CHAPTER 4: RESEARCH DESIGN AND METHODS

4.1 INTRODUCTION

This chapter provides a description and motivation for the research design and the methods used in this study. Aspects relating to the inquiry strategy and broad research design are, firstly, depicted in the section that follows. This is followed by a discussion of the data source utilised, the sampling and data analysis plan employed. The latter part of the chapter reflects on assessing and demonstrating the quality and rigour of the research design. Lastly, ethical considerations applicable to this study are addressed.

4.2 DESCRIPTION OF INQUIRY STRATEGY AND BROAD RESEARCH DESIGN

The following sections provide, firstly, a description of the study's general strategy of inquiry. This is followed by the basic characteristics associated with quantitative research. Lastly, a classification of the study's overall research design is presented.

4.2.1 A description of the study's strategy of inquiry

To investigate and illustrate the process of index construction and applying it as a method of market segmentation, this study utilised secondary data that were originally gathered during a survey by an independent market research organisation amongst a sample of commercial farming businesses in South Africa. A structured questionnaire was used as data gathering instrument. From the questionnaire, items were identified that on face value were believed could contribute to constructing a particular index of commercial farming sophistication. In the context of this study the availability of the data, as well as

the original questionnaire used and sampling frame targeted during the survey, are therefore considered incidental and serves solely to illustrate the process of index construction as a method of market segmentation. The strategy of inquiry is described as quantitative.

4.2.2 The basic characteristics of quantitative research

Quantitative research, in general, involves data gathering in their primary form from large numbers of individuals or sample elements, with the intention of extrapolating the results to a wider population (Tustin *et al.*, 2005:89). More specifically, quantitative research is associated with numbers as the unit of analysis with the aim of measuring or describing phenomena (Denscombe, 2007:248). When such raw, unanalysed quantitative data becomes available for re-analysis by other researchers it constitutes secondary data. Coyer & Gallo (2005:60) describe secondary data analysis as a method where data collected in another study is used to answer new research questions or use different statistical techniques.

4.2.3 A classification of the study's overall research design

The broad research design of this study can best be described by the following descriptors:

- *Empirical research:* The study analysed secondary numeric data, and can therefore be classified as an empirical study.
- *Basic research:* The aim of basic research is concerned with the advancement of theoretical conceptualisations about a particular topic (Leedy & Ormrod, 2010:44). This study was undertaken to investigate the process of index construction and applying it as method of market segmentation. This is illustrated through the construction of a commercial farming sophistication index and using it to segment the market into homogeneous groups.

- *Descriptive research:* According to Leedy and Ormrod (2010:182), descriptive quantitative research involves either identifying the characteristics of a certain observed phenomenon, or exploring possible associations among two or more phenomena. This is accomplished by examining a phenomenon, namely that of index construction and its application as a method of market segmentation, and by describing important factors associated with this process (Kelley, Clark, Brown & Sitzia, 2003:261).
- *Non-experimental:* This study did not involve changing or modifying the situation being researched, nor to determine cause-and-effect relationships (Leedy & Ormrod, 2010:182).
- *Cross-sectional:* Descriptive quantitative research is mostly concerned with the measurement at a single point in time (Kelley *et al.*, 2003:261). The original design of the survey by MSSA was cross-sectional, having gathered the data over a specific period in the latter part of 2009.
- *Secondary data:* The data utilised in this study were originally gathered during 2009 by an independent research company and further analysed as part of the process of index construction. In the context of this study the secondary data is, however, regarded an incidental source to illustrate the process of index construction and market segmentation.

4.3 SOURCE OF DATA

This section describes how, when, where, and by whom the secondary data used in this study were originally collected. The last section lists the specific questions that were identified from the original survey questionnaire for use in this study.

4.3.1 Data gathering

The data were originally gathered during 2009 by an independent research company, Marketing Surveys and Statistical Analysis (MSSA), amongst a sample of commercial farming businesses across all provinces of South Africa. A survey methodology was employed to guide the data collection process.

A structured questionnaire served as data-collection instrument. This means of data collection is well-suited for collecting data that can be quantified and used to describe the activities and conditions in commercial farming businesses pertaining to levels of sophistication.

Questionnaires were sent out via mail to the selected sample units, as identified from the sampling frame used by MSSA. A personalised covering letter explaining the purpose of the research: that participation is voluntary, ensuring confidentiality and related issues, accompanied the questionnaire. The respondents were requested to complete the questionnaire and return it by a certain date. A self-addressed, stamped return envelope was also included for the convenience of the respondent.

4.3.2 Questionnaire design

The questionnaire used during the original survey was developed with the aim of describing the commercial farming market in South Africa in terms of selective firmographic characteristics as well as certain behavioural aspects of the business. This included investigating structures, systems, strategies and practices employed across various functional areas in the farming business.

4.3.3 Question identification

While the intention of the 2009 survey was not to construct an index of commercial farming sophistication, it offered sufficient scope for the

construction of such a construct. The researcher identified a number of possible questions that were believed could contribute towards measuring farming sophistication or could be used to validate the index and derived market segments. The questions identified were:

- What is the primary farming operation (main source of income) of this farming business?
- What are the secondary branches of farming operations?
- What is the total size of the farms being managed by the farming business?
- What percentage of total income is obtained from other farming-related operations?
- What percentage of total income is obtained from other non-related farming operations?
- What is the annual total turnover of all farming activities?
- How is the primary farming business administrated?
- Does the farming business complete and submit VAT returns?
- Does the farming business manage a cash-flow budget?
- Does the farming business make use of an external/independent accountant to keep financial records?
- Does the farming business make use of any external/independent financial advisor (other than accountant)?
- Does the farming business make use of cell-phone(s) as part of managing and operating the farming operations?
- Does the farming business make use of any 3G cellular services as part of the farming business?
- Does the farming business make use of a PC as part of the farming business?
- Please indicate what the PC is used for as part of the farming business?
- Does the farming business have access, or plan to have access to the Internet in the nearby future?
- What is the main method of paying wages and salaries?
- Indicate the types of insurance paid for by the farming business?

- Does the active farmer have any agricultural related qualifications?
- What is the highest educational level of the active farmer?
- Does the active farmer have an updated will?
- Which medical insurance products has the active farmer made provision for?
- How often does the active farmer review his/her estate planning?

The data reflecting the responses related to these questions were made available to the researcher for further analysis as part of this study. Procedures and methods employed to standardise the data into a suitable format for index construction and segmentation are dealt with in Section 4.5.1.

