A fresh look at the ASISA fund classification standard

By JCDS Frasco

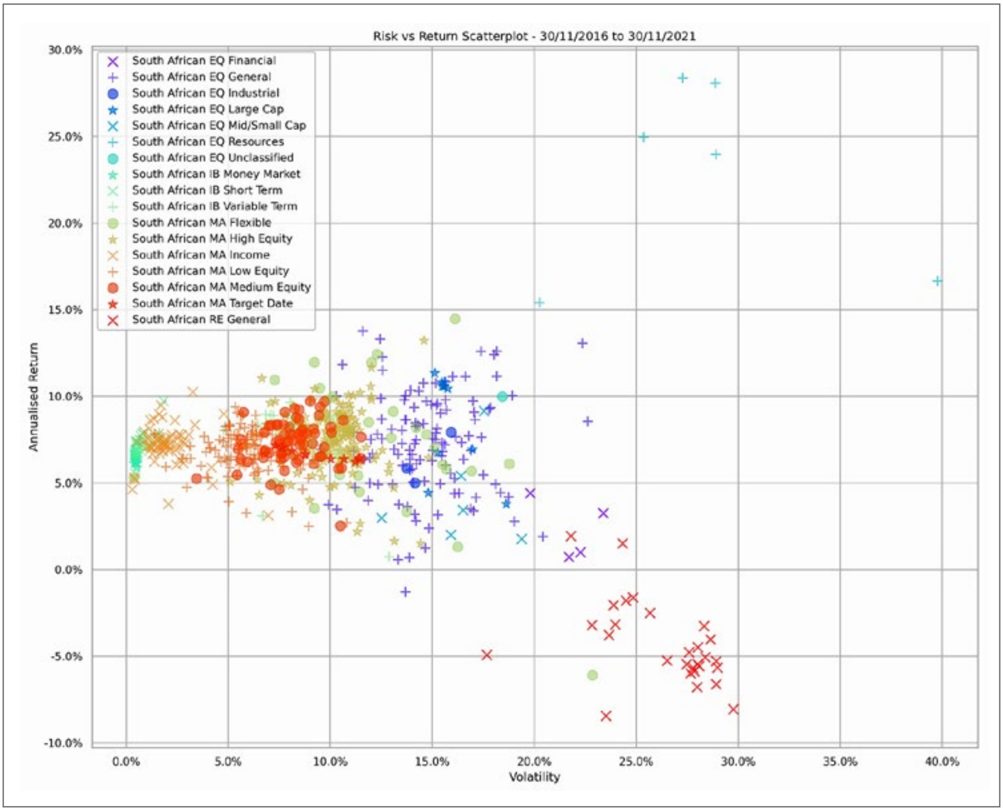
Presented at the Actuarial Society of South Africa's 2022 Convention
Cape Town 25–28 October 2022

ABSTRACT

This paper considers the current Association for Savings and Investment South Africa (ASISA) Standard on Fund Classification for South African Regulated Collective Investment Scheme (CIS) Portfolios, and offers some recommendations on how this could be improved to meet the objective of making portfolios more comparable for users of this information. It does this by considering the objectives of having a classification standard, as provided by ASISA in the document referenced below, and by looking at several different traditional methods that may be considered as appropriate when considering whether portfolios are similar or different. It then introduces some machine learning techniques that may be more up to the task of highlighting the similarities and differences between portfolios, within and across categories in the classification standard, before ending with some thoughts and recommendations on how the classification standard could be altered to better meet the objectives. We hope this is considered by ASISA and the Financial Sector Conduct Authority (FSCA), and the broader investment fraternity, when considering the subject.

KEYWORDS

South African, collective investment schemes, ASISA, classification standard, machine learning, hierarchical clustering, dendrograms, K-means clustering, principal component analysis, singular value decomposition, K nearest neighbours classifiers, decision tree classifiers, confusion matrix, risk versus return scatterplots, multi-asset, equity, interest bearing, real estate



Risk vs return for all of the South African ASISA categories

CONTACT DETAILS

Joao Frasco, Stanlib Multi-Manager and INN8 Invest; Email: joao@frasco.co.za

1. INTRODUCTION

> *The classification of South African regulated collective investment scheme portfolios ("CIS portfolios") provides a framework within which portfolios with comparable investment objectives and investment universes are grouped together. It is a key tool for investors and their advisers in that it provides useful information during the consideration of investment choices.*

Text in this paper that appears formatted as above (italic and indented) is a direct quote from the document titled "ASISA standard on fund classification for South African regulated collective investment scheme portfolios", found at: https://www.asisa.org.za/media/y5mcfou4/20210917_asisa-fund-classification-standard-final.pdf.

1.1 The problem statement

> *OBJECTIVE OF STANDARD*
> *The ASISA Standard on Fund Classification for South African Regulated Collective Investment Portfolios ("ASISA Fund Classification Standard") establishes and maintains a classification standard for CIS portfolios in South Africa.*
>
> *The objectives of the ASISA Fund Classification Standard are to:*
> — *Promote investor awareness and understanding of CIS portfolio types;*
> — *Assist with the comparison of CIS portfolios within and across classification categories; and*
> — *Assist with the assessment of potential risks of investing in a particular type of CIS portfolio.*
>
> *The purpose of the ASISA Fund Classification Standard is to:*
> — *Ensure that CIS portfolios adhere to the classification category definitions;*
> — *Standardise applications for approval of the classification of a CIS portfolio; and*
> — *Facilitate the timeous and appropriate classification and reclassification of CIS portfolios.*

Is it time to review the ASISA standard on portfolio classifications? Using the existing classifications of CIS portfolios, traditional techniques, and the latest machine learning techniques, we may find better methods to classify portfolios. According to the quote above, the objective of the classification is to group together "portfolios with comparable investment objectives and investment universes". Doing this subjectively, within very broad asset class constraints can be very difficult, and can certainly lead to portfolios that behave very differently (from a return perspective), despite perhaps having similar investment objectives and universes, defeating the objective to "Assist with the assessment of potential risks of investing in a particular type of CIS portfolio".

One way to analyse how portfolios behave, is to consider their returns (say monthly), over a period of time (say three, or five years). Although portfolio behaviour may, and likely will, differ in future to how it has behaved in the past, analysing historic returns may be better than using the alternative methods of classifying portfolios used currently, which can be demonstrated to show very different results for two portfolios within the same category, making risk assessment by classification at best spurious, and at worst dangerous.

We will limit the analysis of ASISA categories to looking only at the South African region (in tier 1 of the classification standard). The same methods and principles could be applied more generally to the rest of the regions and categories within them. South African categories generally allow up to 30% in foreign assets, and an additional 10% in other African assets.

1.2 Structure of this paper

This paper is split into five main sections.

In section 2, we explore some of the traditional techniques that can be used to compare and contrast portfolios. This includes asset allocation, portfolio-based analysis, transaction-based analysis, and returns-based analysis. We discuss the shortcomings of each of these methods, before turning to returns-based analysis for the rest of the paper, although we come back to asset allocation throughout.

In section 3, we explore some machine learning techniques available to assist in understanding how portfolios may be grouped based on nothing but returns, using an appropriate frequency and time horizon, and possibly their existing category classification. Some of these tools have existed for a long time, but are making a comeback with the recent surge in popularity of machine learning.

In section 4, we look at the current classification standard, and how homogeneous or heterogeneous it is, by looking at the range of results obtained from the portfolios classified with the existing classifications. This demonstrates some of the potential deficiencies, which we explore in section 5.

In section 5, we build on the previous three sections to propose how we could think about a new classification standard. Where a classification exists that would assist in grouping comparable portfolios, this could be easily adopted. Where no good classification exists, a catch-all for ill-defined portfolios could exist as per the current system, i.e. label these portfolios as Unclassified.

In section 6 we summarise our main recommendations for the changes to the classification standard for the South African tier. These can be used by practitioners in thinking about portfolios and their structure when considering alternatives. They can also be used by ASISA and the FSCA when thinking about redesigning the existing classification standard to better meet its objectives of assisting investors and their advisers.

1.3 The objective

The objective of the paper and the recommended changes to the standard are intended to show how portfolios within most categories, through the comparison of historical returns, could be made more comparable in future.

This is an important point that requires some more attention. We discuss the current or historical behaviour of portfolios to decide how similar or different the portfolios have been. The implication is that this will provide guidance on the similarity and differences of those same portfolios in future, which is ultimately the subject of interest. We do not look at predictive analytics or methods to make predictions of how portfolios will behave relative to each other. While this certainly could be done using many of the techniques introduced here, and many others, that is not the objective of this paper.

As an incidental benefit, the same system can then also be used to classify new portfolios, using very few returns and some of the supervised machine learning techniques introduced below. Perhaps new portfolios that do not fit *a priori* into one of the existing categories remain Unclassified until they have, say, 12 months of returns, and then move into a specific category, if the algorithms deem the fit to be good. This can equally apply to any other portfolios already in the Unclassified category, and the methods could be applied to portfolios that are not behaving as expected within their categories.

Figure 1 lists all of the ASISA categories and the number of registered portfolios with retail fee classes, ignoring institutional-only portfolios. It also shows the assets under management (AuM) in each category, across all portfolios and fee classes, retail and institutional, as well as the number of portfolios with at least five years of historical returns, which will be used in the analysis throughout this paper.

This paper will only focus on South African portfolios in the first tier of the classification standard, thereby ignoring the global, worldwide, and regional tiers.

2. TRADITIONAL TECHNIQUES

2.1 Asset allocation

There are essentially three different methods that asset allocation could be used to compare portfolios.

2.1.1 CURRENT ASISA CLASSIFICATION STANDARD

The first method, which is used by ASISA in the current classification standard, is to provide limits or constraints on the allocation to some or all of the different asset classes available to investors. This raises a number of issues.

One issue is having to define precisely each and every asset class, or combining some. You may need to include a category for 'Other' as a catch-all for everything not specified, and this may include very heterogeneous assets. You will then be required to define minimums and maximums for each of these specified asset classes with narrow bands, or lose the benefit of having homogeneous groups if the bands are too wide and overlap between categories. You also have the problem of portfolios being allowed to drift outside of the bands for a considerable period of time, essentially negating the bands. The previous two issues are issues with the current classification standard.

You also have an issue regarding the relative homogeneity, or lack thereof, of the asset classes being used. It is easy to think about local listed equity as a single asset class but a detailed analysis of shares and portfolios will quickly highlight just how different these can behave, for example, think about resources versus financials.

With access to this classification standard, one could then assume that portfolios within the same category should be similar, and the category fairly homogeneous, but it would require well-defined and narrow limits as raised above. Many constraints may not be specified, making any comparison potentially superficial at best – this is a problem with the current classification standard, especially for the Multi Asset tier.

ASISA category	Number of portfolios	Assets under management	Number of portfolios with five years of history
South African MA High Equity	210	R701 308 493 073	152
South African EQ General	170	R342 911 620 712	117
South African MA Low Equity	158	R282 129 643 085	118
South African MA Income	110	R291 937 651 773	67
South African MA Medium Equity	98	R99 246 542 364	72
South African MA Flexible	52	R49 839 726 381	37
South African IB Short Term	39	R263 118 754 366	27
South African IB Variable Term	39	R83 614 340 456	24
South African RE General	37	R33 396 568 254	28
South African IB Money Market	33	R186 383 480 722	28
South African EQ Large Cap	12	R15 400 565 256	11
South African MA Target Date	8	R8 107 455 422	8
South African EQ Mid/Small Cap	7	R6 668 066 315	6
South African EQ Resources	6	R5 488 946 533	6
South African EQ Financial	4	R1 122 371 184	4
South African EQ Industrial	3	R2 358 579 323	3
South African EQ Unclassified	1	R205 339 779	1
Global EQ General	90	R213 239 883 325	43
Global MA Flexible	49	R49 849 621 729	22
Global RE General	17	R9 112 369 579	13
Global MA High Equity	16	R44 948 837 770	11
Global MA Low Equity	10	R9 515 113 000	9
Global IB Variable Term	5	R1 747 052 882	2
Global MA Income	3	R2 525 761 170	2
Global IB Short Term	3	R770 570 097	3
Global MA Medium Equity	2	R205 448 006	0
Global EQ Unclassified	1	R177 958 109	1
Worldwide MA Flexible	92	R69 317 973 702	57
Worldwide EQ General	9	R6 306 158 416	2
Regional EQ General	4	R997 401 084	3
Regional Equity - General - Africa	3	R2 278 993 375	2
Regional MA Flexible	2	R1 204 433 127	1
Regional IB Short Term	2	R490 371 594	2
Regional EQ Unclassified	2	R262 659 210	0
Regional IB Variable Term	1	R373 066 371	1
Unclassified	3	R706 207 090	0

FIGURE 1 Table of ASISA categories with portfolios and AuM as at 30 November 2021

2.1.2 STRATEGIC ASSET ALLOCATION

The second method is to use the portfolio's strategic asset allocation (SAA) to compare portfolios. This has potentially even more drawbacks, including that many portfolios neither have an SAA, nor are managed to one. Even if some portfolios do have an SAA, the asset classes that each one uses could be different, even though you would expect the major asset classes to be present in various portfolios.

You would also find that even those portfolios with an SAA may deviate from it substantially over time, so that two portfolios with the same SAA may behave quite differently to each other because their actual asset allocation has deviated from the SAA, and the portfolio it is being compared to.

You also need to consider the relative homogeneity of the asset classes raised above: a good example being the classification, by many practitioners, of inflation-linked bonds alongside nominal bonds although they are very different asset classes. You would also need to think carefully about how to compare two portfolios given their multiple asset classes; specifically, what makes two portfolios similar or different. Is it just the difference in their equity allocation, or every difference, and how are these differences aggregated?

There are machine learning techniques that will be introduced in the next section that could ameliorate some of these problems, but the sparsity of data would make the exercise of very limited use in most cases.

2.1.3 ACTUAL ASSET ALLOCATION

The third method is to use the portfolio's actual asset allocation, but again we are left with some of the issues that are raised above, including the consistency in the definition of asset classes. The availability of data in any standardised way, given the consistency raised above, is a major problem when comparing portfolios. Another issue is how to deal with the change in the asset allocation over time, although this can be mitigated using machine learning techniques discussed below, if the data is readily available in a consistent form. Finally, the relative homogeneity of asset classes as raised above is another issue to overcome.

2.2 Holdings-based analysis

We could compare portfolios based on their holdings over a period of time. If the period of time is long enough, and the frequency at which holdings are considered is frequent enough, this can lead to mountains of data, making comparison using traditional methods impossible. Machine learning techniques designed for big data would however be up to the task. Unfortunately, holdings data is not readily available for CISs, so this method would be theoretically possible but probably impractical in most situations. It is perhaps important to note that there is commercially available software that does this using quarterly holdings data.

This could however be a longer term project for the industry to start thinking about.

The publishing of daily holdings data, however infrequent and delayed so as to not be used to trade, but rather for analytic purposes by practitioners and academics, could prove very useful. Active ETFs already do this in the US at the end of every single day, with no delay.

2.2.1 TOP 10 PORTFOLIO HOLDINGS

The top ten holdings, disclosed by most portfolios every month on factsheets, can be useful when comparing portfolios, however, they say very little about how those portfolios will behave or have behaved in the past. Having an historical record of these top ten holdings would certainly be a lot more useful, as this would represent many more dimensions that could be used to compare portfolios using advanced machine learning techniques. However, you are still limited by not knowing what the rest of the portfolio looks like, and this could be very different if not well represented by the top ten holdings.

2.3 Transaction-based analysis

Similar to holdings, we could compare portfolios based on their transactions over a period of time. If the period of time is long enough, and the frequency at which transactions are considered is frequent enough, this can lead to mountains of data, making comparisons using traditional methods impossible. Machine learning techniques designed for big data would however be up to the task. Unfortunately, transactional data is not readily available for CISs, so this method would be theoretically possible but probably impractical in most situations.

As for holdings-based analysis, this could be a longer term project for the industry to start thinking about. The publishing of daily transaction data, however infrequent and delayed so as to not be used to trade, but rather for analytic purposes by practitioners and academics, could prove very useful. Active ETFs already do this in the US on a daily basis, so this is neither impossible, nor impractical.

2.4 Returns-based analysis

Returns-based analysis too has many limitations, but it has the benefit of returns being available for every single portfolio. Although daily returns are generally available, and can be even more useful than, say, monthly returns, we will focus on monthly returns as these are generally easier to obtain and provide up to twenty times less data. Certain numbers will also not differ based on the frequency of data being used e.g. the annualised return over a specific time period is the same whether you use daily or monthly returns (specifically time-weighted returns). Volatility is often also the same irrespective of the frequency, but this does require independent and identically distributed returns.

