Abstract: The purpose of this paper was to use the metric Multidimensional scaling (MDS) to explore the ten dire household debt determinants in the context of South Africa. Macroeconomic data used was collected from the South African reserve bank and Statistics South African websites for the first quarters of 1990 to 2013. SPSS 22 was used to execute the analysis. A Standardized Residuals Sum of Squares (STRESS 1) measure calculated as 0.0077 confirmed the best fit of the MDS model and the Tucker’s Coefficient of Congruence implied that 99.9% of variance in the model is accounted for by the two dimensions. This was also a confirmation that the ten selected determinants can better be represented in a two dimensional perpetual map. The findings revealed two profiles of household debts. Gross domestic product and house prices are associated with high levels of household debts. The remainder of the determinants is found to have low effects. MDS demonstrated its effectiveness in classifying household debt determinants according to their contribution. Also revealed is that an MDS is a useful tool to use in quantifying the ubiquitous, but slimy, notion of similarity.

Keywords: Multidimensional scaling, household debts, perpetual map, STRESS measure, dimensions, goodness of fit

1. Introduction

Multidimensional scaling, just like factor and cluster analyses is an exploratory data analysis tool used to condense a large amount of data and presenting it in a simple spatial map. This map communicates important relationships in the most economical manner (Mugavin, 2008). The author further emphasized multidimensional scaling (MDS) as having several advantages such as modelling nonlinear relationships among variables and handling nominal or ordinal data. This technique does not require adherence to multivariate normality and have been found effective in extracting typical information in data exploration according to Johnston (1995) and Steyvers (2002). Giguère (2006) and Tsogo et al. (2000) suggested MDS when the purpose of a research is to find structure in the data. The underlying dimensions extracted from the spatial structure of the data are thought by Ding (2006) to reflect hidden structures, or important relationships within it. This procedure is achieved by rescaling a set of dissimilarities measurements into distances assigned to specific locations in a spatial configuration. The more the points are closer together on the spatial map, the more similar are the objects. As a result, a visual representation of dissimilarities (or similarities) among objects, cases, or more broadly observations, will be provided (Jaworska and Chupetlovska-Anastasova, 2009).

The goal of using MDS in this paper is to explore the South African household debt determinants and assign them to the dimensions accordingly. An important assumption made is that the analyst does not have a prior knowledge about these dimensions. Existing relationships between the variables is automatically identified by the technique and those variables that are similar to one another are unconsciously assigned to respective dimensions. These dimensions present a visual display in a perpetual map and the output provided is highly intuitive to interpret. MDS further has the ability to reveal the findings that have not been considered during hypothesis formulation. Davison and Sireci (2000) warned about a drawback about this technique which is a lack of precision of other statistical techniques. The authors however pointed that the manner in which the technique arranges data is very useful at first glance and makes it easy for conclusions to be drawn easily. Rathod (1981) is in support of the view and emphasised on the stability of the spatial representations by the technique. The author further stated this as an instance where the convergence of representations supports the notion that the structure found is integral in the grid and is not an artifact of the method of analysis.

MDS is an old technique and it has received less attention compared to other statistical methods. It was first applied in 1969. A number of studies applied MDS in different fields such as marketing (Steffire,
1969; Neidell, 1969), computer studies (Green and Carmone, 1969; Venna and Kaski, 2006), data mining and exploration (Silva and Tenenbaum, 2004; Groenen and Velden; 2004, Zhang, 2010) among others. There is dearth of literature and research on the exploration of the determinants of household debts more specifically in South Africa. MDS has not been applied to investigate these underlying debt determinants. Several studies have however looked at the effects of these determinants on household debt using econometric techniques such as the vector error correction and the vector autoregressive models. Though these models perform a statistical significance of individual determinant, they do not provide their profiles accordingly. These models are also too technical and the user needs proper knowledge and understanding of statistics. They also tend to give cumbersome and complicated results. There is therefore a need for a study that explores effective frameworks that may provide the results which non-statisticians may also find easy and interesting to read and understand. An MDS provides a guiding map that helps in reducing the complexity inherent to the proximities by combining the determinants according to the type and the amount of effects in household debts. The study also intends to contribute an understanding regarding the different household debt determinants in South Africa. Further an undermined gap in the literature may be filled when novel and more effective techniques such as the MDS are applied. The application of MDS in this study may further create awareness to scholars whose studies are focused on data summarisation and dimension reduction. The rest of this paper is structured as follows: Section 2 discusses and presents the data and methodological framework, Section 3 provides a deliberation on the results and in Section 4, concluding remarks are delivered. Limitations and acknowledgements are highlighted in Section 5.