4.4 SAMPLING

In practice, it is seldom possible to study all the members of a population that is under consideration (Thompson, 2012:1). The availability of population elements at a single time, as well as budget and time constraints, requires the use of a sample – in order to study a sample of the population. This approach was used during the sampling design of the original study. The sections that follow consider the units of analysis and observation that applied to the original survey and equally apply to the current study. The sampling frame of the original survey is subsequently defined. This detail is provided as background to the secondary data utilised in this study. Of particular importance is the sample size that was available for the current study. It is evaluated against recommended sample sizes needed for index construction.

4.4.1 Units of analysis and units of observation

Units of analysis refer to those elements that are examined, in order to create summary descriptions of all such units (Babbie, 2011:101). In the original survey the unit of analysis was defined as a commercial farming business, and

was subsequently defined similarly for the current study. In some studies, however, the units of analysis do not coincide with the units of observation, with the latter representing those elements from which the data is collected about the unit of analysis. This was the case in the original survey, with the active farmer or farm manager being targeted to provide information on the farming business.

4.4.2 Sampling frame

Although the data collected during the original survey are considered incidental in the context of this study and serves as a valid quantitative basis for index construction, the sampling frame defined as part of the original survey is briefly noted for background purposes.

A mailing list that contained the postal addresses of about 4 500 commercial farming businesses located in South Africa constituted the sampling frame of the original survey. Questionnaires were mailed to these businesses, with a total of 876 completed and valid responses received by the cut-off date. This represented an approximate response rate of 19.4%.

4.4.3 Sample size required for index construction

A number of recommendations are made in the literature regarding the minimum sample size needed when constructing an index variable. However, there seems to be a lack of agreement in this regard, with suggestion ranging between a minimum sample of 150 proposed by Hinkin (1995:973) and 300 proposed by Clark and Watson (1995:12).

The available sample database of 876 records formed the point of reference for the construction of the sophistication index and subsequent segmentation. The research design adopted for this study specified the splitting of the sample database into two sub-sets. One set was to be used for the construction of the

index, while the other would serve as a holdout sample for use as part of evaluating the reliability of measurement. The generating of the two samples was done by means of a randomised split of the original sample, while specifying the appropriate sub-sample sizes needed in each data set. Of the 876 cases in the original sample database, 600 were randomly allocated to the development sample, with the remaining 276 being allocated as holdout sample. The sample of 600 allocated for the index construction therefore met the minimum requirement of 300 suggested by Clark and Watson (1995:12).

4.5 DATA ANALYSIS

Aspects relating to the standardisation of the measurement scales are discusses, followed by a presentation of the data analysis plan for constructing the index and segmenting the market. A statistical software package, IBM SPSS (Version 20) was used for data analysis.

4.5.1 Standardisation of measurement scales

The response categories of all the questions identified from the 2009 survey were coded, so that the categories would be represented by a number of dichotomous (binary) variables, indicating the presence or absence of a specific item within the commercial farming business. The scoring system specified a zero for the absence of an item, and a one for the presence of the item. This, firstly, standardised the scale across all variables; and secondly, it met the measurement requirements for principal component analysis, which were used as part of the index construction.

4.5.2 Data analysis plan for constructing of the index

The following data analysis plan provided the guidelines for the analysis of the data in order to construct the sophistication index and apply it as a method of

market segmentation. The plan followed a multi-step approach, namely: (1) Selecting of items and defining of variables that are most likely to be indicators of commercial farming sophistication; (2) combining of variables into an index; and (3) segmentation and validating of the index.

The words ‘item’ and ‘variable’ are often used interchangeably. In this study, these two concepts have an identical meaning. A variable or item more specifically captures the presence or absence of an item in a commercial farming business.

As noted in Section 4.4.3, the sample of 876 was divided randomly into two groups of 600 and 276, respectively. The construction of the index was done on the sample of 600, while the remaining 276 were used as holdout sample to support index and segment validation.

During the first step, one-way frequency tables were produced based on the sample of 600. Frequency tables provide a proportional distribution of the various dichotomous variables, indicating the percentage of commercial farming businesses where a particular item is present. This is an important preliminary step in the process of index construction, as Babbie (2011:173) emphasises. Only variables that show sufficient variation should be used in the subsequent steps of the index construction process.

Step two involved examining the empirical relationships between the variables. The data analysis was performed as follows:

- Subjecting of the identified variables from step one to an exploratory factor analysis. More specifically, a principal component extraction method was used. Principal component analysis is a technique for finding relationships in multivariate data, with the aim of reducing the dimension of the data without serious loss of information (Porkess, 2004:193). The employment of this data-reduction technique is also supported by Galpin (2006:1), who noted that binary variables place all variables on the same scale, thus

making principal component analysis the appropriate technique to be used for index construction. The scores on the factor that explained the largest proportion of variation in the data formed the initial scores for the index. These scores, or factor loadings, are continuous with a normal distribution, and are typically standardised with a mean of zero and a standard deviation of one.

- The scores were then categorised into 10 equal-sized groups. Inspection of frequency distributions and correlation analysis provided insight into the extent of association amongst the variables and the initial index scores. Variables that showed no or little correlation were excluded from the subsequent analysis.
- A second principal component analysis was performed using the reduced list of variables as input. The scores from the factor that explained the largest proportion of variation in the data were subsequently included in a stepwise regression as dependent variable to identify a set of variables that would help explain any variation in the index scores. Stepwise regression analysis is a multivariate technique that identifies from a set of independent variables, those variables that best predict the dependent variable (Leedy & Ormrod, 2010:282). The new set of scores produced by the stepwise regression was correlated with the scores derived from the second principal component analysis. The scores from the stepwise regression formed the final commercial farming sophistication index. It should be noted that the variables included in the final stepwise regression model were not necessarily only those identified from the second principal component analysis.
- In order to form a final set of commercial farming sophistication segments, the scores from the stepwise regression were categorised into a number of equally sized groups. The proportional distribution of variables across the groups was examined, and where the proportions were similar, adjacent groups were combined. The forming of segments were thus guided by some basic statistical analysis and criteria specified by the researcher.

- Forming of segments. Lower and upper cut-off points were identified for each segment, providing segmentation rules for researchers and practitioners.

The last step in the analysis involved the segmentation and validation of the index. Validity was confirmed by firstly examining the relationship between segments across those variables that comprise it; and secondly, amongst other variables that were expected to also be correlated with the levels of sophistication.

4.6 ASSESSING AND DEMONSTRATING THE QUALITY AND RIGOUR OF THE RESEARCH DESIGN

This section considers issues relating to assessing the quality of the secondary data used as part of this study. It also addresses issues of reliability of measurement, particularly concerning the construction of the commercial farming sophistication index. The last section deals with validity of measurement.

4.6.1 Quality of secondary data

When utilising secondary data as part of a study, it is important to assess if the data are sufficient to address the research objectives (Tustin *et al.* 2005:132). In other words, an evaluation of the accuracy of such data is an essential step in the design of the study. This includes evaluating the source of the data and the quality of the data.