There are many traditional methods of using returns to compare portfolios, but we will focus on what we think are the main ones used by typical advisers, asset consultants, multi-managers, and investors more generally.

2.4.1 RISK VERSUS RETURNS SCATTERPLOTS

This is probably the most used (and even overused) method to compare portfolios against themselves over time, or against other portfolios over the same period of time. You will find it in every system that stores and allows users to analyse portfolio returns, as well as most, if not every, survey of portfolio returns. The returns are typically used to see how well or badly portfolios have performed over the period (annualised return achieved), and the risk is used to establish how much risk was taken in delivering the return.

The return is usually just the annualised time-weighted return of the portfolio over the period under consideration, and the risk is typically the annualised standard deviation of the same returns (technically the natural log of returns), commonly referred to as volatility. Although the time period typically focuses on three years of returns, any other period can be used, although you need to be careful about calculating standard deviation/volatility with very little data. The fewer data points used, the higher the standard error of the statistic being calculated.

While many criticise volatility as not being risk, we think this criticism is largely misplaced. Risk is inherently very difficult to establish absolutely in most cases, but certainly in investments. It can therefore be very subjective, unless you decide to use a metric like volatility. Volatility considers risk as uncertainty, as it measures the spread of returns around the average of the returns. Low/high volatility would therefore imply low/high uncertainty in returns delivered. If you then assume that the past is your best estimate of the future, which is not a bad starting assumption in the absence of any other information, then, by definition, that uncertainty translates directly into risk.

Obviously you can have situations where volatility does not imply risk, where it is a very bad proxy for risk, and where volatility can be very misleading. One example would be where prices do not reflect the underlying value of the instrument, perhaps because the instrument is very thinly traded, or not traded at all. Similarly for prices that are marked to model as in private equity and infrastructure or unlisted debt. In these cases, volatility, and hence risk, is typically understated, as prices are not changing as frequently or by as much as they would otherwise do.

The biggest problem however comes with strategies that count on very low probability events, and the period under consideration does not include at least one of these events, because they occur very infrequently. Writing deep out of the money options is one such strategy that is often referred to as 'picking up pennies in front of a steamroller'. You may make money consistently for many days, weeks, months, or even years, and volatility would not reflect the inherent risk in such a strategy. When the event then occurs, which it inevitably does with a higher frequency than was initially estimated, the results are invariably catastrophic.

But let us not throw out the proverbial baby with the bath water, as volatility is very useful in many other situations, for example, in deep and liquid markets. When combined with a risk-free return (again proxied), it leads to the Sharpe Ratio, which is extensively used in analysing portfolio performance. The Sharpe Ratio measures the excess return of the portfolio over the risk-free rate, per unit of risk as measured by volatility. It is therefore referred to as a risk-adjusted return measure.

So while we can find fault with volatility not always representing risk, by sometimes understating it, you will never find it overstating risk, which is often the context used when criticism is levelled at it. This is often how it is criticised, where it does not suit the manager because their strategy has very high volatility, and they do not believe that this is indicative of the risk of the portfolio or the strategy.

The reason that risk versus return scatterplots are so useful and used, is because risk is often the most important ex-ante consideration that investors and their advisers think about when investing. Financial advisers are required to establish the risk profile of their clients before recommending an appropriate portfolio or solution. Volatility can be translated into potential losses of capital under certain assumptions (which are sometimes not valid) and this can be very useful for investors to compare to their appetite for risk.

Two portfolios with very different volatilities are therefore generally considered to be very different, at least when considered from certain dimensions, and two portfolios with similar volatilities are considered to be similar. The judgement is often made that although returns in future may not look exactly like the past, the risk is probably well represented by the historical volatility. This thinking should only be applied when the period over which volatility is measured is 'long enough', and the period includes at least some unexpected events.

In our opinion, you should use periods of at least three years, but preferably five years. As a rule of thumb, when measured over three years, you can expect volatility to fluctuate from a maximum value around three times the minimum value e.g. from 3% p.a. to 9% p.a. Most of the time however (say 80% of the time), you can expect it to fluctuate from a maximum of twice the minimum e.g. 4% p.a. to 8% p.a. So be very careful when you see a volatility number calculated over three years at a point in time, and think about extrapolating this number into the future.

As far as including rare and surprising or unexpected events, the last five years are appropriate as they included the market turmoil brought on by the Covid pandemic. This event had a significant impact on returns from all major asset classes, including local and global equities, bonds, and property. Money market rates also plummeted as central banks lowered their lending rates to unprecedented lows, and even negative rates.

2.4.2 CORRELATIONS AND CORRELATION MATRICES

The correlation between two portfolios is an often used metric of the similarity of two portfolios. Correlation matrices just generalise this to many portfolios, providing pairwise correlations between every pair of portfolios. This can be beautifully visualised in a table of the matrix, along with a heatmap of the correlations, where high numbers (from zero to one) are typically coloured in increasingly darker shades of red to reflect bad diversification, and low numbers (from zero to minus one) might be coloured green to reflect good diversification. Of course this is often a misunderstanding of how correlations actually work, so let us begin by defining correlation.

Correlation is a statistical measure of the linear relationship between two variables. Although there are several types, we will focus on the most commonly used one here (Pearson's). It measures how two portfolios' returns co-vary, in relation to their averages (means), and standardises this by the variability of each portfolio's returns (standard deviation). You end up with a number that ranges from: minus one (perfect negative correlation, implying that the returns move in opposite directions, one portfolio up while the other down and vice versa), through zero (no correlation, implying that the two portfolios returns appear to have no relationship, meaning that sometimes they move in the same direction and sometimes they move in opposite directions), to plus one (perfect positive correlation, implying that the two portfolios' returns move in the same direction, up and down together).

Heatmaps should reflect this by colouring numbers around zero green, as they represent good diversification (no correlation), and numbers closer to one or minus one red, as they represent high correlation (either positive or negative respectively).

While correlation may represent a good first approximation as to the similarity of two portfolios, it comes with some limitations. The first is that you should be comparing two portfolios that are similar (perhaps in terms of objectives, or the asset classes in which they invest) to begin with, otherwise the correlation could be spurious. Correlation also ignores the beta in the relationship between the two portfolios in a linear regression.

The beta measures the sensitivity or rate of change of one portfolio to another. You could theoretically have two portfolios with very high correlation but very low beta. These portfolios would not be similar in any meaningful way. A contrived example of this may be two portfolios from a single manager, where the first portfolio is 100% invested in local equity, and the other portfolio is 50% invested in the same local equity, and 50% in cash. The two portfolios would have a correlation close to one, but would deliver very different returns and have very different risk.

The other issue with correlation matrices is that while they can be useful when comparing any two portfolios, it becomes very difficult to understand all of the relationships when the number of portfolios increases by even a small number, as the number of pairwise correlations increases proportionally to the square of the number of portfolios. This is again where machine learning techniques become useful in understanding the structure (similarities and differences) of many portfolios, and we will look at this in the next section.

2.4.3 STYLE-BASED ANALYSIS

Style based analysis compares portfolio returns to the returns of style indices, either individually, or in combination, to measure the sensitivity (beta) of the portfolio to each of the styles being used in the analysis. Portfolios can then be compared based on these sensitivities (betas) and assessments can be made on the similarity or differences between portfolios.

This may be especially relevant and useful when looking at a specific asset class, specifically equities, where style indices typically exist, focused on styles supported by loads of research. Unfortunately, you still then need to decide how you will compare portfolios based on these sensitivities, and the machine learning techniques introduced below may provide some methods to do this.

These methods are generally more useful when analysing a single portfolio to understand the manager's philosophy and process. Comparing two or more portfolios using these methods can also be very useful, but it typically does not provide a unique perspective on the similarity and differences between portfolios. Bear in mind that having summarised sensitivities results in a substantial loss of data, that when used directly can be more beneficial when making comparisons.

We will therefore not be deploying this technique within this paper, but thought that it was important to raise it as a legitimate tool when analysing and comparing specific portfolios.

2.4.4 FACTOR-BASED ANALYSIS

Factor-based analysis is similar to style-based analysis, except more generic factors are used in the analysis, instead of styles which are typically equity-centric. Factors could include things like exchange rates, gross domestic product, interest rates (generally, but specifically some version of them, like the repo rate), inflation, commodity prices, etc.

The same issues as with style-based analysis arise, in that you still need to decide on how to compare the sensitivities of different portfolios, and the same solutions to these issues exist.

We do not deploy this technique in this paper, although it is a legitimate tool when analysing and comparing specific portfolios.

2.4.5 NOMINAL VERSUS ACTIVE RETURNS

Before moving on to the next section, perhaps a few words on active versus nominal returns is appropriate. In this paper, we consider only nominal returns for a couple of reasons, but first let us define these terms so that we better understand the differences.

Nominal returns are the actual time-weighted returns achieved by the portfolios, monthly, or over longer time periods, in which case they are simply linked or compounded.

Active returns are nominal returns relative to some benchmark, which could be an index or composite of indices, or anything else. The way the relative calculation is done can be important. It is sometimes taken as the simple difference of the two returns (portfolio minus benchmark), and sometimes taken as the ratio (1+portfolio return divided by 1+benchmark return, all minus one). Ratios are preferred for analytical purposes because they can be aggregated for longer periods, whereas differences cannot be (although you can still calculate the longer term returns for the portfolio and the benchmark and difference them).

While comparing portfolios based on their active returns is intuitively more appealing, because you are considering how the portfolios behave relative to a benchmark, and how this makes them similar or different, there are some major issues. By far the most important is that the portfolios being compared may not be managed to the same benchmark. While this may seem like a trivial matter when comparing two, or a very few, portfolios, as they could all be selected on the basis that they share the same benchmark, this cannot be assumed when looking at the entire population of portfolios in an ASISA category, which is what we will be doing in this paper.

It is also important to realise that in many situations the results do not differ at all, or materially, whether you use active or nominal returns. This may not seem obvious or intuitive, but can be easily demonstrated by running the analysis on both sets of returns and seeing if the results differ materially.

If you use active returns for your analysis, you should ensure that the benchmark is appropriate for all the portfolios being compared, as the analysis could otherwise be invalidated for the portfolios for which the benchmark is not appropriate. An appropriate benchmark must be one that the manager manages the portfolio against, so this is not a trivial hurdle.

3. MACHINE LEARNING TECHNIQUES

This is not an exhaustive, or a very detailed and technical exposition of all of these methods, but rather a brief introduction with some explanations to aid understanding. There are, for example, many 'hyper parameters' that can be 'chosen' for many of these algorithms that are not mentioned. Please do not misunderstand this as a reflection that these decisions are not important, as they could lead to very different results, especially when considering the finer detail. Unfortunately, the topic is just too broad and detailed to cover completely in this paper.

3.1 Unsupervised learning

Unsupervised learning refers to a set of techniques that are applied on unlabelled data, where there is no target label or value as the result of all of the feature data. You simply provide the algorithm with the data, raw or manipulated, and ask it to provide insights into the structure of the data, including patterns and relationships. In our case the raw data are monthly returns (features) for a selection of portfolios (samples), but it could have been holdings or transactions, as mentioned in the previous section.

3.1.1 HIERARCHICAL CLUSTERING

Hierarchical clustering is a technique for grouping portfolios (any elements in general, but we are focused here on portfolios) based on their similarities or differences. If you have a distance matrix between every pair of portfolios, you could join the two closest portfolios into a single cluster, and then calculate the centroid of that cluster. You could then continue by looking for the next two closest portfolios, or a portfolio and the centroid of an existing cluster. In this way, you will begin with every portfolio being in its own cluster, and eventually end with every portfolio being in a single cluster.

This allows you to stop anywhere along the way when you have a required number of clusters, or to just analyse how portfolios cluster together by considering how similar they are. The distance between portfolios can be calculated in many different ways, but Euclidean distance of returns in n-dimensional space, where n is the number of monthly returns, is typical. You could also perform a Principal Component Analysis (PCA) to reduce the number of dimensions first, and then calculate the Euclidean distance in the reduced dimensionality space.

There are many other different ways to calculate distances, and to group the elements into a hierarchical structure, but we will focus on the method described above. Technically, this is known as an unsupervised learning method as it discovers the structure in the data without requiring any sort of label, so you do not need to teach the algorithm the structure of the data.

If monthly returns are a valid way to think about the similarities and differences between portfolios, this method could yield interesting results in terms of classifying portfolios based on their returns. It is reasonable, for example, to assume that high equity portfolios will experience higher/lower returns than low equity portfolios, when equity returns are high/low (negative). This should place high equity portfolios closer to each other than to low equity portfolios, and two clusters could be expected to form if the period under consideration included periods of high and low equity returns, when analysing portfolios with high, and portfolios with low equity allocations.

The same could be said for any other asset class that can experience a large range of returns. In reality, it is not the range of returns that is the single most important factor. If you were to compare money market portfolios only, you could still perform hierarchical clustering, even though the range of returns would be considered quite low. It is the relative difference in returns that is important. So, while comparing equity portfolios to money market portfolios would be dominated by how much equity the portfolios have, thereby clustering equity portfolios together and money market portfolios together, comparing just money market portfolios would still produce a hierarchical structure, albeit that the distances between portfolios would be a lot shorter.

Unfortunately, this method doesn't specify how many clusters should be considered, and the number of clusters can range from one to as many clusters as there are portfolios in the data being considered, with each portfolio in its own cluster. The hierarchical clustering produced by this method can be visualised using a dendrogram, as will be seen below.

This method and the visualisation thereof using a dendrogram can be combined with a correlation matrix that shows all pairwise correlations. This can be useful, if not overwhelming, as the relationship between most pairs of portfolios is lost once the hierarchical clustering has been performed. It is however very useful because large correlation matrices contain too much information for a user to make sense of.

3.1.2 K-MEANS CLUSTERING

With K-means clustering on the other hand, you can specify exactly how many clusters to divide the data into. This could be easy to decide if you could first visualise the data, if for example it only had two or three dimensions, but is a lot more difficult for higher dimensional data. Simply deciding on an arbitrary number of clusters may not yield the results you want as some clusters may only have one or two portfolios, if they are sufficiently far (different) from the rest of the portfolios.

The method is otherwise very similar to hierarchical clustering, although it won't tell you the relationship between any two portfolios, just that they are either in the same cluster or not. Again, it could be combined with the correlation matrix to overcome this shortcoming, by grouping the portfolios in the same cluster together in the matrix. This would allow you to not only see the correlation between portfolios in the same cluster, but also across clusters.

The actual algorithm is fairly straightforward. You begin by randomly selecting the same number of portfolios as the number of clusters that you want. These will represent the centroids of the clusters to begin. You then allocate each portfolio to the closest cluster centroid, and then calculate the new cluster centroid based on all the portfolios that were closest to that cluster. As the centroid of each cluster is likely to have changed as it went from having a single portfolio to perhaps having many, you repeat the process.

You keep repeating the process until portfolios no longer move from their existing clusters, which could take many iterations depending on the number of portfolios and clusters. Although this will give you a result, it may not be the 'best' result. You sum the distances of every portfolio from their cluster centroid, and rerun the entire process from the very beginning to see if you can get a lower sum. You need to decide how many times you want to run the process before settling on the lowest sum achieved, and this will be the result of your K-means clustering exercise.

Unfortunately, there is no useful way to visualise the results of this technique, as you simply have the elements of each cluster, but at least you get to see which portfolios fall into which cluster, implying some level of similarity, and which fall into separate clusters, implying differences.

This is also an unsupervised learning technique as you are not providing any label for the portfolios. You are simply asking the algorithm to consider the monthly returns and provide a structure for the data, in this case, which portfolios belong in which cluster, implying their similarity to some extent.

It is important to note that you don't have a sense for how similar portfolios within a cluster really are, although you could look at their distances to establish this. If some of the clusters only have one or two portfolios that were very different from all the others, you could end up with one big cluster that is not very homogeneous, so it is important to understand the overall structure when inferring similarities within a cluster.

3.1.3 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a technique that reduces highly dimensional data into new dimensions (linear combinations of those provided) that describe the most variability in the data, with the first principal component describing the most variability, and each subsequent principal component describing the next most, while being orthogonal to all previous principal components.

This can be very useful on its own to describe the similarities and differences between portfolios, by focusing on these primary principal components, if they explain a high proportion of the variability of the portfolios' returns. It can also be very useful as a dimensionality reduction technique, where we can focus on a small number of dimensions to describe these differences, instead of having to comprehend 60 months' worth of returns for every single portfolio, and how they are similar or differ.