2. Data and Methodology

This section defines the data used in the study and the motivation for the variables. Also viewed is the methodological framework the study adopted.

Data: MDS analyses the data called proximities. These proximities specify the overall similarity or dissimilarity of the objects under investigation (Wickelmaier, 2003). An MDS framework does the analyses by looking for a spatial configuration of the objects. Ultimately the distances between the objects will match their proximities as closely as possible. This study uses data collected from the South African Reserve Bank and Statistics South Africa for the period 1990 Q1 to 2013 Q1. The data consists of ten macroeconomic and financial determinants of household debt. Literature suggested numerous theories which explain household indebtedness. This study consulted two of these theories and related literature to help in identifying the determinants of household debt. One of the theories called the Keynesian theory was developed by Keynes (1936). The author thought of the subject in the form of the absolute income hypothesis (AIH) and Modigliani (1986) made a follow-up with the life cycle hypothesis (LCH). Friedman (1957) thought the AIH can further be developed and suggested the permanent income hypothesis (PIH). To date, a modified Keynesian theory is being used by economists in macroeconomic related studies.

It is reported in the literature that the level of household indebtedness is determined by supply and demand. Meng et al. (2011) highlighted that most of the households enter into debt due to the availability of funding, by, for instance, credit providers. As a result, this study analysed the factors affecting borrowing and/or lending and adopted the approach used by Meng et al. (2011). The following were identified as potential household debt determinants; house prices (HP), consumer prices (CP), household income (INC), interest rates (IR), gross domestic product (GDP), household consumption (HC), household savings (HS), unemployment rates (UR) and exchange rates (ER) and tax rates (TAX). It is important to note that these variables were also used in the study “Household Debts-and Macroeconomic factors nexus in the United States: A Cointegration and Vector Error Correction Approach”, conducted by Moroke (2014).

Methodology: Multidimensional scaling (MDS) is a statistical technique suitable to provide a visual representation of objects that are similar or dissimilar. This technique also represents objects as points in a dimensional map. Similar objects are positioned next to each other and dissimilar objects are placed far apart. The first step in MDS is to produce matrix of distances between \( n \) objects. The number of dimensions for the mapping of objects is fixed for a particular solution. The following are general steps for carrying out a MDS analyses:

Step 1: Set up the \( n \) objects in \( p \) dimensions, where coordinates \((x_1, x_2, ..., x_p)\) are assumed for each object in \( p \)-dimensional space. In this study, \( n = 9 \) determinants of household debt each with 93 observations and \( p \) is determined from the analysis of the results.
Step 2: Euclidian distances between the objects are calculated for the assumed configuration using the formula:

\[ d_{ij}^2 = \sum_{i=1}^{p} (x_{ij} - y_{ij})^2, \]  

where \( d_{ij}^2 \) is a square of Euclidian distance between points \( i \) and \( j \). \( x_{ij} \) and \( y_{ij} \) are the coordinates on an axis. These objects are grouped according to the similarities or dissimilarities between them. MDS is of two types, the metric and non-metric. The former occurs when the actual magnitudes of the original series are used to obtain a geometric representation in \( p \) dimensions (Johnson and Wichern, 2007). Non metric MDS is as a result of the use of ordinal information to obtain geometric representation. Therefore, a positive monotone transformation is applied to a non-metric MDS for scaling into spatial distances and linear transformation function is applied to a metric MDA. The reader is referred to Giguère (2006) for more details about these transformations.