Coyer and Gallo (2005:62) talk about both practical and methodological issues that should be evaluated when a researcher chooses to conduct an analysis of secondary data. Firstly, the investigator must check for the fit between the original data and the new research questions. In the case of the current study, the aim of the original study was to gather data to profile the commercial

farming market in South Africa in terms of selective firmographic characteristics as well as certain behavioural aspects of the business, which included investigating the structures, systems, strategies and practices employed across various functional areas in the farming business. The construction of a sophistication index was therefore not one of the original outcomes envisaged. However, the secondary data were evaluated by the researcher to be suitable for the construction of such an index. This judgement was based on criteria that included the identification of variables that the researcher believed could contribute towards measuring farming sophistication. The question formats were such that standardisation of data was possible, which was a necessity for employing certain advanced statistical techniques as part of index construction. A fit therefore existed between the original data and the new research question.

Secondly, the operational definitions of the variables used in the original study should be applicable and relevant to the requirements of the current study. With the secondary data serving as incidental source for illustrating the process of index construction as a method of market segmentation, the definitions concerning units of analysis as well specific variable definitions were adopted for the current study.

Thirdly, the original data should be evaluated in terms of sample selection bias. Sampling bias may be defined as a condition that disturbs the randomness by which a sample is selected (Leedy & Ormrod, 2010:215). Forza (2002:164) also notes that poor sample design affects the generalisability of the results. In the original study, a mail survey was used to gather the data, where potential respondents were identified from a list that contained the postal addresses of about 4 500 commercial farming businesses located in South Africa. The sampling employed was judged to be a non-probability method, which typically restricts generalisation to the population. Exploratory analysis conducted by MSSA found that the sample data were suitable for weighting to reflect population estimates (MSSA, 2009:8). According to Statistics South Africa (2009:10), there were an approximate 39 000 commercial farming businesses in

South Africa in 2007. Other characteristics released by Statistics South Africa indicated the proportional distribution of the population across province, size of farming businesses, type of primary farming activity and how farming businesses are administered. A set of weights were calculated by MSSA so that the original sample would reflect the approximate 39 000 commercial farming business population as well as corresponding firmographic characteristics. This process would contribute towards reducing some of the sampling bias.

4.6.2 Reliability

Reliability is concerned with the accuracy and precision of a measurement procedure (Cooper & Schindler, 2003:231). In other words, a measure is considered reliable if it yields consistent results. A measure that is stable produces consistent results – with repeated measurements of the same target population and same instrument. Cooper and Schindler (2003:238) acknowledge that this is a difficult task in survey research – due to time and financial constraints when surveying the same population twice.

As was noted in Section 4.4.3, the available sample database of 876 records was split into two sub-sets, with one set used for the construction of the index, while the other served as holdout sample to evaluate the reliability of measurement. In other words, the aim was to check if similar results could be obtained when applying the outcomes from the index construction and segment formation on a separate independent sample.

Lastly, the set of weights that were calculated for the original sample to reflect Statistics South Africa population estimates, were adjusted by MSSA for the two sub-samples. Although not specified as a requisite in general for index construction, being able to present the size of each market segment based on estimated population numbers offer significant value for marketing practitioners in establishing the purchasing power and potential profits of a segment. This is also one of the key success factors of a market segmentation study and is

referred to as the measurability of market segments (Canever *et al.*, 2007:514; Freathy & O'Connell, 2000:103). This aspect is noted and illustrated in Chapter 5 as part of presenting the results.

4.6.3 Validity

Cooper and Schindler (2003:231) distinguish between two main forms of validity, namely: external and internal. External validity is concerned with the representativeness and generalisability of the results. This is largely addressed by the sampling process, and already covered in the preceding section. Internal validity looks at the ability of the measurement to measure what it is supposed to measure. Two specific forms of validity are relevant in studies where constructs are measured: content and construct validity.

Content validity considers the extent and appropriateness with which the variables provided adequate coverage of the investigative construct. Construct validity addresses the question of what is actually measured. Determination of both the content and construct validity is crucial (Cooper & Schindler, 2003:232). While the researcher identified variables from the original survey that he believed could contribute towards measuring farming sophistication, this constituted face validity. For the purpose of this study though, ensuring face validity was deemed sufficient. Future research studies might be considered where the content and construct validity of commercial farming sophistication be established.

4.7 ETHICAL CONSIDERATIONS

A number of ethical issues pertaining to this study were considered and adhered to.

- *Right to privacy and confidentiality.* This study utilised secondary data and access to personal details regarding the original survey participants were restricted. This eliminated compromising an individual's confidentiality.
- *Use of data source:* Approval was granted by MSSA to use the secondary data in further analysis with the aim of constructing an index of commercial farming sophistication. Ethical clearance was also obtained from the University of Pretoria regarding the use of the secondary data.
- *Use of the findings:* The results of this study are to be used for academic purposes only and may be published in an academic journal.

4.8 CHAPTER SUMMARY

This chapter provided a discussion of the research design and methods employed in this study. Aspects relating to data source utilised, the sampling and data analysis plan employed were addressed. The next chapter will provide a discussion of the research results, with the focus on index construction and its application in segmenting the commercial farming market.

CHAPTER 5: RESEARCH RESULTS

5.1 INTRODUCTION

This study set out to investigate and illustrate the process of index construction as a method of market segmentation. The results pertaining to the construction of such an index, namely one that measures the levels of commercial farming sophistication, as well as the processes followed to segment the market are reported in this chapter. The construction of the sophistication index utilised secondary data that were originally collected by Marketing Surveys and Statistical Analysis. The construction of the index was based on a sub-sample of 600, while the remaining sample of 246 served as holdout for index and segment validation.

5.2 INDEX CONSTRUCTION AND SEGMENTATION

The data analysis followed a multi-step approach, namely: (1) Item selection and defining of variables that are most likely to be indicators of commercial farming sophistication; (2) combining of variables into an index; and (3) segmentation and validating of the index. The sections that follow report the results for each of these steps.

5.2.1 Step 1: Item selection and standardisation of measurement scales

The response categories of all the questions identified from the 2009 survey were coded, so that the categories would be represented by a number of dichotomous (binary) variables, indicating the presence or absence of a specific item within the commercial farming business. The first step of index construction, however, involved the identification specific binary variables that were believed could contribute towards measuring farming sophistication.

Following the identification of these variables, one-way frequency tables were produced using the unweighted data set (n = 600). Only variables that showed sufficient variation were used in the subsequent steps of the index construction and are reported in Table 10.