By way of example, when applying this technique to the Multi Asset tier of categories and their portfolios, the first principal component could reduce 60 months of returns to an equity dimension (principal component), as this will help to describe the most variability in portfolio returns. Thinking back to March and April 2020, the single most important factor that drove portfolio returns was the amount of equity exposure in those portfolios.

It is unfortunately not that straightforward, as each principal component will be a combination of all 60 months of returns, and the interpretation of what the principal component represents can be somewhat subjective. It could for example include both local equity and global equity. It could also include growth assets like property, although this is likely to have been reflected in a separate principal component (possibly the second), because of the huge impact from property exposure over the past five years.

So while this technique is very useful in uncovering the structure of data, it does require exploration and judgement to confer meaning to the structure. There are several ways that this can be ameliorated. The first is to know something about the underlying data being analysed. Knowing the portfolios and what makes them unique and different can sometimes help understand why the principal components are what they are.

The second is to include pure proxies for the dimensions that we think will be important drivers of the differences that will be observed. You could for example include an equity index and a property index in a multi asset portfolio analysis to see if these indices were reflected in the principal components.

Unfortunately, including these indices in the data can change the structure as they now form part of the data and they may be the cause of the most variability in the data, especially because they represent pure exposure to that index. A better approach would be to use the linear combinations of returns that the principal component represents, what are technically known as the loadings, to project the indices onto the visualisation.

Yet another method would be to project the features loadings, in this case the months, onto the visualisation. Unfortunately this may be less useful when dealing with features like months, as they may not be very useful e.g. March 2020 may not be very meaningful to the typical reader as different asset classes behaved in different ways. You also have the problem of many features that are being projected, and cluttering the chart, so you may want to only project the features with the largest loadings.

Finally, it is also important to understand how much of the variability is explained by each of the principal components. Clearly a principal component that explains 80% of the variability is much more important than one that explains only 10%. When using the principal components to visualise the data, you also want to understand how much of the variability is explained by each of the principal components, to establish whether the visualisation captures enough of the variability in the data.

3.1.4 SINGULAR VALUE DECOMPOSITION

Singular value decomposition (SVD) is similar to PCA in that it is used to understand the structure of data based on the inputs, but it differs in that you do not need to decide the variable of interest (portfolios) and the features of the data (monthly returns) separately as in PCA. In PCA, you could decide that you wanted to understand the structure of portfolios, by looking at monthly returns which are called the features of the data. You could instead decide that you wanted to understand the structure of monthly returns based on a bunch of portfolios, and these would now be the features. In both cases, you could get very different results if you changed the features.

For example, we use the last five years of monthly returns in our analysis. We may get different results if we instead chose to use returns from ten years ago that included the global financial crisis, as different asset classes behaved differently to how they behaved during the Covid pandemic. Similarly, we would get very different results if we used equity portfolios to understand the structure of monthly returns over the past five years, compared to the results we would get if we used interest-bearing portfolios instead, or multi asset portfolios.

In SVD, you don't need to decide on the variable of interest and the features, as it provides the structure of data using both, allowing you to interpret the data from either dimension as they are linked. So in our case, the structure of the portfolios is based on the monthly returns, and the structure of the monthly returns is based on the portfolios, all provided for in a single solution that again derives the principal components that explain the most variability in the data.

Although we wanted to cover this method in this section as it is very useful in analysing returns and even holdings data, and preferable in many ways to PCA, we will not be using it in the analysis section as the results are very similar to what we would get from PCA.

3.2 Supervised learning

Supervised learning refers to a set of techniques that are applied on labelled data, and hence the learning is supervised. In this case you provide the algorithm with the input data, raw or manipulated, and the output label, and ask the algorithm to learn the patterns that link the input data with the output label. In our case the raw data are monthly returns (features) for a number of portfolios (samples), but it could be holdings or transactions, as mentioned in the previous section, and the output label will be the ASISA category to which the portfolios belong.

3.2.1 K NEAREST NEIGHBOURS CLASSIFIERS

Whereas hierarchical clustering, K-means clustering, PCA and SVD are referred to as unsupervised learning techniques, because you ask the technique to give you the structure of the data based on the values supplied (returns), K nearest neighbours (KNN) differs in that you ask the model to learn the structure of the data based on the features and the labels provided, and hence it is a supervised learning technique.

For example, let's say you are interested in finding the 'best' category to classify a specific portfolio, assuming you either don't know it, or you want to ignore its current category, in case it has been 'misclassified'. You could provide the labels for many different portfolios, along with the returns of each of those portfolios, and ask the model to predict the label for your portfolio of interest.

The technique assumes that the data is correctly labelled to start, otherwise it will be learning from 'bad' data, and as the adage goes, garbage in equals garbage out. The problem we are trying to address however is exactly this: is the existing classification appropriate or are some portfolios potentially misclassified?

It is important to note that by 'misclassified', we do not mean some errors are being made with the classification, or even deliberate action by managers when deciding in which category to include their portfolios. We mean that the portfolios are not like others in the category, which is trying to group homogeneous portfolios for the benefit of investors and their advisers.

We therefore apply this technique in a different way to how it was intended to be used. We ask the technique to predict the category of each portfolio for which we already know the category. The implication is that if the prediction is different to reality, this could imply that the portfolio is misclassified. This then gives us information on the quality of the classification standard, if it allows portfolios that are very different to all have the same classification.

This method is not useful in deciding how many clusters (categories) of portfolios there should be, but rather under which category a particular portfolio should be classified. Using the process above, you could end up with certain categories disappearing as the portfolios that were previously in those categories get reclassified into categories into which they fall more closely, and we see this in the analysis below.

You do need to be careful that low frequency categories may get under-represented. This often appears as bias in the results as over-represented categories get chosen more frequently as there are more neighbours to be close to.

So although this technique is usually used as a predictive tool, to make predictions for new data, we will use it in a different way to reclassify portfolios, or think about how they are classified. Although we begin with the assumption that the category for each portfolio is correct, we do not assume that the incorrect 'recall' of the category of a portfolio when doing a prediction means that the model has made a 'wrong' prediction, but rather that the category being 'recalled' is a better fit to the data. We define what recall means for this technique below when we apply it.

3.2.2 DECISION TREE CLASSIFIERS

A decision tree is one of the most commonly used supervised machine learning techniques, and there are two main reasons for this. First, it is intuitive, and easy to understand the results (it is not a black box), and second, it provides great results in many cases 'straight out of the box'. The algorithm is pretty straightforward. It considers each feature in the data and chooses both the feature and the value on which to split the data into two different branches of the tree that will yield the best results i.e. learn the best way to classify the data.

As with KNN, this is a supervised learning technique as we supply the labels (categories) for the data (portfolios), along with the features (monthly returns). Traditionally, you could then apply the model to new unseen data (portfolio returns) to decide on the category (label) for the portfolio. As with KNN, we will not be using this technique as such.

We make the assumption that the labels are correct to begin with, but as with KNN we use the trained model on the same data set to see how it predicts the category of each portfolio. A failure to predict is not seen as a model error, but rather as a potential misclassification of the portfolios. This can be very useful, as will be seen below, in deciding on the quality of the existing classification standard.

We should not expect to get the same results from these two techniques, as the algorithms are completely different, but intuitively, we may expect to get similar outcomes and conclusions around certain portfolios possibly being misclassified and the classification standard possibly having some shortcomings.

This method also has many hyperparameters, which we explore in more detail when we run the analysis, providing some justifications for the values chosen. The same issues exist as with KNN in that the parameters can lead to overfitting, in which case the predictions will be the same as the current classification. For categories that have a big differences in the number of portfolios in the each category, like Equity, the parameters could also lead to all (or most) portfolios being predicted to fall into the category with the most portfolios, which may be what we are looking for in our specific use case, but would generally not be ideal.

4. ANALYSIS OF CURRENT CLASSIFICATIONS

4.1 Risk versus return scatterplots for individual categories

So let us begin by using the traditional tools of the trade, and looking at risk versus return scatterplots for each of the South African categories. We compare and contrast portfolios within each category first, and then consider how these categories may overlap. This should help to illustrate two important points. First, some of the categories are not very homogeneous, and may contain some portfolios that may be considered outliers. Secondly these categories may contain a high degree of overlap with each other, and some portfolios perhaps belong in other categories. Often the problem of demarcation is not an easy one to solve, but perhaps we can do better than the current classification standard.

We use monthly returns for the past five years to the end of November 2021. We could have easily decided to include either more or less data, but wanted to strike a balance of more data providing more information for comparison, but potentially being less relevant, than less and more recent data. The exercise could be repeated using three years of data without loss of some of the generality, but we wouldn't advise going any shorter.

4.1.1 SOUTH AFRICAN – MULTI ASSET

> *Multi Asset portfolios are portfolios that invest in a wide spread of investments in the equity, bond, money and property markets to maximise total returns (comprising capital and income growth) over the long term.*

We begin with the largest of the South African 'Second Tier' classifications: Multi Asset.

4.1.1.1 SOUTH AFRICAN – MULTI ASSET – HIGH EQUITY

> *Multi Asset – High Equity portfolios – These portfolios invest in a spectrum of investments in the equity, bond, money, or property markets. These portfolios tend to have an increased probability of short term volatility, aim to maximise long term capital growth and can have a maximum effective equity exposure (including international equity) of up to 75% and a maximum effective property exposure (including international property) of up to 25% of the market value of the portfolio. The underlying risk and return objectives of individual portfolios may vary as dictated by each portfolio's mandate and stated investment objective and strategy.*

The largest category overall, and within the Multi Asset tier, by number of portfolios and AuM, is the High Equity category, with 210 portfolios and AuM of R701 billion, of which 152 portfolios have five years of returns.

Figure 2 is a scatterplot of risk versus return, where the x-axis is volatility (the annualised standard deviation of monthly returns for each portfolio), and the y-axis is the annualised (time-weighted) return of the portfolio. Each dot represents one portfolio. The volatility ranges from just above 5% p.a. to just under 15% p.a., while the returns range from 1.5% p.a. to 13.2% p.a. This volatility range is important as we will see the overlap with other categories, so please keep it in the back of your mind.

Although two portfolios that are close to each other are probably similar in many different ways, they could also be different and just have happened to have achieved the same numbers for risk and return over the same period. What is however unlikely is for two portfolios that are far from each other to be similar in any meaningful way. This is especially true of portfolios that differ along the x-axis. So while the portfolios are distributed along a continuum from lowish (5%) to highish volatility (15%), you could reasonably conclude that portfolios at the extremes are not like portfolios at the other extreme.

This presents a problem if the objective of a category is to try to create a fairly homogeneous grouping of portfolios. Also note the magnitude of the difference in risk from 5% to 15% (three times more). A two standard deviation event in the left tail of distribution of returns (a fairly common occurrence) would imply a negative return of –10% and –30% respectively. That is a huge difference when thinking about risk.

The problem lies with the definition of the category, which puts an upper limit on the equity allocation (both local and global combined) of 75%, but a lower limit of 0% (essentially, no lower limit). Again, while two portfolios with total equity around 70% may be considered similar (along certain dimensions), as could two portfolios with total equity around 40%, two portfolios with a difference of 30% total equity exposure could hardly be considered to be similar.

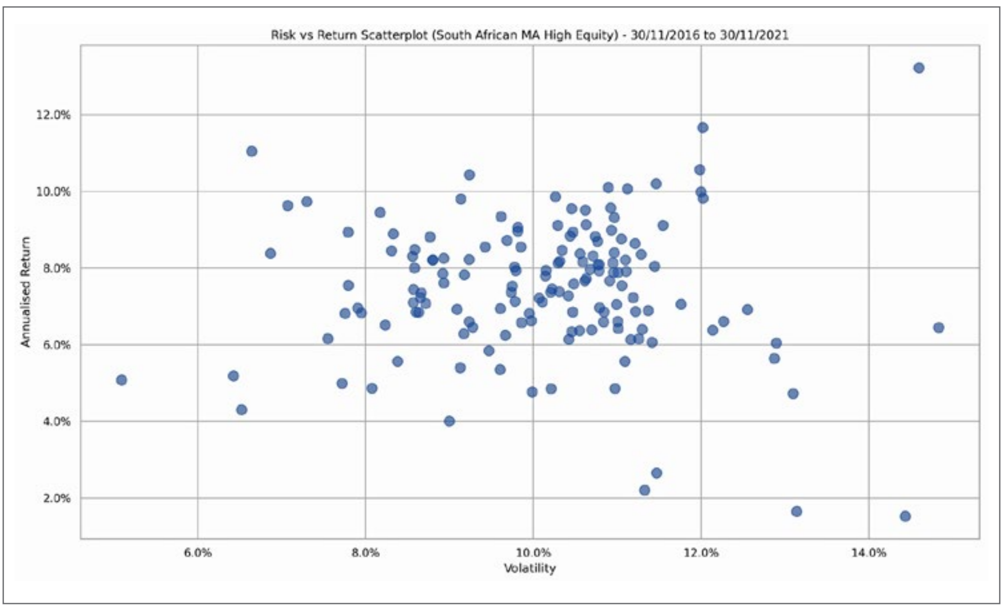


FIGURE 2 Risk vs return – multi asset – high equity

4.1.1.2 SOUTH AFRICAN – MULTI ASSET – MEDIUM EQUITY

> *Multi Asset – Medium Equity portfolios – These portfolios invest in a spectrum of investments in the equity, bond, money, or property markets. These portfolios tend to display average volatility, aim for medium to long term capital growth and can have a maximum effective equity exposure (including international equity) of up to 60% and a maximum effective property exposure (including international property) of up to 25% of the market value of the portfolio. The underlying risk and return objectives of individual portfolios may vary as dictated by each portfolio's mandate and stated investment objective and strategy.*

Now let us compare and contrast the above results to the Medium Equity category with 98 portfolios and AuM of R99 billion, of which 72 portfolios have five years of returns. The Medium Equity category has a maximum total equity exposure of 60%, versus the 75% for the High Equity category.

Figure 3 is the risk versus return scatterplot over five years for the Medium Equity category. We only find a single portfolio with lower volatility than the lowest volatility portfolio in the High Equity category, with volatility around 3.5%. The highest volatility portfolio also has lower volatility than the highest volatility portfolio in the High Equity category, but not much lower at around 11.5%. It is certainly way above the lowest volatility portfolio in the high equity category, and comparing the dots on both figures should reveal a high degree of overlap in both returns and volatility, but we make this even clearer when we include them in a single chart below.

This implies that many portfolios that are in different categories are actually similar along these two dimensions. You could in fact reasonably argue that two portfolios in these two different categories are more similar to each other, than they are to other portfolios in their same categories. Clearly the category label is less useful in describing these portfolios than these two dimensions. We see this pattern of overlap and similarity again when we consider the Low Equity category, and yet again in the Income category.

Let us put this in practical terms so that the problem becomes a lot clearer. An adviser is looking for an appropriate moderate risk solution (whatever that means) for their client, whom they have risk profiled. Should they look in the Medium Equity category, and feel comfortable that all the portfolios in that category have similar risk? Should they look in other categories, like the High Equity and the Low Equity, which will both have portfolios with risk similar to the portfolios in the Medium Equity category. From this analysis it should be obvious that looking at the category first may not be the 'best' approach.

If you instead decided to use some other measure of risk to find the appropriate solution for the risk profiled investor, say total equity exposure, then you would be better served by looking at the total equity exposure of each portfolio, not only at a single point in time, but over a sufficiently long period of time. In this case, the category of the portfolios is again

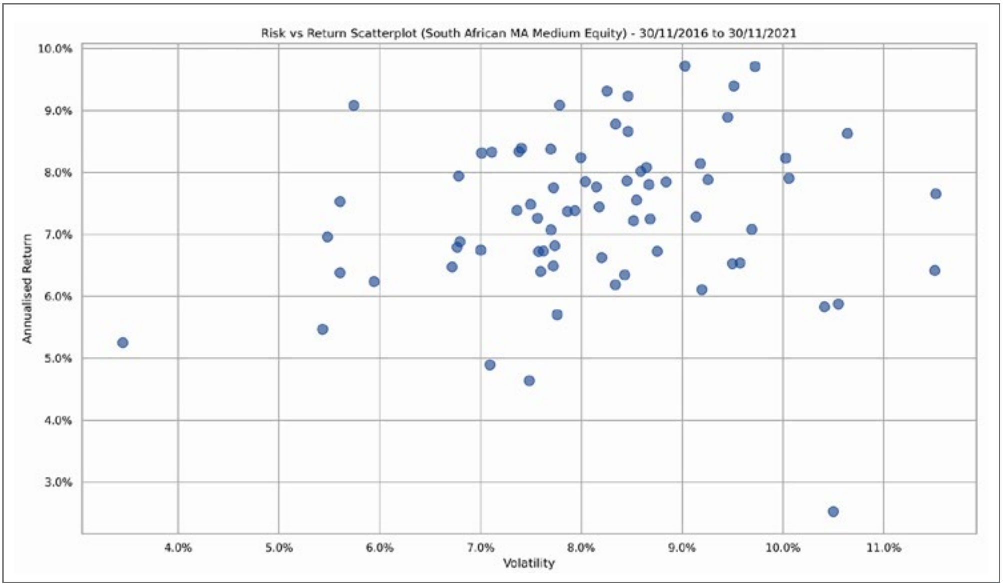


FIGURE 3 Risk vs return – multi asset – medium equity

useless, except to put a theoretical upper limit on the total equity exposure, and hence should not be used by advisers when looking for portfolios.

