intergrating the psychological conversion model with database information to measure and manage customer equity. *Journal of Database Marketing*, 3(1).

Hofmeyr, J and Rice, B. (2000) *Commitment-Led Marketing. The key to brand profits is in the consumer's mind.* Chichester: Wiley.

Honeywill, R and Blyth, V (2001) *I-cons: The Essential Guide to Winning and Keeping High-value Customers.* Sydney: Random House

Jacoby, J. (1971) A model of multi-brand loyalty. *Journal of Market Research Society*, vol.13, Number 1: 25-31

Jacoby, J. and Chestnut, R. (1978) *Brand Loyalty measurement and management.* New York: Wiley.

Jacoby, J. and Kyner, D.B. (1973) Brand loyalty vs repeat purchasing behaviour. *Journal of marketing research*, vol X (Feb), 1-9.

Joreskog, K G. (1993) Testing Structural Equation Models. In Bollen, K. A. and Long, J.S. Editors *Testing Structural Equation Models.* Newbury Park, California: Sage

Kamakura, W. A. and Wedel, M. (2000) Factor analysis and missing data. *Journal of Marketing Research*, Vol 37, 490-498ss

Kiesler, S. and Sproull, L. S. (1986) Response effects in the electronic survey. *Public Opinion Quarterley* 50, pp 402-13

Lansing, J.B. and Morgan, J.N. (1955) Consumer Finances over the Life Cycle *Consumer Behaviour Volume II,* New York Uni Press

Mehta, R. and Sivadas, E. (1995) Comparing response rates and response content in mail versus electronic mail surveys *Journal of the Market Research Society*, October v37 n4 p429(11)

Mittal, V. and Kamakura, W.A. (2001) Satisfaction, repurchase intent and repurchase behaviour: Investigating the moderating effects of customer characteristics. *Journal of Marketing Research*, vol 38, 131-142

Moore, D.S. and McCabe, G.P. (1997) *Introduction to the Practice of Statistics.* Second Edition. New York, W.H.Freeman and Co

Morgan, R.M. and Hunt, S.D. (1994), The commitment-trust Theory of relationship Marketing, *Journal of Marketing*, 58 (July), 20-38.

Morrison, D.G. and Schmittlein, D.C. (1988) Generalizing the NBD model for customer purchases: what are the implications and is it worth the effort? *Journal of Business and Economic Statistics,* April 1988, vol 6, No2, p145-158.

Neter, J., Wasserman, W. and Whitmore, G.A. (1988). *Applied Statistics* Massachuesetts: Allyn and Bacon

Nunnaly, J. (1978) *Psychometric theory.* New York: McGraw-Hill

Oliva, T.A., Oliver, R.L. & MacMillan, I.C. (July 1992) A Catastrophe model for developing service satisfaction strategies. *Journal of Marketing*, vol. 56, 83-95.

Olsson, U. (1979). On the robustness of factor analysis against crude classification of the observations. *Multivariate behavioural Research,* 14, 485-500

Papatla, P. and Krishnamurthi, L. (1992) A probit model of choice dynamics. *Marketing Science* Vol 11 No2 Spring.

Pritchard, M.P., Havitz, M.E. and Howard, D.R. (1999) Analysing the commitment-loyalty link in service contexts. *Journal of Academy of Marketing Science*, 27 (3), 333-348

Reichheld, F.F. (1993) Loyalty-Based Management. *Harvard Business Review* (March-April) 64-73

Reinartz, W. J. and Kumar, V. (2002). The mismanagement of customer loyalty. *Harvard Business Review,* 1/7/02

Reinartz, W.J. and Kumar, V. (2000). On the profitability of long-life customers in a non-contractual setting: An empirical investigation and implications for marketing. *Journal of marketing*, vol 64 (Oct) p17-35.

Rice, B. (22-24 June 1999) Understanding Commitment: the key to customer relationship management. *Pan-Pacific Direct Marketing Conference*, Sydney Australia.

Rigdon, E.E and Ferguson, C.E. (1991) The performance of the Polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research*, November, Vol 28 Issue 4, p491-498

Robertson, T.S. (1976). Low-commitment consumer behaviour. *Journal of Advertising Research,* vol 16 Number 2 April

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: J.Wiley & Sons

Santos, J.R.A. (1999) Cronbach's Alpha: A tool for assessing the reliability of scales. *Journal of Extension* Vol 37, No2

Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data.* Suffolk, UK: Chapman and Hall.

Schafer, J.L. and Olsen, M.K. (1998) Multiple imputation of multivariate missing data problems: A data analyst's perspective. *Multivariate Behavioural Research*, 33 (4), 545-571

Schmittlein, D.C., Morrison, D.G. and Colombo, R. (1987). Counting your customers: who are they and what will they do next? *Management Science*, vol 33 No1, Jan, p1-24.

Schulz, D.E. (1998) Are we too loyal to our concept of loyalty. *Marketing News,* vol 32 Issue 13, p11

Sharma, S. (1996) *Applied Multivariate Techniques.* Wiley, USA

Sheth, J.N. and Parvatiyar, A. (1995) Relationship in Consumer Markets: Antecedents and Consequences. *Journal of the Academy of Marketing Science*, 23 (4), 255-71.

Storbacka, K. and Lehtinen, J.R. (2001) *Customer relationship management: creating competitive advantage through win-win relationship strategies.* Singapore: McGraw-Hill.

Szymanski, D.M. and Henard, D.H. (2001) Customer satisfaction: A meta-analysis of the empirical evidence. *Journal of Academic Marketing Science*, vol 29 No1 p 16-35

Teo, T.S.H. and Lim, V.K.G. (2001) The effects of perceived justice on satisfaction and behaviour intentions: the case of computer purchase. *International Journal of Retail and Distribution Management* Vol 29, Number 2, 109-124

Tse, A.C.B (1998) Comparing the response rate, response speed and response quality of two methods of sending questionnaires: e-mail vs. mail. *Journal of the Markiet Research Society,* vol 40, p353 (9)
s

Walsh, J.P., Kiesler, S., Sproull, L and Hesse, B. (1992) Self-selected and randomly selected respondents in a computer network survey. *Public Opinion Quarterly*, 56, pp 241-244.

Wells, W.D. & Gubar, G. (1966) Lifecycle concept in marketing research. *Journal of Marketing Research*, vol III, November, 355-63.

Werner, R. K. (2002) The mismanagement of customer loyalty. *Harvard Business Review* (July)

Wheat, R.D. and Morrison, D.G. (1994) Regularity, recency and rates. *European Journal of Operational Research,* 76, p283-289

Wyner, G.A. (1999) Customer relationship measurement. *Marketing Research*, Summer, Vol.11, Issue 2, 39-41.