Table 10: Frequency distribution of initial set of variables (n = 600)

Variables	No	Yes
Primary farming operation: Stock farmer	33.0%	67.0%
Have any secondary branches of farming operations	24.0%	76.0%
Operate other farming-related business	78.7%	21.3%
Primary farming operation administrated: Sole Ownership	33.8%	66.2%
Registered for VAT	3.2%	96.8%
Have cash-flow budget	27.0%	73.0%
Make use of accountant (external)	71.7%	28.3%
Use a cell-phone as part of farming operations	8.7%	91.3%
Use PC as part of farming business/farm management	27.2%	72.8%
Use PC for business management	68.5%	31.5%
Use PC for animal/irrigation management	75.2%	24.8%
Use PC for financial management	57.5%	42.5%
Use PC for VAT/tax management	60.7%	39.3%
Have Internet access	38.8%	61.2%
Pay wages via Internet	71.0%	29.0%
Have short-term insurance for farming business	12.3%	87.7%

5.2.2 Step 2: Combination of the variables into an index

Step two involved the statistical process of combining variables into an index by examining the empirical relationships between them. Various inter-relating sub-steps are distinguished and these will be discussed in the sections that follow.

5.2.2.1 Step 2.1: First principal component analysis

Firstly, the data derived from the identified variables from step one were subjected to a principal component analysis. The scores on the factor that explained the largest proportion of variation in the data formed the initial scores

for the index. These scores, or factor loadings, are continuous with a normal distribution, and are typically standardised with a mean of 0 and a standard deviation of 1.

The results obtained from the principal component analysis are reported in the tables below. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy shown in Table 11 represents an index value for comparing the magnitudes of the observed correlation coefficients with the magnitudes of the partial correlation coefficients. Pallant (2010:183) recommends a minimum value of 0.6 for a good analysis. A high value, such as the 0.82 reported in Table 11, is indicative of the fact that the correlations between pairs of variables can mostly be explained by other variables. This result suggests that one could continue with the analysis.

A further indication of whether the data used are considered suitable for such an analysis is the result of Bartlett's test of sphericity (Pallant, 2010:183). This test is used to test the hypothesis that the correlation matrix is an identity matrix. In other words, that all the diagonal terms are one and all off-diagonal terms of zero. If the resulting correlation matrix is considered an identity matrix, continuing with a principal component analysis is not advisable. However, the *p*-value of 0.000 reported in Table 11 suggests that the null hypothesis can be rejected and that the correlation matrix yielded coefficients large enough to help explain variation in the data, and assist in identifying any underlying emerging factors.

Table 11: Results from first principal component analysis - KMO measure of sampling adequacy and Bartlett's test of sphericity

Kaiser-Meyer-Olkin measure of sampling adequacy		0.820
Bartlett's test of sphericity	Approximate Chi-Square	1 657.72
	Degrees of freedom	120
	Sig.	0.000

Table 12 shows that the first factor from the analysis captured the highest proportion, namely 21.0% of the variability in the data. This is in comparison slightly higher than the 19.7% obtained from the corresponding first step during the construction of the African Response Business Sophistication Measure (African Response, 2006:21); and this was, accordingly, considered to be a positive indication of the potential for index construction.

Table 12: Results from first principal component analysis - Variance explained

	Rotation sums of squared loadings		
Factor	**Total**	**% of variance**	**Cumulative %**
1	3.368	21.0%	21.0%
2	1.351	8.4%	29.5%
3	1.291	8.1%	37.6%
4	1.171	7.3%	44.9%
5	1.108	6.9%	51.8%

Seven of the 16 variables subjected to the analysis were more closely related with the first factor, forming the basis of the newly created index. The factor loadings of these seven variables are listed in Table 13.

Table 13: Results from first principal component analysis – variables loading on first factor

Variables	Factor loadings
Use PC as part of farming business/farm management	0.786
Have Internet access	0.734
Use PC for financial management	0.734
Use PC for VAT/tax management	0.651
Use PC for business management	0.634
Pay wages via Internet	0.611
Use PC for animal/irrigation management	0.550

Lastly, reporting of the Cronbach alpha value provides a means for evaluating the internal reliability consistency of the newly created index. Pallant (2010:97) suggests that ideally, the coefficient should be above 0.7. An alpha value of

0.813 was reported, which is indicative that the seven variables forming the basis for deriving the initial scores for the index could be considered highly internally reliable.

5.2.2.2 Step 2.2: Distilling of variables for further analysis

The scores obtained from the first factor were, subsequently, categorised into 10 equally sized groups. Inspection of the frequency distributions (refer to Table 14) and results from a correlation analysis (refer to Table 15) provided insight into the extent of association amongst the variables and the initial banded index groups.

Table 14: Percentage frequency distribution of initial set of variables by 10 banded groups (n = 600)

Variables	Group:									
	1	2	3	4	5	6	7	8	9	10
Primary farming operation: Stock farmer	63.1%	89.1%	83.6%	54.2%	70.0%	56.7%	73.8%	55.0%	50.8%	75.0%
Have any secondary branches of farming operations	96.9%	50.9%	60.7%	79.7%	83.3%	71.7%	83.6%	83.3%	78.0%	68.3%
Operate other farming-related business	13.8%	10.9%	6.6%	30.5%	23.3%	23.3%	21.3%	30.0%	32.2%	21.7%
Primary farming operation administrated: Sole Ownership	86.2%	76.4%	91.8%	54.2%	70.0%	61.7%	60.7%	56.7%	50.8%	51.7%
Registered for VAT	89.2%	98.2%	95.1%	94.9%	98.3%	96.7%	96.7%	100.0%	100.0%	100.0%
Have cash-flow budget	80.0%	54.5%	55.7%	66.1%	66.7%	80.0%	82.0%	70.0%	89.8%	83.3%
Make use of accountant (external)	18.5%	45.5%	24.6%	22.0%	13.3%	28.3%	31.1%	18.3%	35.6%	48.3%
Use a cell-phone as part of farming operations	76.9%	80.0%	86.9%	93.2%	100.0%	96.7%	98.4%	93.3%	93.2%	95.0%
Use PC as part of farming business/farm management	-	-	32.8%	96.6%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Use PC for business management	-	-	-	11.9%	15.0%	35.0%	36.1%	48.3%	81.4%	88.3%
Use PC for animal/irrigation management	-	-	-	8.5%	23.3%	18.3%	32.8%	35.0%	45.8%	85.0%
Use PC for financial management	-	-	-	15.3%	25.0%	55.0%	60.7%	81.7%	91.5%	96.7%
Use PC for VAT/tax management	-	-	1.6%	18.6%	38.3%	45.0%	49.2%	71.7%	76.3%	93.3%
Have Internet access	-	1.8%	19.7%	64.4%	70.0%	76.7%	86.9%	95.0%	98.3%	100.0%
Pay wages via Internet	-	-	3.3%	3.4%	10.0%	25.0%	47.5%	51.7%	66.1%	83.3%
Have short-term insurance for farming business	100.0%	94.5%	52.5%	94.9%	88.3%	90.0%	88.5%	88.3%	91.5%	88.3%

Variables that showed no or little correlation (where p-value > 0.05) were excluded from subsequent analyses.