The original data used in this paper is known thus a metric MDS is used instead. However, Young and Harris (2004) technically recommend the use of dissimilarities as input to the MDS program. This is due to the relationships revealed by the similarities in the distances which are direct and positive. The authors pointed out that the higher the dissimilarity, the larger the perceived distance. Otherwise, if the analysis chooses to use similarities, they must be transformed by subtracting the original data values from a constant which is higher than all collected scores (Kruskal and Wish, 1978). The similarities and/or dissimilarities are calculated using a Euclidian model in [1] as a basis to compute optimal distances between objects in an \( n \)-dimensional stimulus space (Giguère, 2006). This distance is defined as the length of the hypotenuse linking two points in a hypothetical right triangle. The data used for the current study has variables measured in different scales. It is expected that these data are correlated with each other and thus evolves the correlation matrix. The correlational data can be visualised in this matrix as another application of MDS. Irrespective of the number of variables, such matrix becomes complex making it difficult to detect patterns of correlation. An MDS solution plots the objects on a map, so that their correlational structure is accessible by visual inspection (Wickelmaier, 2003).

Step 3: Fit a regression of the distance according to input data. The fitted distances obtained from the regression equation assuming a linear regression are called disparities which are scaled to match the configuration distance as closely as possible.

Step 4: Use a suitable statistic to judge the goodness-of-fit between the configuration distances and the disparities. Kruskal (1964) recommended the use of Standardized Residuals Sum of Squares (STRESS) measure to full fill this duty. STRESS measure is also useful in determining the appropriate number of dimensions to include in the model (Mahole, Moroke and Mavetera, 2014). This measure is calculated using the formula:

\[ \text{STRESS} 1 = \sqrt{\frac{\sum (d_{ij} - \hat{d}_{ij})^2}{\sum \hat{d}_{ij}^2}} \]  

STRESS measure ranges between 0 (perfect fit) and 1 (worst possible fit). Any value less than 0.1 is typically taken to mean a good representation of the objects by the points in a given configuration. If two dimensions are used, a STRESS value below 0.05 is generally considered to be satisfactory (Mazzocchi, 2008). According to Hair et al. (2010), STRESS is minimized when the objects are placed in a configuration so that the distances between them best match the original distances. Arce and Garling (1989) and Kruskal and Wish (1978) suggested that iterative process is run until the STRESS function has been minimised. The authors further explain the purpose of this process as being to find successive approximations aimed towards minimising this STRESS. Ideally, the purpose of MDS is to find representation of objects as points in \( p \) dimensions in such a way that STRESS is as small as possible (Johnson and Wichern, 2007). Table 1 provides a summary of informal guidelines suggested by Kruskal (1964) to help interpret the STRESS measure:
Table 1: Guidelines for interpreting STRESS

<table>
<thead>
<tr>
<th>STRESS</th>
<th>Goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>Poor</td>
</tr>
<tr>
<td>10%</td>
<td>Fair</td>
</tr>
<tr>
<td>5%</td>
<td>Good</td>
</tr>
<tr>
<td>2.5%</td>
<td>Excellent</td>
</tr>
<tr>
<td>0%</td>
<td>Perfect</td>
</tr>
</tbody>
</table>

Source: Kruskal (1964)

This goodness of fit refers to a monotonic relationship between the similarities and the final distance. Takane et al. (1977) introduced an augmented measure of discrepancy and is more preferred than Kruskal’s. This measure is denoted by SSTRESS and replaces \(d_{ij}\)'s and \(\tilde{d}_{ij}\)'s in [2] by their squares. The mathematical formula thus becomes:

\[
SSTRESS = \frac{\sum (d_{ij}^2 - \tilde{d}_{ij}^2)^2}{\sum d_{ij}^2}. \tag{3}
\]

Note that same guidelines as in Table 1 can be used to interpret this measure as well.