4.1.1.3 SOUTH AFRICAN – MULTI ASSET – LOW EQUITY

> *Multi Asset – Low Equity portfolios – These portfolios invest in a spectrum of investments in the equity, bond, money, or property markets. These portfolios tend to display reduced short term volatility, aim for long term capital growth and can have a maximum effective equity exposure (including international equity) of up to 40% and a maximum effective property exposure (including international property) of up to 25% of the market value of the portfolio. The underlying risk and return objectives of individual portfolios may vary as dictated by each portfolio's mandate and stated investment objective and strategy.*

Let us compare and contrast the above results to the Low Equity category with 158 portfolios and AuM of R282 billion, of which 118 portfolios have five years of returns. The Low Equity category has a maximum total equity exposure of 40%, versus the 75% for the High Equity category, and 60% for the Medium Equity category.

Figure 4 shows the risk versus return for the past five years. While the annualised returns range from about 2.5% p.a. to almost 10% p.a., the volatility ranges from about 1.5% p.a. to around 11% p.a. (more than seven times as much). It should be obvious that many portfolios in this low equity category are more risky than many portfolios in the

Medium Equity category, and even many in the High Equity category that had volatility as low as 5% p.a.

You will again note the overlap with the Medium Equity category, which is perhaps by now not surprising. What may however be surprising is the overlap with the High Equity category. This will again be a lot clearer when we present this data in a single chart. The reasons for this may also not be obvious given the total equity exposure limit, but this will be explained below in the recommendations.

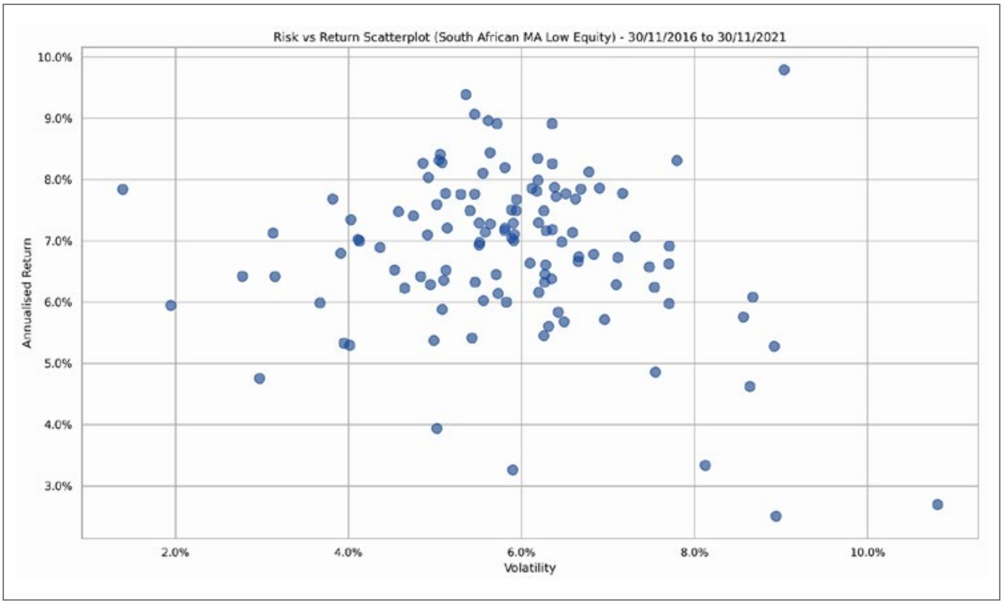


FIGURE 4 Risk vs return – multi asset – low equity

4.1.1.4 SOUTH AFRICAN – MULTI ASSET – INCOME

> *Multi Asset – Income portfolios – These portfolios invest in a spectrum of equity, bond, money market, or real estate markets with the primary objective of maximising income. The underlying risk and return objectives of individual portfolios may vary as dictated by each portfolios mandate and stated investment objective and strategy. These portfolios can have a maximum effective equity exposure (including international equity) of up to 10% and a maximum effective property exposure (including international property) of up to 25% of the market value of the portfolio.*

Now we compare and contrast the above results to the Income category with 110 portfolios and AuM of R292 billion, of which 67 portfolios have five years of returns. The Income category has a maximum total equity exposure of 10%, versus the 75% for the High Equity category, 60% for the Medium Equity category, and 40% for the Low Equity category.

Figure 5 shows the risk versus return for the past five years. While the annualised returns range from about 3% p.a. to just over 10% p.a., the volatility ranges from less than 0.5% p.a. to just under 5% p.a. (ten times as much). It should be obvious that many portfolios in this income category are more risky than some portfolios in the Low Equity category, and even one in the Medium Equity category. Again it may not be obvious how this could happen in this category with a much lower total equity exposure limit, but we explain this below, so please bear with us.

These four categories considered so far are meant to be well defined with increasing or decreasing total equity exposure upper limits. There is a category which is even more loosely defined though, the Flexible category. We should probably consider where the portfolios in this category fall on the above risk versus return spectrum, to again check for overlap with the above categories, and whether it represents a homogeneous collection of portfolios, so we will do that next.

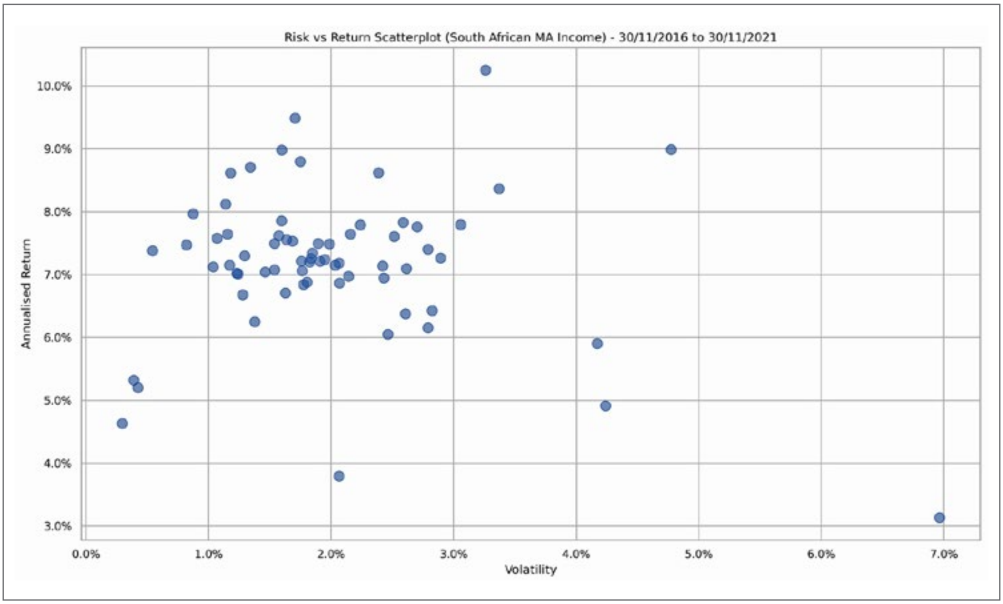


Figure 5 Risk vs return – multi asset – income

4.1.1.5 SOUTH AFRICAN – MULTI ASSET – FLEXIBLE

Multi Asset – Flexible portfolios – These portfolios invest in a flexible combination of investments in the equity, bond, money and property markets. The underlying risk and return objectives of individual portfolios may vary as dictated by each portfolio's mandate and stated investment objective and strategy. These portfolios may be aggressively managed with assets being shifted between the various markets and asset classes to reflect changing economic and market conditions and *the manager is accorded a significant degree of discretion over asset allocation to maximise total returns over the long term.*

Now let us compare and contrast the above results to the Flexible category with 52 portfolios and AuM of R50 billion, of which only 37 portfolios have five years of returns. The Flexible category has no upper or lower limit on total equity exposure, which means that in theory these portfolios could sit neatly in any of the categories covered above. The exception would be portfolios with more than 75% total equity exposure, which could sit in the Equity General category, with the exception of portfolios with more than 75% but less than 80% total equity exposure, of which there would be few, if any.

Figure 6 shows the risk versus return for the past five years. While the annualised returns range from about –6% p.a. to just under 15% p.a., the volatility ranges from about 0.5% p.a. to about 23% p.a. (around 50 times as much). It should be obvious that the portfolios in this category range from as low as the Income category, to even higher than the High Equity category. Please don't interpret this as individual portfolios having variable volatility over time, as the flexible label may imply, but rather that this category can contain portfolios with varying objectives, i.e. it is the category that is flexible, and not necessarily the portfolios therein.

You have to ask what the point of this category is, if you have portfolios with such a wide range of volatility, and varying objectives. Clearly this is the least homogeneous group of portfolios, and the function of this category is not immediately obvious in the context of the classification standard. This would perhaps be an appropriate category for portfolios

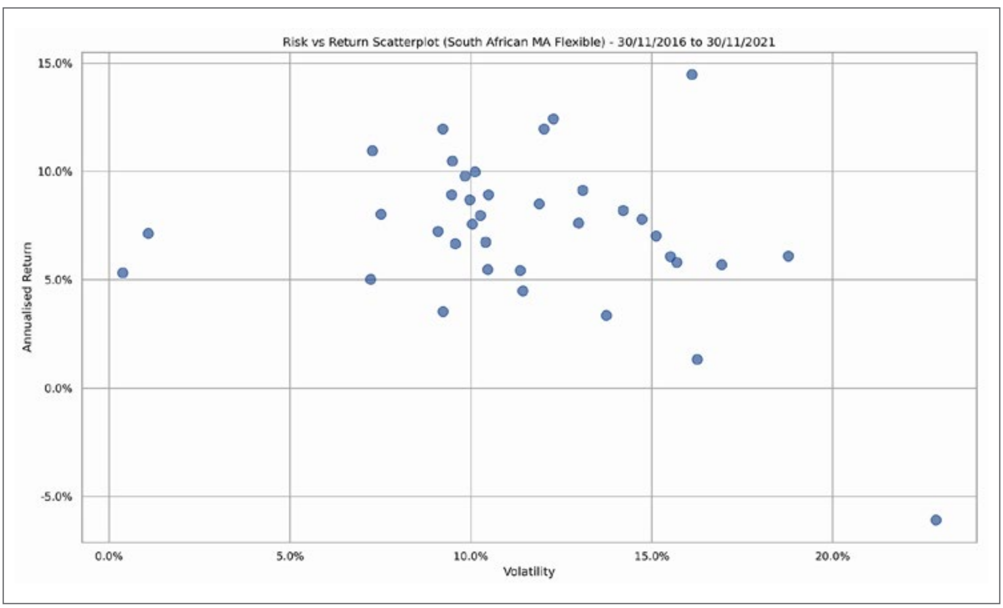


Figure 6 Risk vs return – multi asset – flexible

that would allocate between 0% and 100% to equities over different time periods, making the portfolios flexible, but not appropriate for portfolios that have a consistently high or low equity allocation, as there are already categories for those.

Finally, if we plot these portfolios alongside the portfolios in the categories covered above, you will see that they get lost in the mix. We will get to these comparisons in the next section. One thing does however appear different, and that is that some of the portfolios in this category have higher volatility than some of the portfolios in the High Equity category.

For this reason, and the reason alluded to above that the Equity categories have a minimum allocation to total equities of 80%, we should probably compare the portfolios in the Flexible category with the portfolios in the Equity General category as well. We do that below, although note that this is not a Multi Asset 'Second Tier' classification by definition, although we will discuss in what ways it can be considered to be Multi Asset.

4.1.1.6 SOUTH AFRICAN – MULTI ASSET – TARGET DATE

> *Multi Asset – Target Date portfolios – These portfolios invest in a spectrum of equity, bond, money market, or real estate markets where the asset mix changes over time in a predetermined manner as the target date approaches. Due to the change in asset mix over time, portfolios in this category cannot be compared and consequently cannot be ranked.*

Finally for the Multi Asset tier, let us now compare the results above to the Target Date category with only eight portfolios and AuM of R8 billion, all from a single service provider, and all of them sitting neatly within the range of returns observed for the other categories as shown in Figure 7. We may suspect that this category should probably not even exist, although the portfolios may not sit neatly in one of the other categories for their entire life.

4.1.2 SOUTH AFRICAN – EQUITY

> *These portfolios invest a minimum of 80% of the market value of the portfolios in equities and generally seek maximum capital appreciation as their primary goal.*

So let us turn our attention to the next biggest South African 'Second Tier' classification, Equity.

4.1.2.1 SOUTH AFRICAN – EQUITY – GENERAL

> *Equity – General portfolios – These portfolios invest in selected shares across all industry groups as well as across the range of large, mid and smaller market capitalisation shares. While the managers of these portfolios may subscribe to*

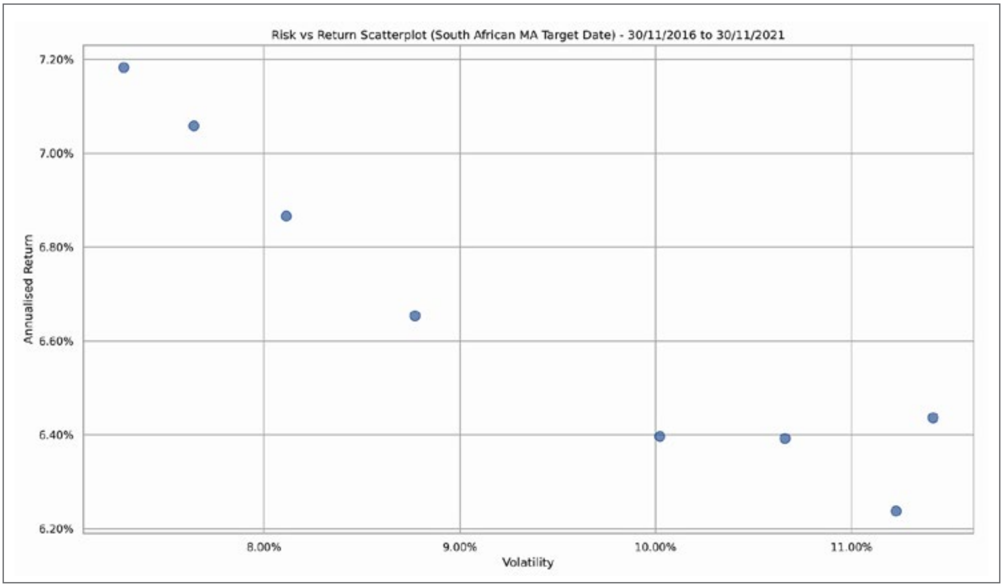


FIGURE 7 Risk vs return – multi asset – target date

different investment styles or approaches, their intent is to produce a risk/return profile that is comparable with the risk/return profile of the overall JSE equities market. The portfolios in this category offer medium to long-term capital growth as their primary investment objective.

Let us begin with the General category with 170 portfolios and AuM of R343 billion, of which 117 portfolios have five years of returns. The category has a lower limit on total equity exposure of 80%, which implies at least some overlap with the Flexible category, but none of the other Multi Asset categories.

For the General category, although one could argue that this is not a multi asset category, there are a couple of reasons why this comparison will be useful. The General category only requires 80% investment in equity, which means that it could include other assets up to 20%, so it could in fact be considered the highest equity multi asset category, but more on this below.

Figure 8 shows the risk versus return for the past five years. While the annualised returns range from more negative than –1% p.a. to just under 14% p.a., the volatility ranges from just under 10% p.a. to just over 22% p.a. (around two times as much). While the portfolios in this category overlap with some of the portfolios in the Flexible category, as expected given the maximum total equity allowances, they also overlap to some extent with the High Equity category (where some portfolios had volatility over 14%), which is perhaps strange given the non-overlapping total equity exposure limits, but this is explored and explained below.