Table 15: Correlation analysis of initial set of variables by banded group – Spearman's rho (n = 600)

Variables	Correlation coefficient	Sig.
Primary farming operation: Stock farmer	-0.089	0.029
Have any secondary branches of farming operations	0.007	0.865
Operate other farming-related business	0.131	0.001
Primary farming operation administrated: Sole Ownership	-0.244	0.000
Registered for VAT	0.143	0.000
Have cash-flow budget	0.154	0.000
Make use of accountant (external)	0.088	0.031
Use a cell-phone as part of farming business	0.187	0.000
Use PC as part of farming business/farm management	0.763	0.000
Use PC for business management	0.635	0.000
Use PC for animal/irrigation management	0.535	0.000
Use PC for financial management	0.734	0.000
Use PC for VAT/tax management	0.649	0.000
Have Internet access	0.725	0.000
Pay wages via Internet	0.614	0.000
Have short-term insurance for farming business	0.018	0.654

5.2.2.3 Step 2.3: Second principal component analysis

A second principal component analysis was performed using the reduced list of variables as input. The scores on the first factor, which explained the largest proportion of variation, formed a second set of initial scores that could be used for the construction of the final index values as part of step 2.4.

Table 16 reports a Kaiser-Meyer-Olkin measure of sampling adequacy of 0.825 and a significance value of 0.000 for Bartlett's test of sphericity. Both these results are indicative that the data are suitable for principal component analysis.

Table 16: Results from second principal component analysis - KMO measure of sampling adequacy and Bartlett's test of sphericity

Kaiser-Meyer-Olkin measure of sampling adequacy		0.825
Bartlett's test of sphericity	Approximate Chi-Square	1 578.37
	Degrees of freedom	91
	Sig.	0.000

The first factor from the second principal component analysis captured 24.3% of the variability in the data, as shown in Table 17. This is higher than the 21.0% obtained from the first factor of the first principal component analysis, and is an improvement to the extent that the variables explain the overall variation in the data.

Table 17: Results from second principal component analysis - Variance explained

	Rotation sums of squared loadings		
Factor	**Total**	**% of variance**	**Cumulative %**
1	3.400	24.3%	24.3%
2	1.429	10.2%	34.5%
3	1.145	8.3%	42.7%
4	1.077	7.7%	50.4%

Seven of the 14 variables subjected to the analysis were associated significantly with the first factor. The factor loadings are shown in Table 18.

Table 18: Results from second principal component analysis – variables loading on first factor

Variables	**Factor loadings**
Use PC as part of farming business/farm management	0.820
Have Internet access	0.760
Use PC for financial management	0.719
Use PC for VAT/tax management	0.665
Pay wages via Internet	0.602
Use PC for business management	0.586
Use PC for animal/irrigation management	0.531

5.2.2.4 Step 2.4: Stepwise regression and deriving of the final index scores

While steps 2.1 to 2.3 assisted in the deriving of various sets of initial scores for representing levels of commercial farming sophistication, step 2.4 is concerned with deriving of a final set of index scores, as well as identifying a final set of variables that best explains the variation in the scores. The final outcome of this step would be a mathematical equation that would allow researchers to calculate scores for the commercial farming sophistication index.

In order to derive the equation and calculate the relevant weights, the scores from the first factor obtained from the second principal component analysis were subsequently subjected to a stepwise regression. Using the weighted sample to represent population estimates, the scores were specified as dependent variable, while the original set of 16 variables served as predictor variables. This resulted in a new set of scores. As expected, the new set of scores produced by the stepwise regression correlated highly with the scores from the second principal component analysis, with an adjusted R-square of 0.995 as shown in Table 19. The adjusted R-square provides an indication of the percentage variation explained in the data by the regression function, when adjusting for the number of predictor variables. A value of 1 is indicative of a perfect fit.

Table 19: Results from stepwise regression - Model summary

R	R Square	Adjusted R Square	Standard error of the estimate
0.997	0.995	0.995	0.07021459

The analysis of variance (ANOVA) was used to test the hypothesis that the coefficients included in the final model were significantly different from zero. The results, as reported in Table 20, show that the null hypothesis cannot be

accepted (p = 0.000), suggesting, therefore, that all the coefficients are significantly different from zero.

Table 20: Results from stepwise regression - ANOVA

	Sum of squares	Degrees of freedom	Mean square	F	Sig.
Regression	38 254.734	10	3 825.473	775 944.167	0.000
Residual	192.219	38 989	0.005		
Total	38 446.953	38 999			

The predicted scores calculated from the stepwise regression also formed the values for the final commercial farming sophistication index. It should be noted that the variables included in the final stepwise regression model were not necessarily the same as those identified from the second principal component analysis. The variables, which best predicted the index scores associated with commercial farming sophistication, are reported in Table 21.

Table 21: Results from stepwise regression - Coefficients

	Unstandardised coefficients		Standardised coefficients		
	B	Standard error	Beta	t	Sig.
Constant	-2.148	0.002		-906.173	0.000
Use PC as part of farming business/farm management	0.586	0.001	0.263	472.160	0.000
Use PC for financial management	0.400	0.001	0.200	422.913	0.000
Pay wages via Internet	0.369	0.001	0.165	390.553	0.000
Use PC for VAT/tax management	0.515	0.001	0.254	590.748	0.000
Use PC for animal/irrigation management	0.470	0.001	0.207	525.594	0.000
Have Internet access	0.431	0.001	0.212	397.098	0.000
Primary farming operation: Stock farmer	0.300	0.001	0.145	390.479	0.000
Use PC for business management	0.277	0.001	0.129	301.476	0.000
Use a cell-phone as part of farming business	0.303	0.001	0.085	230.231	0.000
Registered for VAT	0.305	0.002	0.055	151.241	0.000

In conclusion then, the following equation with coefficients is derived for the calculating of index scores:

Equation 1: Calculation of commercial farming sophistication index scores

(Primary farming operation: Stock farmer * 0.300453229780414)
+ (Registered for VAT * 0.304917832846113)
+ (Use a cell-phone as part of farming business * 0.302976592721769)
+ (Use PC as part of farming businesses and farm management * 0.585889113099588)
+ (Use PC for financial management * 0.399969689359916)
+ (Use PC for VAT/tax management * 0.514673779212451)
+ (Use PC for animal/irrigation management * 0.470287613864481)
+ (Use PC for business management *0.277265640843654)
+ (Have Internet access * 0.431459486184270)
+ (Pay wages via Internet * 0.368603037234416)
- 2.1476667354589.