**Step 5:** The final coordinates in Step 1 are changed slightly in such a way that STRESS is reduced.

Using coordinates in Step 5, a map that shows how the objects are related is drawn. Literature recommends fewer numbers of dimensions such as two as opposed to many. A scree plot suggested by Kaiser (1974) may be used to identify the optimal number of dimensions. Wickelmaier (2003) noted that the absolute amount of STRESS gives only a vague indication of the goodness of fit. To correct this shortfall, Borg and Groenen (1997) and Hair et al. (1998) recommend an additional technique for judging the adequacy of an MDS solution such as the scree plot. This plot further shows the points of dimensionality versus fit and is useful in selecting the appropriate number of dimensions when there is a clear elbow indicating that increasing dimensions do not affect STRESS in any substantial way (Mahole et al., 2014). Due to a monotonically decrease of STRESS with increasing dimensionality, one is looking for the lowest number of dimensions with acceptable STRESS. An “elbow” in the scree plot indicates that more dimensions would yield only a minor improvement in terms of stress. As a result, the best fitting MDS model has as many dimensions as the number of dimensions at the elbow in the scree plot. It is assumed that the initial similarity values are symmetric, i.e., no ties and no missing observations. To deal with asymmetries, ties and missing observations, the reader is referred to Kruskal (1964). Another way of assessing MDS model appropriateness is by interpreting the squared correlation index \(R^2\) also known as the Tucker’s coefficient of congruence. This measure indicates the proportion of variation brought about by the input data which is accounted for by the MDS procedure. A minimum 0.6 of this measure is considered an acceptable fit according to Meyer et al. (2005).

3. Empirical Results and analysis

This section discusses the results obtained through the guidance of the objectives and the methodology discussed. The results are presented in tables and on figures. Main interest is a perpetual map in which the variables are positioned. The results on individual analysis, aggregate analysis and individual difference scales are discussed separately.

**Individual analysis:** All the variables used in the analysis are continuous and therefore a metric MDS is chosen. The analysis involves the ten household debt determinants each with 93 observations. The initial analysis in MDS model is to generate the proximities, which is also a matrix of similarity distances between two points. The matrix is generated from the ten household debt determinants, each with ninety three observations using a Euclidian distance measure. The main intention of the matrix is to identify the determinants with similar effects on household debts levels. The coefficients of this matrix help in identifying such similarities. Determinants with high coefficient measure are similar to others and the opposite is true about those with low coefficients. The results of a proximity matrix are summarized on Figure 1.
It is clear that the determinants of household debt can be represented in clusters due to the variety of the coefficients displayed on Figure 1. Euclidian distances between the determinants range between 0.313 and 18.686. This explains the diversity of the determinants and suggests that some groupings may be expected.

The results on the correlational measure between the ten variables are summarized as Figure 2.

From the data in Figure 2, it cannot be clearly deduced which household debt determinants are related. There is a variety of correlations ranging between -0.112 and 0.999. This proves that some of the determinants are highly related while others have low relationships. It is however difficult despite the size of the matrix to identify those clusters of the determinants. The results in Figure 2 are in support of those in Figure 1. Presented in Figure 3 is a scree plot to help identify the number of dimensions or clusters. This figure plots STRESS values against the dimension number.
An elbow starts at the second point of the plot implying that the ten determinants of household debt can better be represented with two dimensions. Further analyses are carried out with this information in mind.

**Table 2: Stress and Fit Measures**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Raw Stress</td>
<td>0.00077</td>
</tr>
<tr>
<td>Stress-I</td>
<td>0.02782</td>
</tr>
<tr>
<td>Stress-II</td>
<td>0.04258</td>
</tr>
<tr>
<td>S-Stress</td>
<td>0.00143b</td>
</tr>
<tr>
<td>Dispersion Accounted For (D.A.F.)</td>
<td>0.99923</td>
</tr>
<tr>
<td>Tucker’s Coefficient of Congruence</td>
<td>0.99961</td>
</tr>
</tbody>
</table>

As shown in Table 2, the measure-of-fit for this solution, normalised raw STRESS and STRESS 1 according to [2], gives a value 0.00077 and 0.0278 respectively. These measures fall within the range of excellent and perfect according to by Kruskal’s (1964) rule of thumb as shown in Table 1. The STRESS 1 value confirms the best fit of the MDS model and the Tucker’s Coefficient of Congruence 0.999 illustrates that 99.9% of variance in the model is explained by the two dimensions. This therefore allows the variables to be represented in two dimensions with no doubt. Both the scree plot and the STRESS 1 measure provide similar results in accordance with Borg and Groenen (1997) and Hair et al. (1998). This section gives the results categorising the determinants into the two dimensions identified on Figure 3.