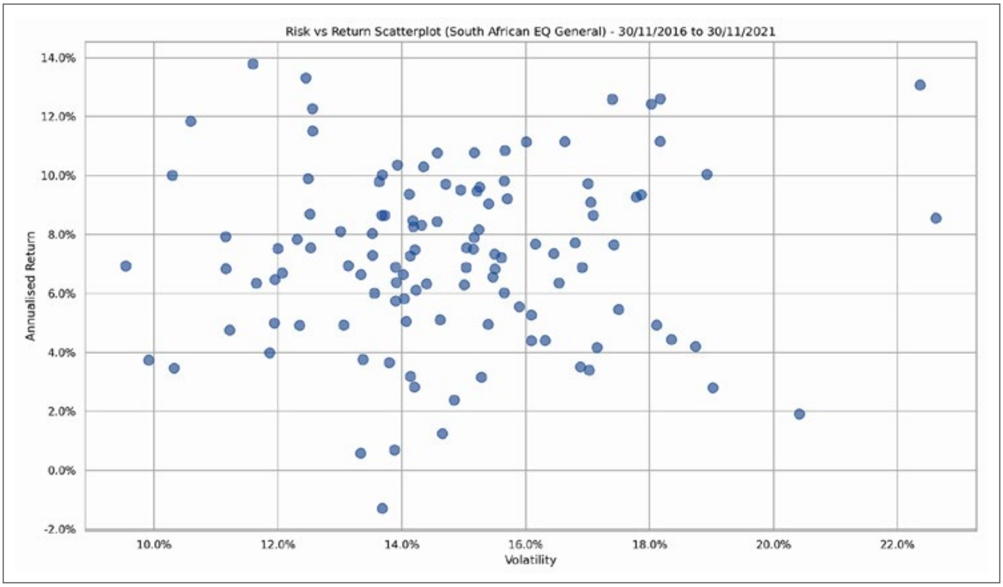


FIGURE 8 Risk vs return – equity – general

You could argue that the range of volatility is still very high, and whether this category represents a homogeneous group of portfolios with such a wide range of volatility. We will see when exploring this data with other techniques that you could certainly cluster these portfolios in various ways to create more homogeneous groups, but perhaps for now just appreciate how the volatility range compares to some of the Multi Asset categories covered above e.g. three times for the High Equity category (from 5% to 15%), and 50 times for the Flexible category (from 0.5% to almost 25%). Also keep this range in mind when considering the other Equity categories discussed below.

4.1.2.2 SOUTH AFRICAN – EQUITY – LARGE CAP

> *Equity – Large cap portfolios – These portfolios invest at least 80% of the market value of the portfolios in large market capitalisation shares which have a market capitalisation greater than or equal to the company with the lowest market capitalisation in the FTSE/JSE Large Cap Index, or an appropriate foreign index published by an exchange. 100% of share purchases must be in this investable universe at time of purchase.*

Now let us look at the Large Cap category with 12 portfolios and AuM of R15 billion, of which 11 portfolios have five years of returns. The volatilities of these portfolios sit well within the range of the portfolios for the General category as can be seen in Figure 9. While the annualised returns range from just under 4% p.a. to just over 11% p.a., the volatility ranges from just under 15% p.a. to just over 18.5% p.a. The few portfolios grouped together at the top left are probably all index trackers, whereas the rest are probably active managers trying to do something different.

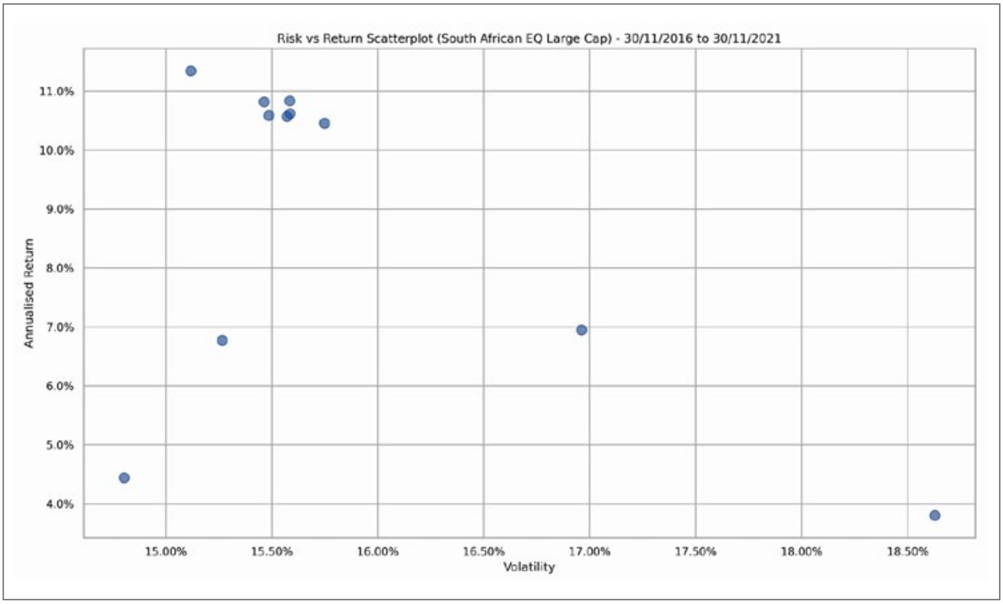


FIGURE 9 Risk vs return – equity – large cap

4.1.2.3 SOUTH AFRICAN – EQUITY – MID/SMALL CAP

> *Equity – Mid & Small cap portfolios – These portfolios invest at least 80% of the market value of the portfolios in shares which are in the FTSE/JSE Small Cap Index or FTSE/JSE Mid Cap Index, or in shares that have a market capitalisation smaller than the company with the lowest market capitalisation in the FTSE/JSE Large Cap Index, or an appropriate foreign index published by an exchange. 100% of share purchases must be in this investable universe at time of purchase. Due to both the nature and focus of these portfolios, they may be more volatile than portfolios that are diversified across the broader market.*

The Mid/Small Cap category only has seven portfolios and current AuM of R6.7 billion, of which 6 portfolios have five years of returns. Figure 10 shows the risk versus return scatterplot and the results may surprise some readers. Although the annualised returns are not very dissimilar to the General and Large Cap categories, ranging from just under 2% p.a. to just over 9% p.a., the volatility is perhaps unexpectedly low if you had to assume that mid and small cap shares are more risky. The volatility range is however also in line with the other equity categories, ranging from a low below 13% p.a. to a high above 19% p.a.

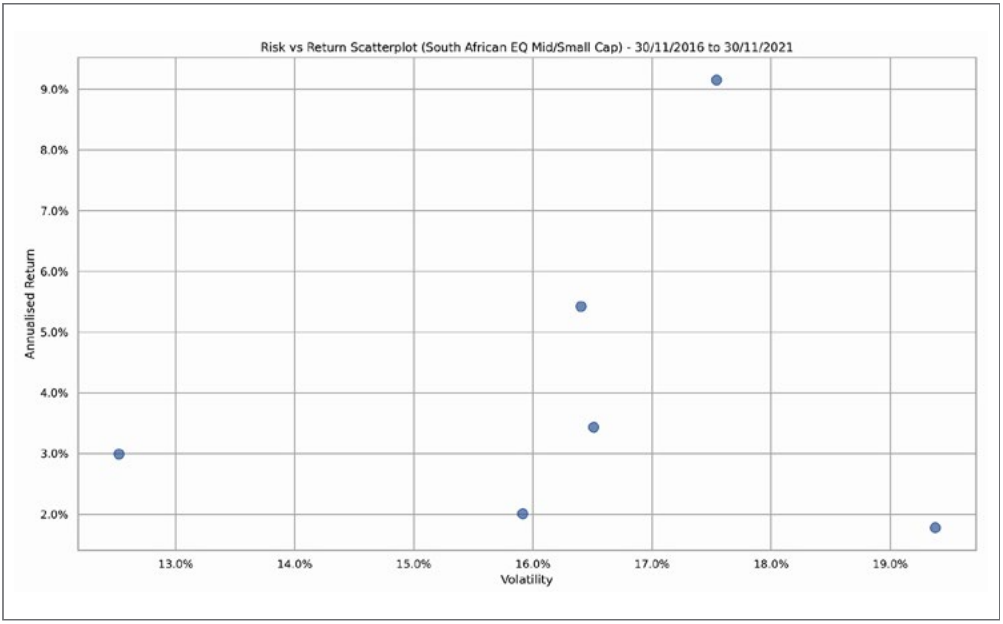


FIGURE 10 Risk vs return – equity – mid/small cap

4.1.2.4 SOUTH AFRICAN – EQUITY – RESOURCES

> *Equity – Resources portfolios – These portfolios invest at least 80% of the market value of the portfolios in shares listed in the FTSE/JSE Oil & Gas and Basic Materials industry groups or in a similar sector of an international stock exchange. Up to 10% of a portfolio may be invested in shares outside the defined sectors in companies that conduct similar business activities as those in the defined sectors. Due to both the nature and focus of these portfolios, they may be more volatile than portfolios that are diversified across a wider range of FTSE/JSE industry groups.*

By far the standout equity category will be Resources, with six portfolios all of which have five years of returns, and AuM of R5.5 billion. From Figure 11 we can see the annualised returns range from below 16% p.a. to over 28% p.a. reflecting the run in resources over parts of the last five years. The volatility is distinctively different from the rest of the equity categories ranging from a low of just over 20% p.a. (not too far off), to a high of just under 40% p.a. (significantly higher than any other number we have seen). This is actually a Gold portfolio, so perhaps not as well diversified as some of the other resources portfolios, and hence the higher volatility.

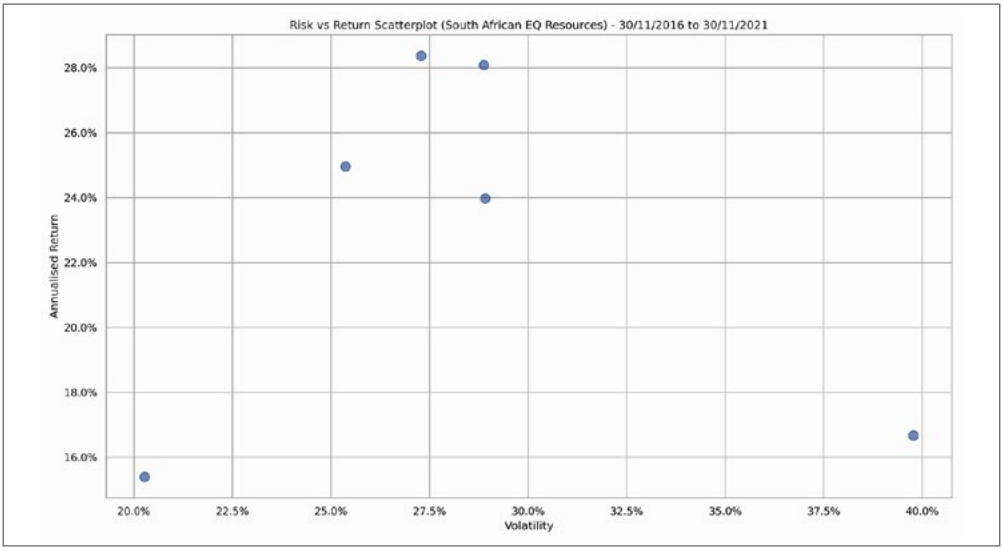


FIGURE 11 Risk vs return – equity – resources

4.1.2.5 SOUTH AFRICAN – EQUITY – INDUSTRIAL

> *Equity – Industrial portfolios – These portfolios invest at least 80% of the market value of the portfolios in industrial shares listed on the Johannesburg Stock Exchange or in a similar sector of an international stock exchange. Industrial shares include all companies listed on the JSE other than those shares listed in the FTSE/JSE Oil & Gas, Basic Materials, and Financials industry groups.*

The Industrial category has even fewer portfolios at only three, all of which have five years of history, and AuM of R2.4 billion. From Figure 12 we can see the annualised returns range from 5% p.a. to 8% p.a., and the volatility ranges from under 14% p.a. to just under 16% p.a. Perhaps the ranges are so narrow because there are so few portfolios, but this category and these portfolios should not exist, despite the AuM.

4.1.2.6 SOUTH AFRICAN – EQUITY – FINANCIAL

> *Equity – Financial portfolios – These portfolios invest at least 80% of the market value of the portfolios in shares listed in the FTSE/JSE Financials industry group or in a similar sector of an international stock exchange. Up to 10% of a portfolio may be invested in shares outside the defined sectors in companies that conduct similar business activities as those in the defined sectors. Due to both the nature and focus of these portfolios they may be more volatile than portfolios that are diversified across a wider range of FTSE/JSE industry groups.*

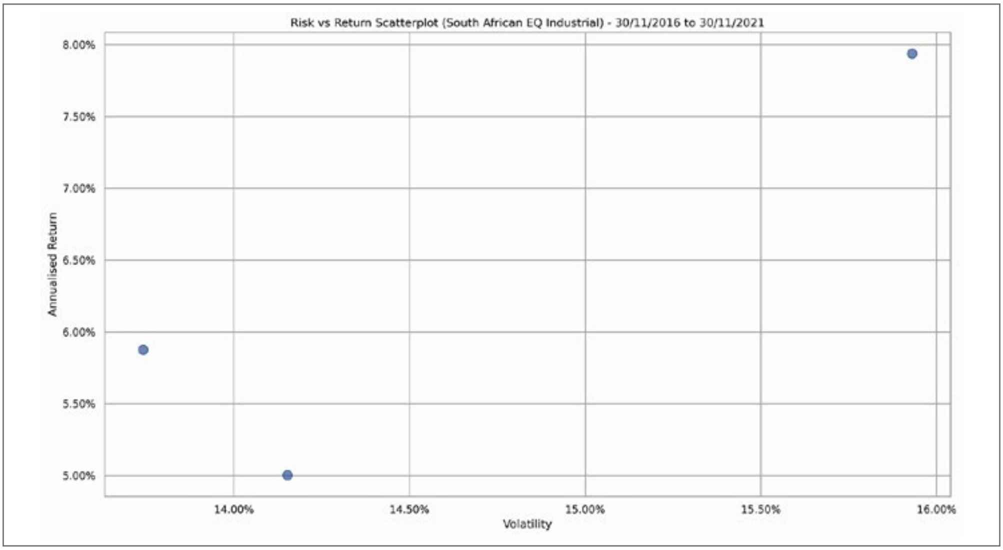


FIGURE 12 Risk vs return – equity – industrial

The Financial category also has fewer portfolios at only four, all of which have five years of history, and AuM of R1.1 billion. From Figure 13 we can see the annualised returns range from under 1% p.a. to under 4.5% p.a. (the worst sector performance over the past five years), and the volatility ranges from under 20% p.a. to just under 23.5% p.a., which is more risky than the Industrial category, but less than the Resources category on average. Perhaps this category and these portfolios should not exist either, despite the AuM.

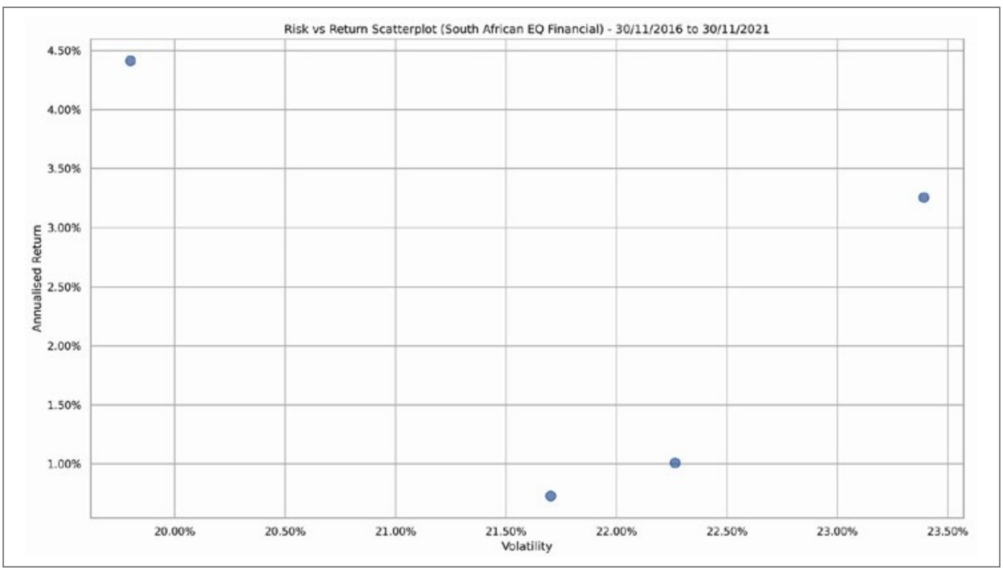


FIGURE 13 Risk vs return – equity – financial

4.1.2.7 SOUTH AFRICAN – EQUITY – UNCLASSIFIED

> *Equity – Unclassified portfolios – These portfolios invest in a single industry or sector or in companies that share a common theme or activity as defined in their respective mandates. Due to both the nature and focus of these portfolios, they may be more volatile than portfolios that are diversified across the broader market. The performance of these portfolios cannot be compared to others in this category. Should it be considered appropriate, where five or more portfolios focus on a particular theme a new category will be created and the portfolios transferred.*

There is a single portfolio in this category that although it has an annualised return and volatility that fits within the range of other equity portfolios, it is probably justified that it remains in the Unclassified category.