These index scores provide a continuous scale for indicating the levels of commercial farming sophistication. Higher scores are associated with higher levels of sophistication. The next phase in the study concerns the forming and profiling of meaningful segments, in order to present homogenous groups of farming businesses.

5.2.3 Step 3: Segmentation and index validation

The last step in the analysis involved two sub-steps. Firstly, the raw index scores were examined and grouped into a meaningful number of market segments that are homogeneous in terms of farming sophistication. Secondly, the validity of the index was considered by examining the relationship between formed segments across those variables that formed the basis of the sophistication index, as well as amongst other variables that were also expected to be correlated with levels of sophistication. This last step also allowed for the profiling of the segments.

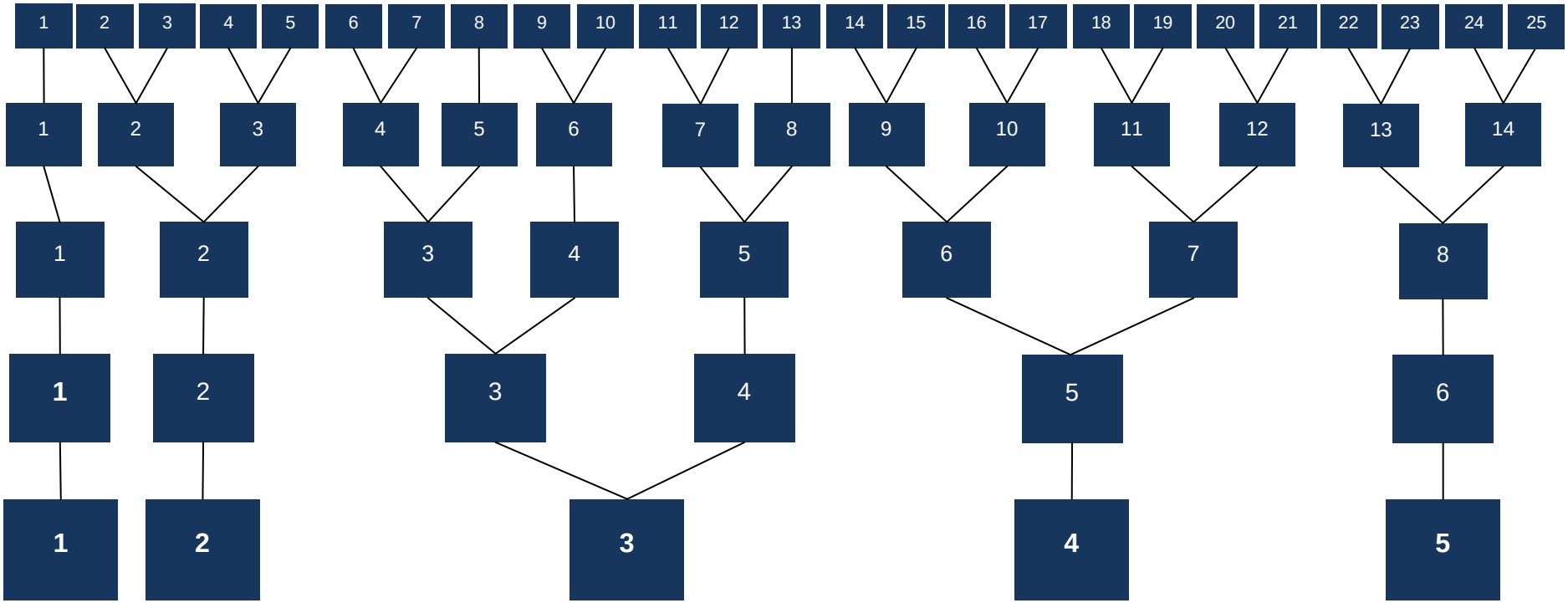
5.2.3.1 Step 3.1: Forming of segments

To form a final set of commercial farming sophistication segments, the scores obtained from the stepwise regression were categorised into a number of approximately equally sized groups, based on the weighted sample. A total of 25 groups were initially formed, which allowed for retaining sufficient group size, while allowing for examining variability in the data. Next, the proportional distribution of variables that formed the basis of the sophistication index was examined across the 25 groups. Where adjacent groups revealed similar proportional characteristics, they were collapsed to form a new reduced set of segment groups. The proportional distribution of groups was again inspected and the same process repeated. This process of evaluating adjacent groups was repeated until a final set of segments emerged. The following criteria were adopted from Jensen *et al.* (2002:92) to guide the process:

- ensuring useful discriminating power across all segments in relation to those variables that were used to construct the index, as well as other variables that are expected to be associated with the index;
- having a compact set of segments to ensure substantiality given the total estimated population of 39 000 commercial farming businesses in South Africa; preferably between four and seven segments;
- having enough segments to permit debates about sophistication thresholds relevant for a particular marketing strategy; and
- have a bottom and top segment that contains the smallest proportions of the population, which serve as the extreme groups along the index scale.

Figure 9 shows the process that was followed for final segment formation.

Figure 9: Process showing collapsing of 25 groups to form a final set of five market segments



Five segments were eventually formed. Lower and upper cut-off points were identified for the intervals that specify each segment, providing segmentation rules for researchers and practitioners alike. The intervals are:

Segment 1: Lowest through -1.542295672832370
Segment 2: -1.539772309891020 thru -0.483595582926958
Segment 3: -0.483595582926948 thru 0.347833592617228
Segment 4: 0.347833592617238 thru 1.131593949484590
Segment 5: 1.131593949484600 thru highest

Table 22 reports the final estimated segment sizes in the South African commercial farming market (Weighted N = 39 000), based on the categorised levels of sophistication. An approximate 6.5% of commercial farming businesses in South Africa are classified as having the lowest levels of commercial farming sophistication, compared with about 12.2% of the market being classified on the highest level of sophistication. Segments 2, 3 and 4 represent an estimated 81.4% of the commercial farming market in South Africa.

Table 22: Final estimated segment sizes (Weighted N = 39 000)

	Estimated size (N)	**%**
Segment 1	2 521	6.5%
Segment 2	9 930	25.5%
Segment 3	10 508	26.9%
Segment 4	11 300	29.0%
Segment 5	4 741	12.2%
Total	**39 000**	**100.0%**

5.2.3.2 Step 3.2: Segment profiling and validation

In the last step, the relationships between segments across those variables that formed the basis of the construction of the sophistication index, as well as amongst other variables that are expected to be correlated with levels of sophistication, were inspected as part of the validation process. It also confirmed meeting the criteria that was set for segment forming.

In Table 23, the estimated percentage of farming businesses in South Africa where a particular characteristic is evident, according to the sophistication segment, are reported. The characteristics reported are those that form the final set of indicator variables used to calculate the raw sophistication index scores from the weighted sample.