The strategy is to look for groups of variables in a certain dimension. It is further necessary to look for distant variables in a map. Two default stimulant coefficients that pull apart the ten determinants are summarized. The two dimensions may be interpreted such that the first is associated with high and the second with low levels of household debts. For instance, a group of determinants located at the upper right corner of the perpetual map represent those with extremely high household debt levels. Presented in Figure 4 is a map of the ten determinants using individual analysis.

**Figure 4: Individual analysis perpetual map**

![Individual analysis perpetual map](image)

Shown in a figure is a group of four determinants of household debt such as exchange rate, household savings, consumer prices and interest rates associated with high and extreme household debts in South Africa. Household consumption expenditure and gross domestic products are far from each other and are also situated on different dimensions.

**Aggregate analysis:** In this analysis, the 93 proximity matrices were combined by computing the average value of each cell. A metric MDS model with Euclidian distances was used to present data. The results are summarised in Table 3 and Figures 5 and 6. Table 3 shows the weights each variable has in a given dimension.
The results show much greater variation in the locations of the determinants of household debts on dimension one than on dimension two. However, gross domestic product seems to be far apart from other determinants as the associated coefficient reveals. The magnitude of this determinant implies that it had more effect on household debt during the chosen period. Household income appears strong in the first dimension. Although not extremely strong, consumer price index, interest rates, household savings, exchange rates, unemployment rates and income tax rates have the same contribution on dimension one. The interpretation of these results may lead to misleading conclusions. As a result to avoid confusion and misrepresentation about the determinants, the study interprets the perpetual map represented as Figure 5. Note that the weights in Table 3 are used to generate the derived perpetual map also known as the stimulus configuration. This map gives a good understanding of the dissimilarity or similarity between the variables as compared to the coordinates in Table 3. The complexity of the results in Figures 1 and 2 is also simplified by presenting Figure 5. The distances in this figure correspond to the correlation coefficients represented in Figure 2. A high correlation is represented by a small distance, and vice versa. In addition to the graphical representation, an MDS analysis provides an explanation of the correlations by interpreting the axes of the associated space.

### Table 3: Final Aggregate Coordinates

<table>
<thead>
<tr>
<th></th>
<th>Dimension 1</th>
<th>Dimension 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP</td>
<td>.209</td>
<td>.085</td>
</tr>
<tr>
<td>CPI</td>
<td>-.481</td>
<td>-.044</td>
</tr>
<tr>
<td>INC</td>
<td>.705</td>
<td>.089</td>
</tr>
<tr>
<td>IR</td>
<td>-.481</td>
<td>-.044</td>
</tr>
<tr>
<td>GDP</td>
<td>1.405</td>
<td>-.305</td>
</tr>
<tr>
<td>HC</td>
<td>.566</td>
<td>.394</td>
</tr>
<tr>
<td>HS</td>
<td>-.481</td>
<td>-.044</td>
</tr>
<tr>
<td>ER</td>
<td>-.481</td>
<td>-.044</td>
</tr>
<tr>
<td>UR</td>
<td>-.481</td>
<td>-.044</td>
</tr>
<tr>
<td>TAX</td>
<td>-.481</td>
<td>-.044</td>
</tr>
</tbody>
</table>

As revealed by the figure, there is one clumping in dimension one of CPI, IR, UR, HS, ER and TAX. All these determinants mentioned cluster between the values 0 and -.5. According to the definition of this dimension, these determinants are viewed as those associated with extremely low levels of household debts. HC and INC are associated with low levels of household debts. House price determinant have extremely high effect on debts and growth domestic product is proven to have high but not extreme influence. The figure reveals less congruence with the clustering in other dimensions. The results revealed that the iteration process stopped at twenty as improvement (0.00007) has become less than the convergence criterion (0.00077). The clumping on dimension one which was revealed by the
individual analysis has shifted to dimension two. Other determinants also rotated from their original dimensions. Figure 6 illustrates a departure from linearity measured by the STRESS 1 and Tucker's Coefficient of Congruence. The fit plot shows the transformed data plotted on the vertical axis and the distances from the model on the horizontal axis.