4.1.3 SOUTH AFRICAN – INTEREST BEARING

> *Interest Bearing Portfolios are collective investment portfolios that invest exclusively in bond, money market investments and other interest earning securities. These portfolios may not include equity securities, real estate securities or cumulative preference shares.*

So let us turn our attention to the next South African 'Second Tier' classification, Interest Bearing. We begin with the lowest risk category within this tier, and work our way up the risk spectrum.

4.1.3.1 SOUTH AFRICAN – INTEREST BEARING – MONEY MARKET

> *Interest Bearing – Money market portfolios – These portfolios seek to maximise interest income, preserve the portfolio's capital and provide immediate liquidity. This is achieved by investing in money market instruments with a maturity of less than thirteen months while the average duration of the underlying assets may not exceed 90 days and a weighted average legal maturity of 120 days. The portfolios are typically characterised as short-term, highly liquid vehicles.*

The Money Market category has 33 portfolios and AuM of R186 billion, of which 28 portfolios have five years of returns. Despite the apparent spread of the risk versus returns in Figure 14, careful attention should be given to the values along both axes to appreciate that these portfolios are relatively closely clustered together. The returns range from about 5.9% p.a. to just over 6.8% p.a., while the volatility ranges from about 0.38% p.a. to about 0.52% p.a., both fairly tight ranges.

You can see by the definition of the category that there are many constraints that impact risk directly, in this case both market risk and liquidity risk. This makes these portfolios very conservative, and they typically do not provide any negative returns, even when measured over single days, although it is important to know that they can. We raise this now to contrast these with the portfolios in the next category that can take on a lot more risk; yet we find many that do not, again raising the question about their classification.

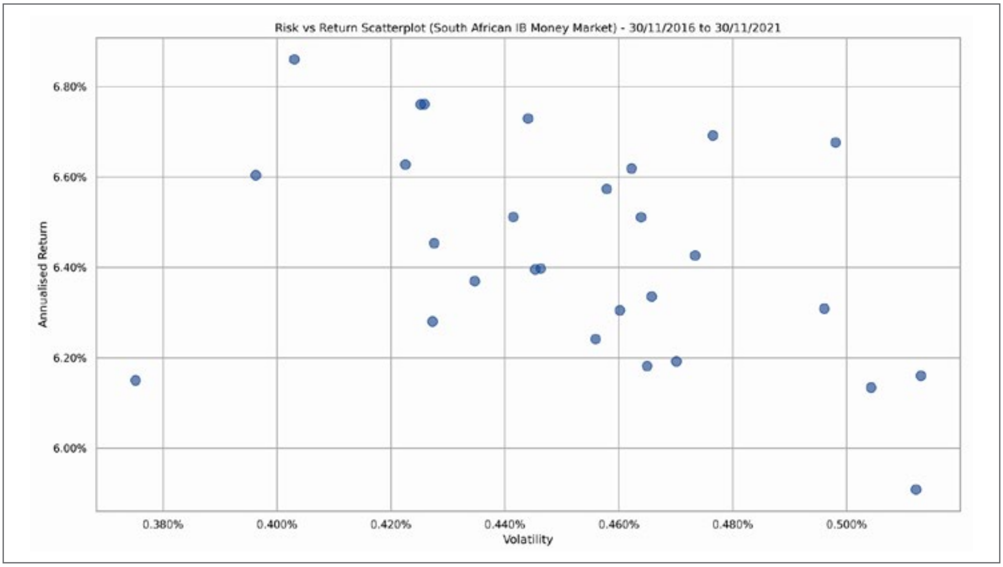


FIGURE 14 Risk vs return – interest bearing – money market

4.1.3.2 SOUTH AFRICAN – INTEREST BEARING – SHORT TERM

> *Interest Bearing – Short Term portfolios – These portfolios invest in bonds, fixed deposits and other interest earning securities which have a fixed maturity date and either have a predetermined cash flow profile or are linked to benchmark yields, but exclude any equity securities, real estate securities or cumulative preference shares. To provide relative capital stability, the weighted average modified duration of the underlying assets is limited to a maximum of two. These portfolios are less volatile and are characterised by a regular and high level of income.*

The Short Term category has 39 portfolios and AuM of R263 billion (making it into the top five categories by AuM), of which 27 portfolios have five years of returns. Although the spread of returns and volatility for this category is much wider (relatively speaking) than the Money Market category, the lower volatility and return values appear to overlap as can be seen in Figure 15. The annualised returns range from about 5.5% p.a. to just under 10% p.a., and the volatility ranges from about 0.45% p.a. to over 1.8% p.a.

So while certain portfolios in this category seem to be making the most of the risk limits imposed, others seem content to look more like Money Market portfolios. We find this behaviour somewhat bizarre as these portfolios do not get appreciated for their lower risk taking, given the category limits.

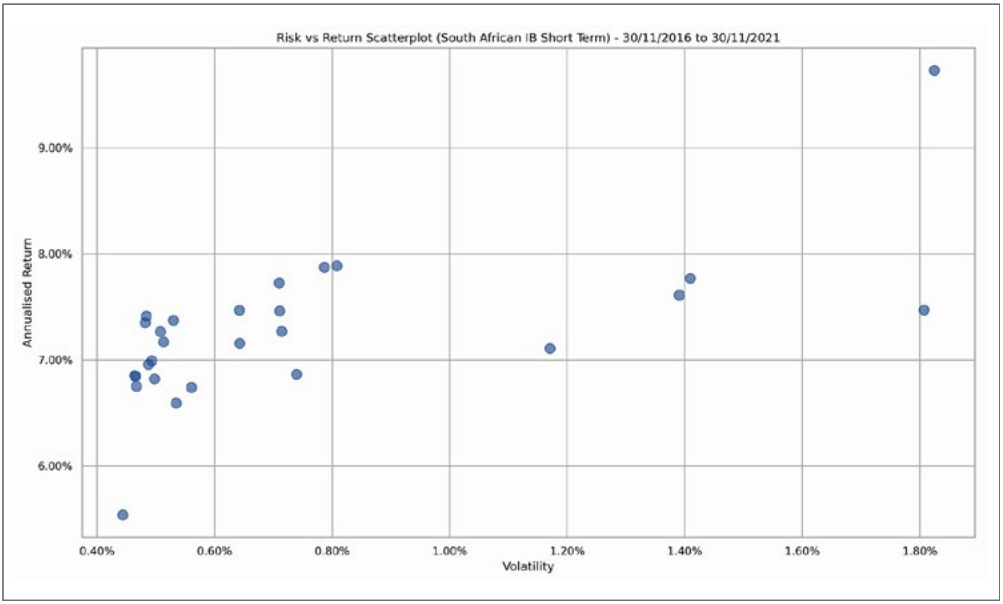


FIGURE 15 Risk vs return – interest bearing – short term

4.1.3.3 SOUTH AFRICAN – INTEREST BEARING – VARIABLE TERM

> *Interest Bearing – Variable Term portfolios – These portfolios invest in bonds, fixed deposits and other interest-bearing securities. These portfolios may invest in short, intermediate and long-dated securities. The composition of the underlying investments is actively managed and will change over time to reflect the manager's assessment of interest rate trends. These portfolios offer the potential for capital growth, together with a regular and high level of income. These portfolios may not include equity securities, real estate securities or cumulative preference shares.*

The Variable Term category has 39 portfolios and AuM of R84 billion, of which 24 portfolios have five years of returns. With the exception of a few portfolios that should be moved out of this category, it would actually represent a fairly homogeneous category, as can be seen in Figure 16.

Unfortunately, the inclusion of some inflation-linked bond portfolios, and some flexible duration portfolios (managed thus, not defined thus), throws out the range of returns and volatility. Including all the portfolios in this category, the annualised returns range from less than 1% p.a. to just under 10% p.a. (a range of 9% p.a.), and the volatility ranges from less than 2% p.a., to about 11% p.a. (also a range of 9% p.a.). Looking at just the portfolios managed to the All Bond Index, the annualised returns range from just under 8% p.a. to just under 10% p.a. (a range of 2% p.a.), and the volatility ranges from about 6.5% p.a. to less than 9% p.a. (a range of less than 2.5% p.a.). This is probably one of the best examples of classifications gone wrong.

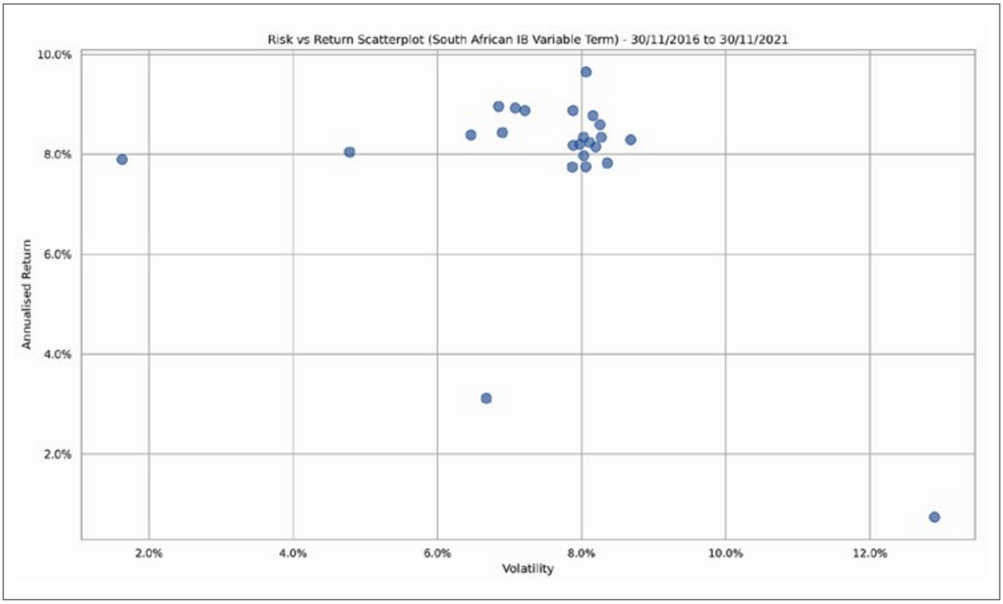


FIGURE 16 Risk vs return – interest bearing – variable term

4.1.4 SOUTH AFRICAN – REAL ESTATE – GENERAL

> *Real Estate – General portfolios – These portfolios invest in listed property shares, collective investment schemes in property and property loan stock and real estate investment trusts. The objective of these portfolios is to provide high levels of income and long-term capital appreciation. These portfolios invest at least 80% of the market value of the portfolio in shares listed in the FTSE/JSE Real Estate industry group or similar sector of an international stock exchange and may include other high yielding securities from time to time. Up to 10% of a portfolio may be invested in shares outside the defined sectors in companies that conduct similar business activities as those in the defined sectors.*

The Real Estate category has 37 portfolios and AuM of R33 billion, of which 28 portfolios have five years of returns. The risk versus return scatterplot in Figure 17, shows a range of annualised returns from –8.5% p.a. to 2% p.a. (probably the worst performing sector/

category/asset class over the past five years), and the volatility ranges from just under 18% p.a. to just under 30% p.a.

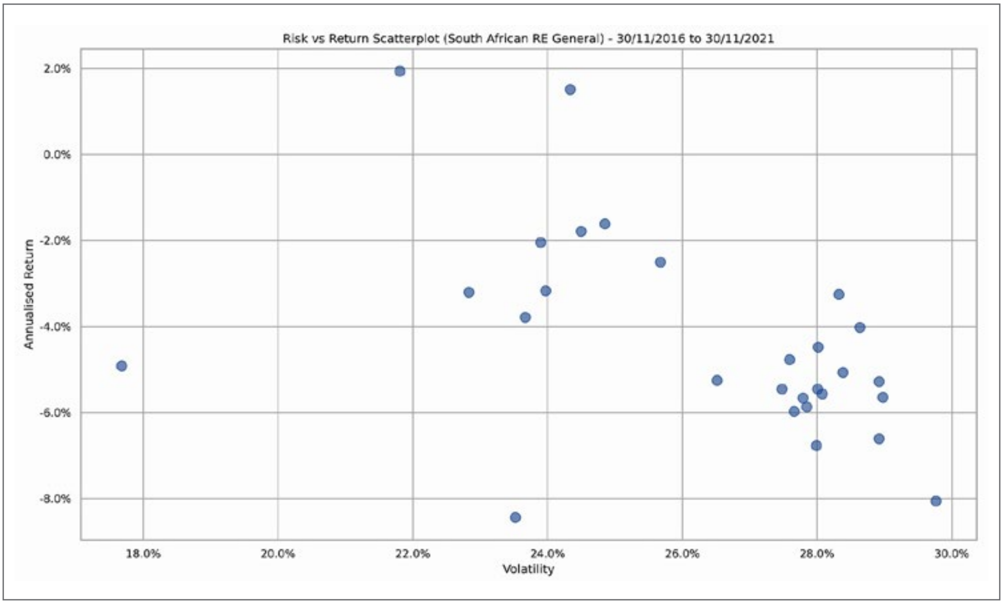


FIGURE 17 Risk vs return – real estate – general

The high average (and individual portfolio) volatility for this category is important, and has been alluded to before when discussing the overlap in certain of the Multi Asset categories. Property has had a horrible five years of returns, but previously enjoyed phenomenal returns outperforming every other asset class over the long term. We get back to the implications of this later in this paper, but please bear these numbers in mind.

4.2 Risk versus return scatterplots for multiple categories

4.2.1 SOUTH AFRICAN – MULTI ASSET

Figure 18 compares the High Equity and Medium Equity categories, to show the level of overlap on the same risk versus return chart. From this chart it should be clear that the degree of overlap is large, and the categories do not do a good job differentiating between portfolios along the risk spectrum or dimension. To be fair, the objective is for each category to be fairly homogeneous, not for categories to also not overlap, but this is somewhat implied, although we will concede not the same.

As highlighted above, part of the issue creating this overlap is that both categories allow total equity exposure between 0% and 60% (the overlap), the upper limit for the Medium Equity category. There seems to be little point in allowing high equity portfolios to not have a lower total equity exposure limit. Some managers may argue that they do in fact wish to use the 0% lower limit (or part thereof), when perhaps they think that all equities are too

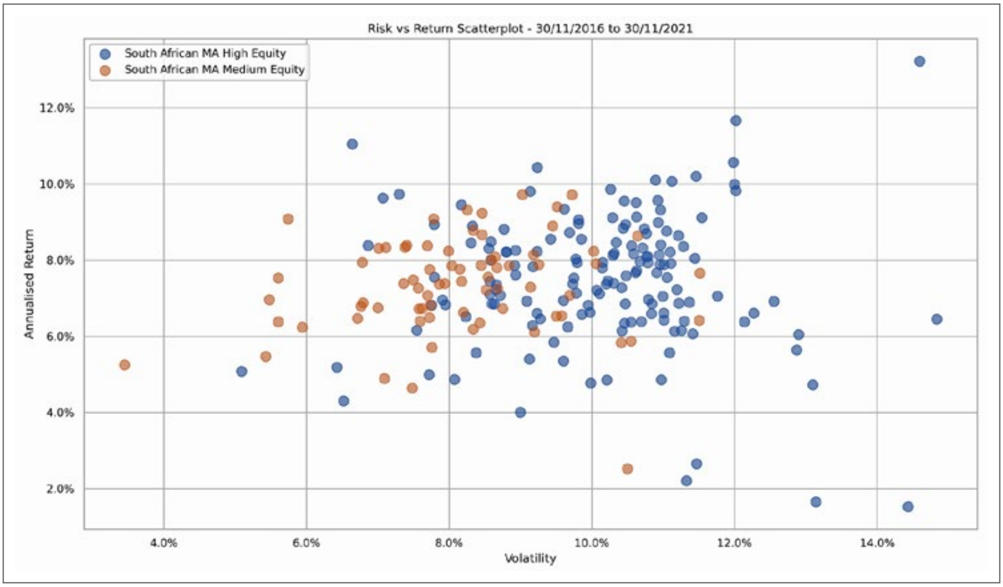


FIGURE 18 Risk vs return – multi asset – high, & medium equity

'expensive', and that you can't find enough 'cheap' equities to invest in. This argument is however flawed from two different perspective.