Table 23: Distribution of selected index variables across segments (Weighted N = 39 000)

	Segment					
	1	**2**	**3**	**4**	**5**	**Total**
Primary farming operation: Stock farmer	81.8%	69.2%	63.5%	53.8%	70.2%	**64.1%**
Use a cell-phone as part of farming business	25.7%	96.1%	97.1%	94.2%	97.8%	**91.5%**
Use PC as part of farming business/farm management	-	19.3%	99.3%	100.0%	100.0%	**72.8%**
Use PC for business management	-	0.9%	22.4%	48.4%	88.1%	**31.0%**
Use PC for animal/irrigation management	-	0.8%	22.2%	30.5%	87.3%	**25.6%**
Use PC for financial management	-	1.0%	31.4%	79.9%	98.0%	**43.8%**
Use PC for VAT/tax management	-	0.6%	33.0%	67.0%	92.7%	**39.8%**
Have Internet access	-	15.1%	65.7%	94.6%	100.0%	**61.1%**
Pay wages via Internet	-	1.8%	10.5%	53.4%	69.1%	**27.2%**
Registered for VAT	71.8%	97.2%	96.9%	100.0%	100.0%	**96.6%**

An inspection of Table 23 reveals that commercial farming businesses in South Africa forming part of the first segment (an estimated 6.5% of the total market) are characterised as having the lowest levels of commercial farming

sophistication, and these are predominantly stock farmers (81.8%). An estimated one in four farming businesses (25.7%) makes use of a cell-phone as part of the daily operational and farm management activities. However, the use of a personal computer and having Internet access seem to be largely non-existent. About 71.8% of farming businesses are registered for VAT.

The number of commercial farming businesses forming part of the second segment (an estimated 25.5% of the total market) who are farming primarily with stock are proportionally lower (69.2% compared to 81.8%), with a sizable proportion of farming businesses in this segment (30.8%) being regarded as crop farmers. Where only an estimated one in four farming businesses (25.7%) in the first segment make use of a cell-phone, slightly more than 95% of the farming businesses in the second segment do so. Despite the high proportion of cell-phone users, only one in five farming businesses in this segment (19.3%) use a PC to assist with their operational and farm management activities.

The market penetration of Internet connectivity is also relatively low (15.1%), with only a few farming businesses (1.8%) using this channel to pay wages electronically. As expected, this segment already sees a large percentage of farming businesses (97.2%) registered for VAT.

Of those farming businesses in the third segment (an estimated 26.9% of the total market), an estimated 63.5% farm primarily with livestock, with 36.5% farming with crops. About 97% of farming businesses in this segment make use of a cell-phone. The use of a PC as part of daily farm operational and management activities is evident amongst nearly all of them (99.3%). The use of a PC for various operational and farm management activities emerges within this segment, with 22.4% using one for general farm and business management, 22.2% for animal/irrigation management, 31.4% for financial management, and 33.3% for VAT management.

Nearly 66% have Internet connection, with one in ten (10.5%) paying wages via the Internet. Ninety-seven per cent (96.9%) of farming businesses are registered for VAT.

Of those farming businesses in the fourth segment (an estimated 29.0% of the total market), a much more equal distribution of livestock and crop farming operations is evident (53.8% and 46.2% respectively). About 95% of farming businesses in this segment make use of a cell-phone. The use of a PC as part of daily farm operational and management activities is evident amongst all (100.0%).

The use of a PC for various operational and farm management activities is more established within this segment, with 48.4% using one for general farm and business management, 30.5% for animal/irrigation management, 79.9% for financial management, and 67.0% for VAT management. Nearly 95% have Internet connection, with 53.4% paying wages via the Internet. All the farming businesses within this segment are registered for VAT.

Of those farming businesses in the fifth and last segment (an estimated 12.2% of the total market), a significant proportion of farming businesses are characterised as livestock farmers (70.2%). About 98% of farming businesses in this segment make use of a cell-phone. The use of a PC as part of daily farm operational and management activities is evident amongst all (100.0%). The use of a PC for various operational and farm management activities is established within this segment, with 88.1% using one for general farm and business management, 87.3% for animal/irrigation management, 98.0% for financial management, and 92.7% for VAT management.

All the farming businesses in this segment have Internet connection, with 69.1% paying wages via the Internet. All the farming businesses within this segment are registered for VAT.

Table 24 specifically focuses on the selective firmographic characteristics of commercial farming businesses in South Africa and their relationships with the five segments.

Table 24: Distribution of selective firmographic variables across segments (Weighted N = 39 000)

	Segment					Total
	1	2	3	4	5	
Have secondary branches of farming operations	73.3%	71.6%	75.0%	79.9%	77.5%	**75.7%**
Operate other farming-related business	6.5%	11.6%	24.6%	30.5%	27.2%	**22.2%**
Earn additional income from activities not related to farming operation	54.9%	50.9%	48.3%	53.4%	53.5%	**51.5%**
Primary farming operation administrated: Sole Ownership	90.6%	83.5%	63.7%	54.5%	48.3%	**65.9%**
Primary farming operation administrated: Company	4.3%	1.2%	5.5%	10.5%	19.7%	**7.5%**
Size of farming land: More than 3 000ha	19.2%	20.8%	22.0%	30.7%	28.0%	**24.8%**
Annual turnover: More than R3-million	-	5.2%	9.8%	22.6%	32.7%	**14.5%**
Have cash-flow budget	50.2%	65.1%	71.2%	81.0%	88.7%	**73.3%**
Make use of accountant (external)	34.8%	26.5%	18.8%	33.8%	38.5%	**28.5%**
Make use of independent financial advisor	13.7%	38.6%	43.8%	40.2%	49.6%	**40.2%**
Use: 3G cellular service	-	4.8%	23.9%	30.4%	43.6%	**23.0%**
Have short-term insurance for farming business	74.6%	83.0%	88.6%	90.2%	95.5%	**87.6%**
Operate other farming-related business	6.5%	11.6%	24.6%	30.5%	27.2%	**22.2%**

Inspection of the table reveals that some of the characteristics do not seem to vary notably across the segments. This was confirmed by means of Chi-square tests, using the unweighted database (n = 600) as test data, with the results showing $p > 0.05$:

- Have secondary branches of farming operations;
- Earn additional income from activities not related to farming operations;
- Size of farming land: More than 3 000 ha;
- Make use of accountant (external).

Characteristics that do seem to have a significant relationship with segments ($p < 0.05$) are:

- Operate other farming-related businesses;
- Primary farming operation administered: Sole Ownership;
- Primary farming operation administered: Company;
- Annual turnover: More than R3-million;
- Have cash-flow budget;
- Make use of an independent financial advisor;
- Use: 3G cellular service;
- Have short-term insurance for farming business.