**Figure 6: Transformed proximities residual plot**

All the points lie on a diagonal line on Figure 4, implying that a model exhibit a perfect fit. Very few of the observations reveal some points departing from the diagonal line. These departures represent the residual of the corresponding observation.

**Individual differences:** This section represents both the stimuli in a common MDS space and the individual differences. It is assumed that all the subjects use the same dimensions when evaluating the objects but that they might apply individual weights to these dimensions. By estimating the individual weights and plotting them different groups may be detected. Figure 7 is a representation of the individual differences of the ten household debt determinants and the 93 observations.

**Figure 7: Individual difference MDS**

The map looks exactly the same as the aggregate MDS representation shown in Figure 5. The tight clustering of the observations’ weights reveals that the sample is homogenous. All the observations are
within the boundaries of -500,000 and 500,000. This means that all the observations put the same weight on the two dimensions.

4. Conclusion

The intension of this paper was to determine clusters of the ten determinants of household debts using a metric MDS framework. The data used was collected from the South African Reserve Bank and Statistics South Africa spanning the period first quarter of 1990 to first quarter of 2013. Total of 93 observations were used. A STRESS 1 measure confirmed that the model has a best fit indicating that representing the ten determinants of household debt in two dimensions is almost excellent. Tucker's Coefficient of Congruence also indicated that the two dimensions explain about 99.9% of variance in the model. A transformed proximities residual plot revealed that a STRESS measure and Tucker's Coefficient of Congruence provide a perfect fit confirming the validity of the perpetual map in representing the determinants. The results showed how effective MDS method can be of use in order to give a visual presentation of multivariate data. Also revealed is that MDS is a useful tool to use in quantifying the ubiquitous, but slippery, notion of similarity. The perpetual map helped in reducing the complexity inherent to proximities by combining the determinants according to the distances between them. This map also helped in exploring a space as no prior hypotheses was made about the variables in terms of their similarities. This graphical representation of data is advantageous in that a clear and unambiguous explanation can be given for the separation of the determinants of household debt. This helped in segregating those determinants associated with high levels from those associated with low levels of household debts. A perpetual map provided a better and clear picture about these determinants.

Both the aggregate and the individual difference analyses showed similar results. The six household debt determinants such as the consumer price index, interest rates, unemployment rates, household savings, exchange rates and income taxes were shown to be associated with extremely low levels of household debts. This implies that the South African policy makers may not be concerned about these factors as they do not pose a serious threat on household debt. Same goes for household consumption expenditures and household income, these variables have low effect on household debts and they may also be of less concern in the South African context. However, it is of extreme importance that urgent attention be paid to the determinants such as the house prices and gross domestic product. These have been found to be the most influential factors of household debts in the country. They must be targeted first when dealing with the problem of household debts. A recommendation is made that further studies be conducted where other related determinants will be added to the configuration. If a determinant clusters with those associated with high levels of debt, it will be easy for policy makers in the country to embark on policies that may be used to reduce household debts due to the availability of information. The findings of this study may be used by policy makers in SA to formulate related policies. It is further recommended that factors associated with extremely high and high household debts be given first priority as they pose a serious threat in the ever growing problem household debts. In the long run, policies may be formulated which take into consideration other eight factors of household debts.

Study Limitations: Though MDS is a very practical and straightforward way of exploring the data, it is advisable to always be careful when giving interpretations about the results obtained from this method. This is because the method does not have statistical tests for validating common space interpretations. Furthermore, MDS does not provide certainty about the conclusions drawn from the analysis.

Acknowledgements: The author is thankful to the employer, the NWU for funding the paper and the support provided.

References