Firstly, this is a high equity category, implying high equity exposure, otherwise what is the point of the category. High equity exposure is therefore what investors should expect from portfolios in this category, not low equity exposure portfolios, even if only at certain points in time. Secondly, there is an alternative Multi Asset category, Flexible, which allows portfolios to have equity allocation ranging from 0% to 100%. These points are actually incidental anyway. The objective of having a category is to group similar portfolios, and allowing that large range in total equity exposure defeats this purpose at the definition level.

If the upper limit of the Flexible category represents a problem as portfolios with more than 75% total equity exposure are not Regulation 28 compliant (Regulation 28 of The Pension Funds Act, 1956), and therefore cannot be used as the total solution for members, then we can consider creating a flexible category with the 75% upper limit. This should however be called a Flexible category, and that level of flexibility should not be allowed in the High Equity category (from 0% to 75%). We do not explore that further here, as recommendations are covered in the next section. We would not support this anyway as it would still violate the objective of creating homogeneous groups of portfolios.

Figure 19 compares the Medium Equity and Low Equity categories. You will again notice a similar pattern to Figure 18 with a high degree of overlap between these two categories. What is often surprising is the overlap near the extremes. While we expect to have a level of overlap towards the centre of the chart, we are often surprised when we have overlap at the extremes. Here we mean the overlap between the same extremes of each category, as opposed to the overlap between the upper extreme of one category with the lower extreme of the other, which would be expected.

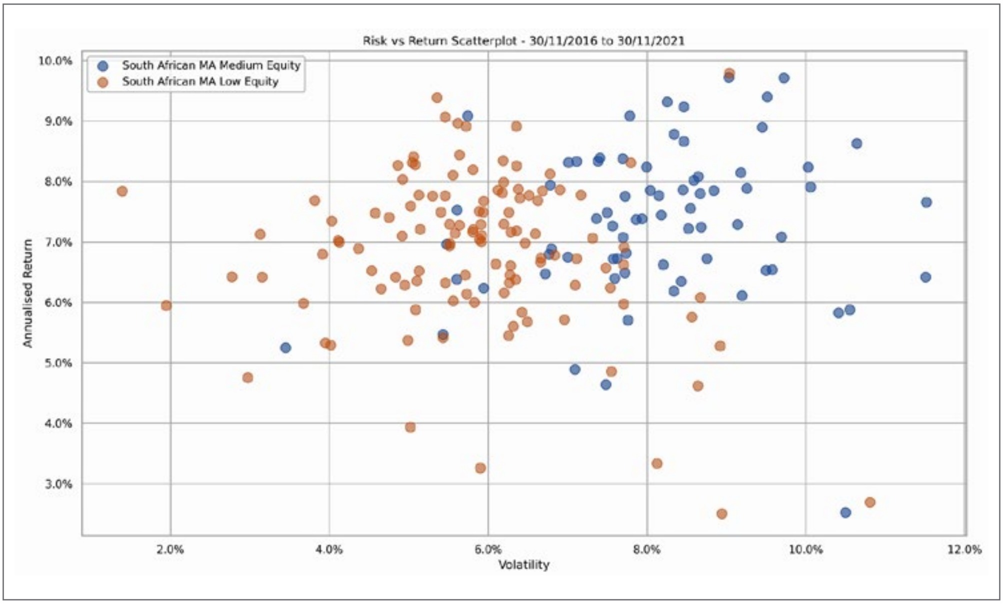


FIGURE 19 Risk vs return – multi asset – medium & low equity

This is especially true at the upper end of the chart on the volatility axis. We could expect to have more overlap at the lower end of the chart because both categories again have a lower limit of 0% total equity exposure, but the Low Equity category is limited to 40% total equity exposure, while the Medium Equity category has an upper limit of 60% (20% more, which is significant).

While this may not make sense intuitively, unless none of the portfolios in the Medium Equity category are utilising the full equity exposure limit, we explore what is perhaps going on here, that may explain at least some of this.

Figure 20 compares these first three categories, namely High Equity, Medium Equity, and Low Equity, in a single chart. You will now be able to see the overlap between even the two extreme categories, High Equity and Low Equity. Separating out these three categories on the basis of volatility alone would be a very difficult exercise. We therefore need to ask about the validity of the classification standard that produces these results.

Unless you want to argue that volatility is not important, which you can certainly do, you would have a problem choosing portfolios by beginning first with the category, and then choosing the portfolio. You would be better served by combining all of these portfolios and deciding on the range of volatility that would be appropriate for your client's risk profile, and then selecting from portfolios within that range.

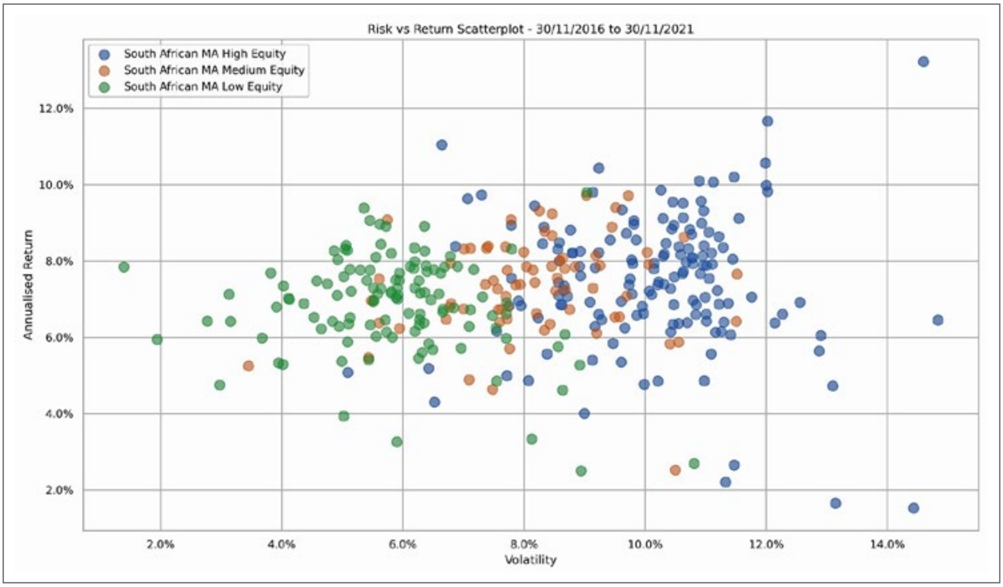


FIGURE 20 Risk vs return – multi asset – high, medium & low equity

Of course, we are deliberately ignoring the fact that portfolios should not be selected on this basis, but past performance (especially risk) is an important consideration to apply along with many other selection criteria. We have previously highlighted that volatility itself is a random variable, subject to variability, so you wouldn't be able to begin with a single volatility number or range anyway, without a fuller appreciation of how the volatility over this specific time period compares to other time periods.

Figure 21 compares just the Low Equity and Income categories, again showing some overlap between these two categories. Specifically note the overlap of the extremes of these two categories: in the Income category, one portfolio has more volatility (7% p.a.) than most of the Low Equity category portfolios, and one Low Equity category portfolio has less volatility (1.5% p.a.) than most of the Income category portfolios.

This may be very surprising as noted above, especially because of the upper limit on total equity exposure for the Income category. It may not be surprising at all, also as noted above, for the Low Equity category as it has no lower limit on total equity exposure. It does however continue to make the case that these categories are not very homogeneous.

Figure 22 adds the Income category to the three previously compared categories (High, Medium, and Low Equity), and the entire picture becomes a lot clearer having all of these categories on a single chart. You may start seeing a pattern that if some limits are changed and some introduced into these categories, we may find the distinction between them becoming a little clearer, which is really the objective of having these categories and the classification standard.

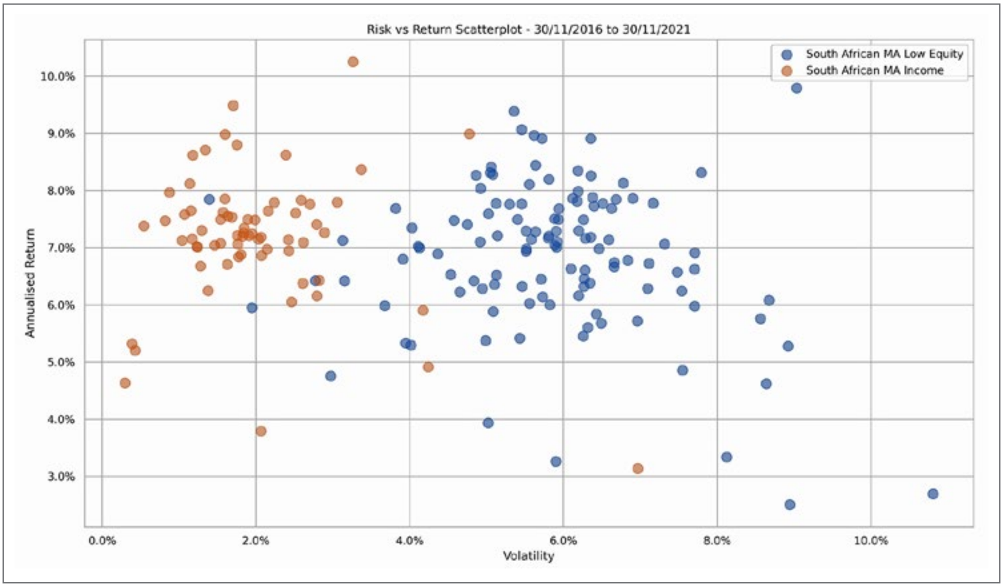


FIGURE 21 Risk vs return – multi asset – low equity & income

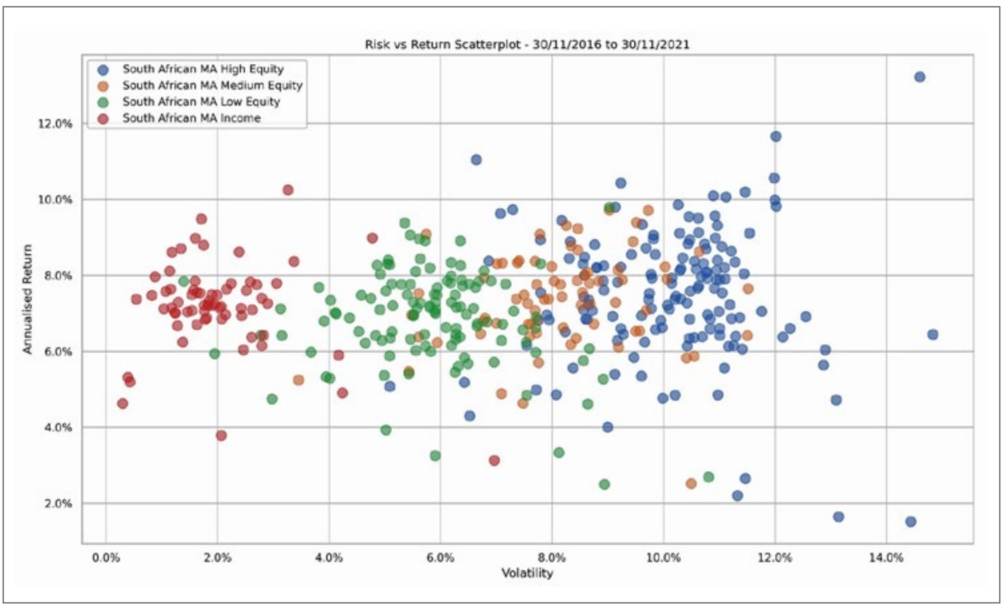


FIGURE 22 Risk vs return – multi asset – high, medium, low equity & income

Figure 23 adds the Flexible category to the rest of the Multi Asset categories considered above, but excludes the Target Date category. While some of these new category portfolios disappear into the mix of the other four categories, others lie distinctly to the right, representing higher risk and alluding to higher total equity exposure.

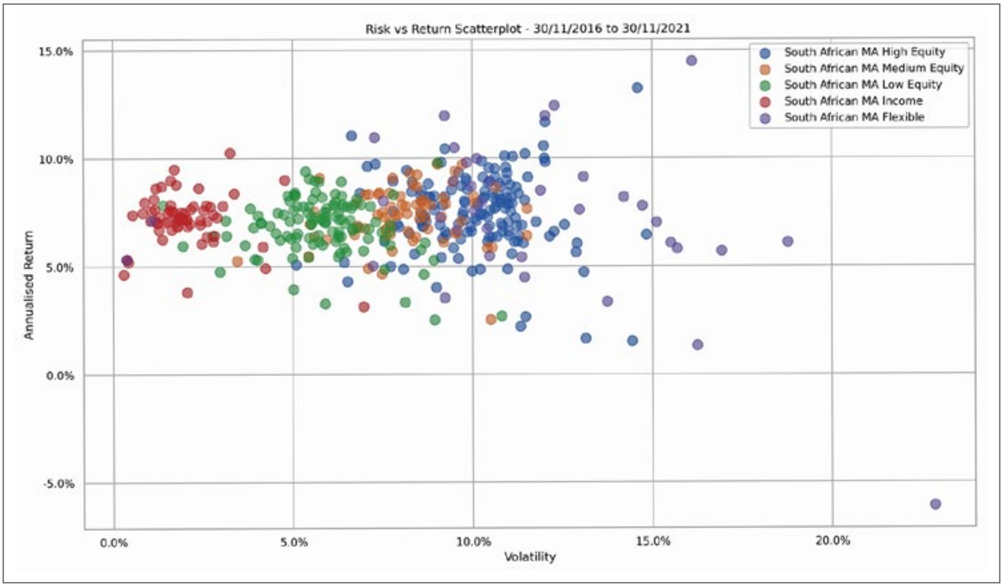


FIGURE 23 Risk vs return – multi asset – all categories (excl. target date)

Finally Figure 24 adds the Equity General category to the Multi Asset tier. Doing this is important for two reasons touched on above. First, note the overlap between this category and both the High Equity and the Flexible categories. Second, this implies that the Equity category can be considered in many ways as an extension of the Multi Asset category.

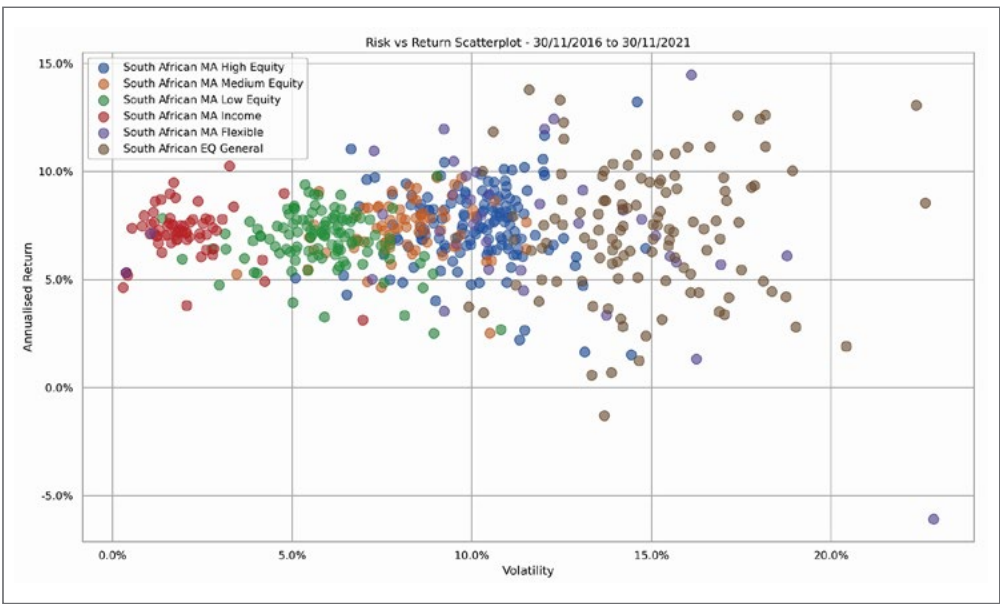


FIGURE 24 Risk vs return – multi asset – all categories (excl. target date) & equity – general

4.2.2 SOUTH AFRICAN – EQUITY

Figure 25 shows the combined risk versus return scatterplot for the entire Equity tier. As previously highlighted, the standout is clearly the Resources category with much higher volatility and returns.