While having secondary branches of farming operations do not seem to differ across the various segments, the prevalence of operating another farming-related business seem to be higher amongst those segments showing higher levels of sophistication. About 6.5% of farming businesses in the first segment also operate another farming-related business, in comparison with 11.6% in segment 2, 24.6% in segment 3, 30.5% in segment 4, and 27.2% in segment 5.

A relationship also exists between the legal structure of a commercial farming operation and segments, with higher levels of sophistication being associated less with sole ownership (48.3% in segment 5 compared with 90.6% in segment 1), but increasingly with that of a company (4.3% in segment 1 compared with 19.7% in segment 5). Segments with higher levels of sophistication also showed higher levels of turnover. No farming businesses in segment 1 reported an annual turnover of more than R3-million, compared with 5.2% in segment 2, 9.8% in segment 3, 22.6% in segment 4, and 32.7% in segment 5.

Commercial farming businesses with higher levels of sophistication are more likely to have a cash-flow budget. About 50.2% of farming businesses in the first

segment reported having a cash-flow budget, compared with 65.1% in segment 2, 71.2% in segment 3, 81.0% in segment 4, and 88.7% in segment 5.

The use of an independent financial advisor is more prevalent amongst segments with higher levels of sophistication. About 13.7% of farming businesses in the first segment reported using an independent financial advisor, compared with 38.6% in segment 2, 43.8% in segment 3, 40.2% in segment 4, and 49.6% in segment 5.

The adoption and use of technologies, such as 3G cellular services are also more prevalent amongst segments with higher levels of sophistication, with about 53.6% of farming businesses in the first segment, compared with 67.5% in segment 2, 76.4% in segment 3, 80.8% in segment 4, and 85.1% in segment 5.

Having short-term insurance is more prevalent amongst segments with higher levels of sophistication. About 74.6% of farming businesses in the first segment have short-term insurance as part of their farming businesses, compared with 83.0% in segment 2, 88.6% in segment 3, 90.2% in segment 4, and 95.5% in segment 5.

The relationship between commercial farming sophistication and the selective demographic characteristics of the active farmer was also explored. Table 25 shows that segments with higher levels of sophistication are more likely to have active farmers with at least a post-matric diploma, a specific agriculture-related qualification and medical cover. Active farmers managing more sophisticated commercial farming businesses are also more likely to regularly review their estate planning.

Table 25: Distribution of selective demographic characteristics of main activate farmer across segments (Weighted N = 39 000)

	Segment					Total
	1	2	3	4	5	
Educational level: Diploma	44.4%	44.9%	57.8%	61.0%	76.5%	**56.9%**
Agricultural related qualification	31.4%	32.2%	51.6%	53.3%	58.2%	**46.6%**
Medical cover	66.6%	86.9%	87.2%	85.7%	88.8%	**85.6%**
Regularly reviews estate planning	85.1%	90.6%	93.1%	96.0%	94.3%	**92.9%**
Educational level: Diploma	44.4%	44.9%	57.8%	61.0%	76.5%	**56.9%**

The last aspect that was considered in the construction of the index and the formulation of segments was the behaviour of data, based on the holdout sample of size n = 276. The holdout sample was independent, as part of the index construction, and this approach allowed for confirming the reliability of the index to serve as a segmentation base. Using the data of the holdout sample, the index values were calculated using Equation 1 (refer to Section 5.2.2.4). The classification rules were applied and five segments were formed.

A comparison of the distribution of selective firmographic characteristics obtained from the holdout sample (n = 276) across the formed segments, as shown in Table 26, and that had been obtained from the sample used to construct the original rules for index calculation (n = 600), shows acceptable levels of correspondence taking into account the sampling errors associated with both samples (approximately 6% and 4% respectively for n = 246 and n = 600).

The average absolute percentage difference between the original estimates reported in Table 24, and those of Table 26, across the 12 variables and five segments, is 6.5%.

Table 26: Distribution of selective firmographic variables across segments derived from holdout sample (Weighted n = 276)

	Segment					Total
	1	2	3	4	5	
Have secondary branches of farming operations	72.7%	73.0%	86.6%	76.2%	72.2%	**76.8%**
Operate other farming-related business	18.2%	13.8%	27.1%	21.8%	36.4%	**23.1%**
Earn additional income from activities not related to farming operation	69.9%	36.2%	52.7%	56.3%	47.4%	**51.3%**
Primary farming operation administrated: Sole Ownership	87.7%	62.5%	72.6%	58.1%	59.9%	**65.9%**
Primary farming operation administrated: Company	-	3.7%	-	11.1%	7.6%	**5.2%**
Size of farming land: More than 3 000ha	14.8%	29.0%	18.8%	26.2%	32.4%	**24.8%**
Annual turnover: More than R3-million	-	4.1%	12.0%	13.2%	39.4%	**13.5%**
Have cash-flow budget	53.3%	71.0%	64.2%	84.9%	91.7%	**74.7%**
Make use of accountant (external)	14.3%	22.6%	21.3%	23.6%	30.3%	**22.8%**
Make use of independent financial advisor	37.2%	35.3%	32.6%	43.5%	52.3%	**39.9%**
Use: 3G cellular service	6.2%	9.2%	18.8%	23.2%	41.2%	**20.5%**
Have short-term insurance for farming business	71.6%	72.5%	83.8%	89.7%	87.5%	**82.2%**

Similarly, the distribution of selective demographic characteristics obtained from the holdout sample (weighted n = 276) across the formed segments, as shown in Table 27, and that were obtained from the sample used to construct the original rules for index calculation (n = 600), shows acceptable levels of correspondence.

The average absolute percentage difference between the original estimates reported in Table 25, and those of Table 27, across the four variables and five segments, is 5.5%.

Table 27: Distribution of selective demographic characteristics of main activate farmer across segments derived from holdout sample (Weighted n = 276)

	Segment					Total
	1	2	3	4	5	
Educational level: Diploma	40.3%	59.0%	68.4%	60.0%	80.9%	**62.7%**
Agricultural related qualification	31.8%	44.7%	61.3%	48.9%	69.0%	**52.1%**
Medical cover	75.0%	79.5%	81.3%	90.5%	87.1%	**83.7%**
Regularly reviews estate planning	83.3%	87.9%	97.7%	96.1%	94.0%	**92.9%**

5.3 CHAPTER SUMMARY

This chapter presented the results pertaining to the constructing of a commercial farming sophistication index for South Africa. This served as basis for investigating and illustrating the process of index construction as a method of market segmentation. Details of the analysis performed in each step of the process was reported and interpreted. A final segmentation of the market was presented.

In the next chapter, conclusions pertaining to the original research objectives are made, areas for future research identified and limitations of the study.