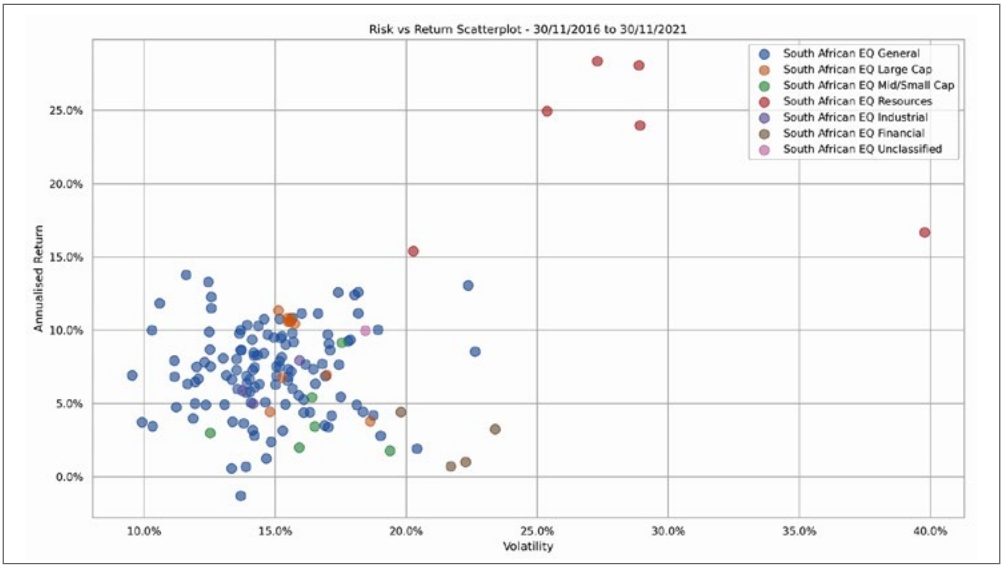


FIGURE 25 Risk vs return – equity – all categories

4.2.3 SOUTH AFRICAN – INTEREST BEARING

Finally, Figure 26 shows the combined risk versus return scatterplot for the entire Interest Bearing tier, which includes just three categories. In this case the categories are generally pretty well defined and the standouts are individual portfolios that should be moved out of the Variable Term category.

4.3 Hierarchical clustering and dendrograms for individual categories

Instead of reducing 60 monthly returns into two single metrics (volatility and return), we could instead use each month as a dimension, independent of every other month, and compare how portfolios deliver returns one month (dimension) at a time, and then compare these returns across all 60 dimensions at once.

4.3.1 SOUTH AFRICAN – MULTI ASSET – HIGH EQUITY

If you think about the similarity/difference between two portfolios as being the Euclidean distance between the monthly returns of the two portfolios, you can create a distance matrix which can be used by an agglomerative algorithm to create a hierarchical cluster or tree, which can be visualised using a dendrogram, shown in Figure 27 for the portfolios in the High Equity category.

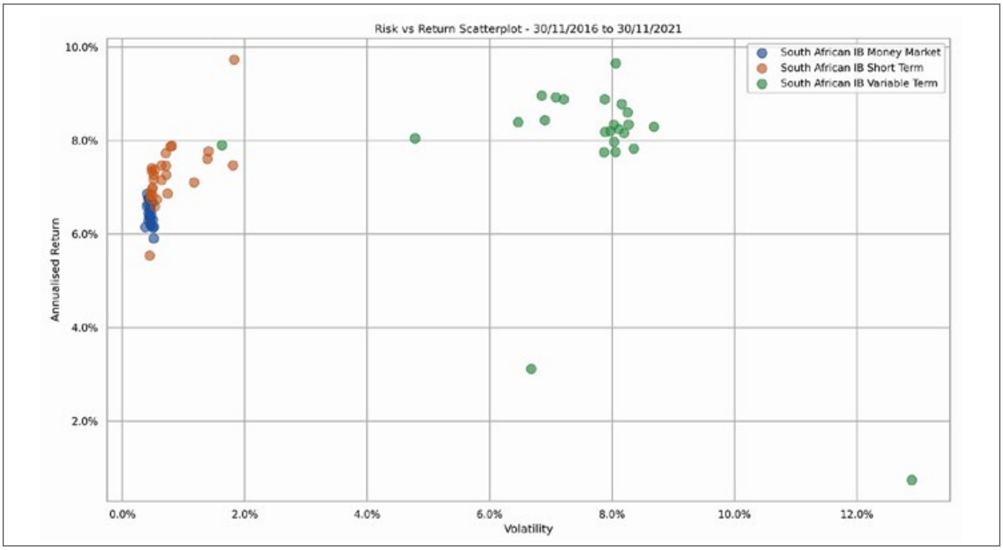


FIGURE 26 Risk vs return – interest bearing – all categories

Starting at the left of the chart, every single portfolio forms its own cluster (is a leaf of the dendrogram), and looking to the right of the chart, every portfolio belongs to the same cluster. Depending on where you want to draw a vertical line between these two extremes, you end up with a number of clusters, counted as the number of times the vertical line crosses the horizontal lines of the dendrogram. You can also then establish the portfolios that make up each of those clusters by following the tree all the way down (towards the left) from each of those lines.

For example, you will notice that the dendrogram has created three different coloured clusters (cyan, red, and green). The red cluster in the middle is the smallest and contains only nine portfolios, and although it joins with the cyan cluster, which includes many more portfolios, it is still obviously very different as they join a long way to the right. The green cluster then joins with the combination of the cyan and red clusters to form the single 'super' cluster. The green cluster can also be subdivided into two distinct clusters. The smaller one consisting of just eight portfolios is just below the red cluster, while the other cluster below that is much larger.

For those who are not too familiar with the structure and meaning of dendrograms, perhaps some explanation would be helpful. The distance from the left at which two portfolios, two clusters, or a portfolio and a cluster join, represents how similar or different they are. It is actually the Euclidean distance between the two portfolios in n-dimensional return space, in this case n equals 60.

If they join very close to the left, they are very similar (close distance in 60 dimensional space), if they join further to the right, they are less similar, or more different. Never read the dendrogram vertically, as two points adjacent to each other may not be similar at all,

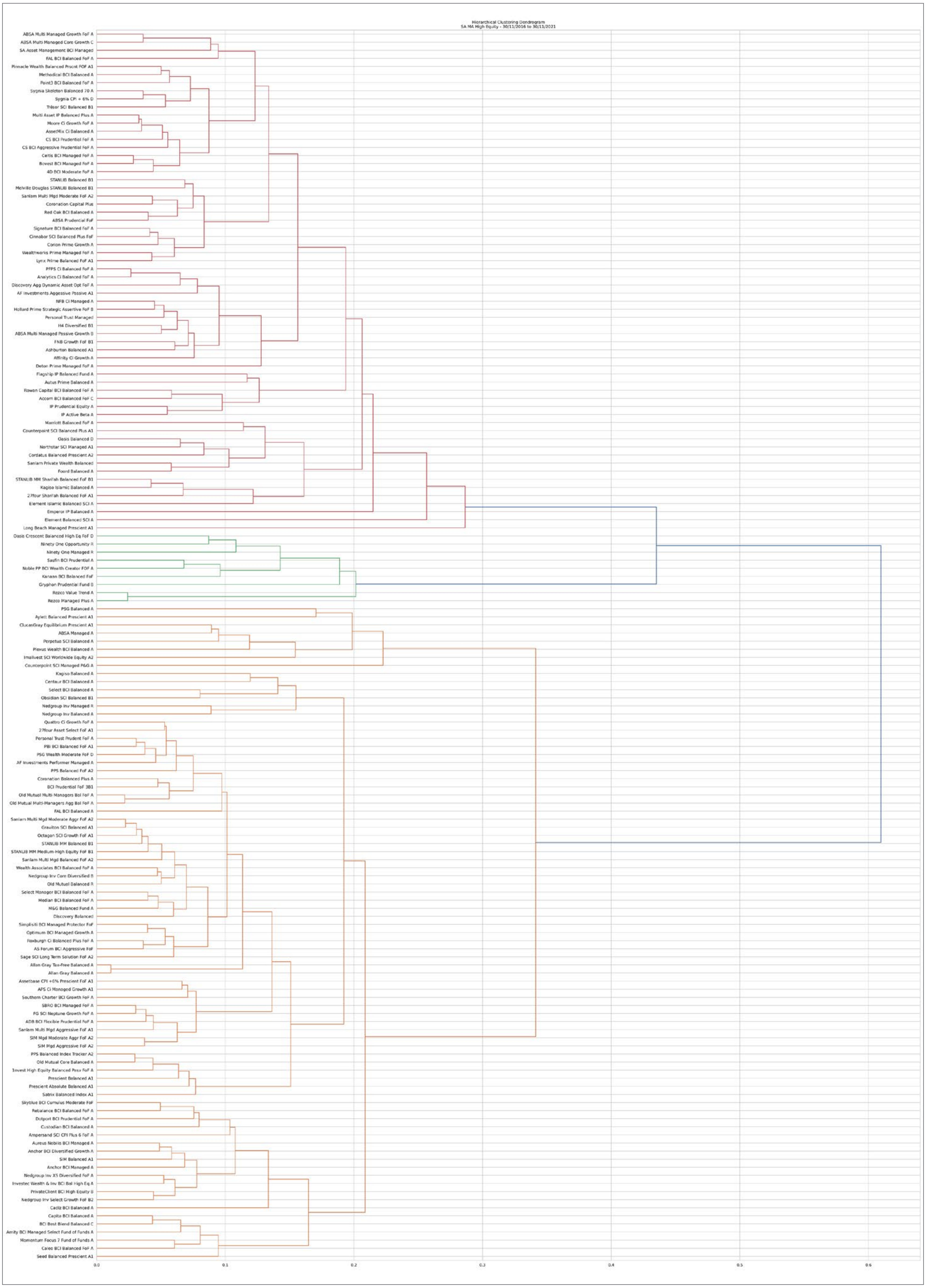


FIGURE 27 Dendrogram – multi asset – high equity

and there is no way of telling except by looking from left to right, and this is only useful if they join directly.

Let us look at an example to clarify these points. Rezco Managed Plus and PSG Balanced lie next to each other in the dendrogram, but their relationship is completely unknown except to say that they are not similar at all because they sit in two completely different clusters. Nothing can be said about how Rezco Managed Plus compares to other portfolios in the red cluster (if you ignore Rezco Value Trend which it is clearly similar to), except that they all sit in the same cluster and are therefore more similar to each other, than to portfolios in other clusters (cyan and green).

Although applying this technique within a single category is of limited use, as we can still get a sense for the structure of portfolios in terms of how similar or different they are, its true power will become more evident when we combine categories, as this may address the questions of overlap between categories found when looking at risk and return. We explore this below.

This technique may also be useful in discovering whether some portfolios are outliers, and perhaps do not belong in a category, perhaps belong in another category, or perhaps do not belong anywhere and would best be described as Unclassified. We explore this in more detail in the next section.

The red cluster is perhaps a good candidate for this as it is both small (outliers by definition should be few in number), and substantially different from other clusters, although it joins with the cyan cluster in this case, whereas it would typically join with the main cluster to which every other portfolio belongs.

It also helps to know the underlying portfolios and managers really well to draw high level inferences from the dendrograms, including some of the reasoning as to why certain portfolios may be part of the same cluster, and why others may not be. This is a drawback of the technique in that it requires a certain level of knowledge of the portfolios to make sense of the clustering. The power of the technique is however undeniable. Armed with nothing but monthly returns you can uncover meaningful structure with the data that represents both actual and practical knowledge of the portfolios being analysed. This power cannot be overstated as should be evidenced by the two Rezco portfolios being placed so closely together in the dendrogram using nothing but returns.

4.3.2 SOUTH AFRICAN – MULTI ASSET – MEDIUM EQUITY

Figure 28 shows the dendrogram for the Medium Equity category. You again have a smaller cluster of potential outliers, in this case the green cluster with 11 portfolios that joins last to form the single cluster.

Judging by our knowledge of some of the portfolios in this cluster, it probably represents the lower risk portfolios in this category, or at least those with lower equity exposure or using derivatives and structures. This cluster also appears to have two sub-clusters which seem to be somewhat different.

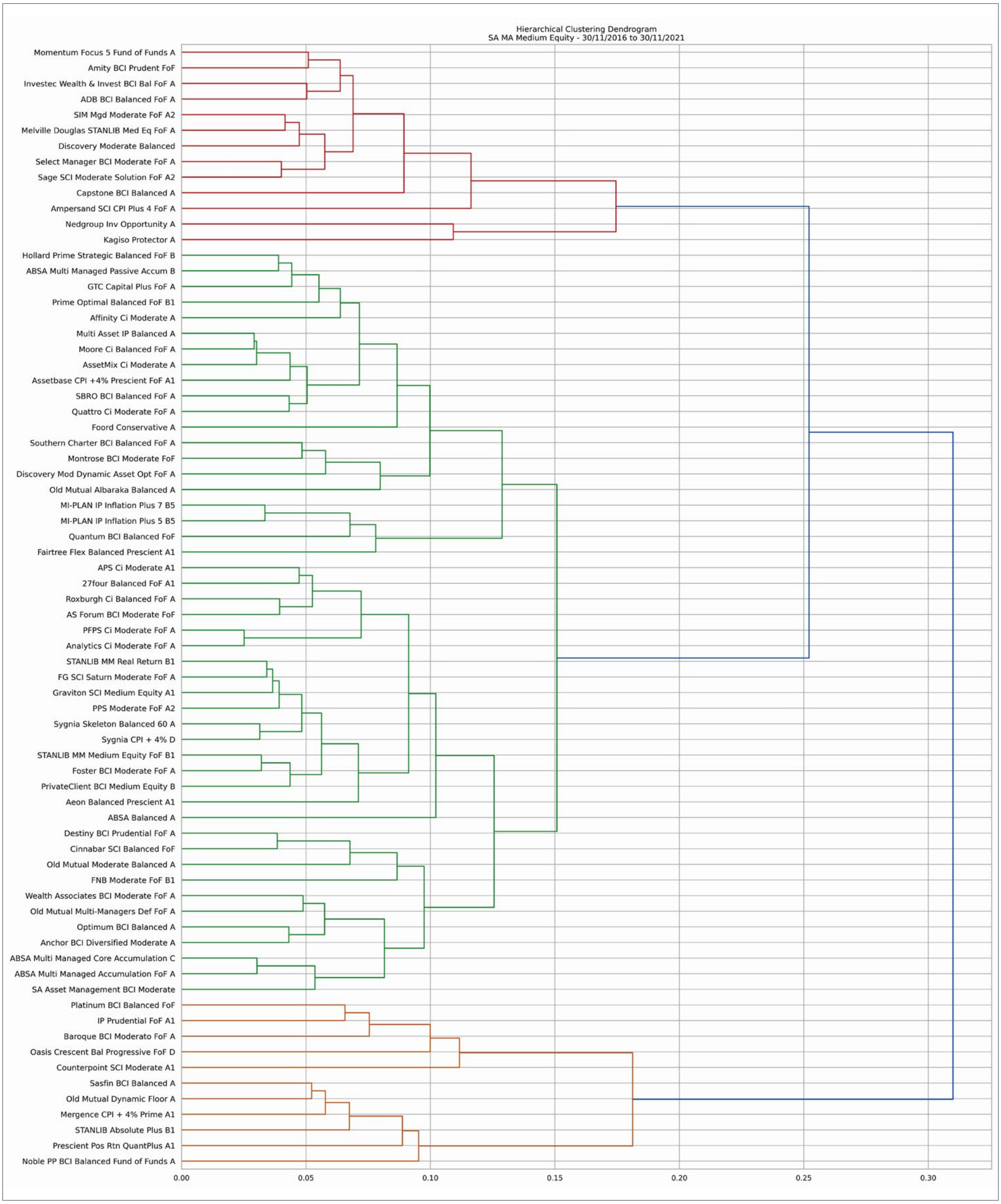


Figure 28 Dendrogram – multi asset – medium equity

The large red cluster probably reflects the typical Medium Equity portfolio, and these portfolios would probably remain in the same cluster when other categories are combined with it (say Low Equity, or High Equity, or both).

4.3.3 SOUTH AFRICAN – MULTI ASSET – LOW EQUITY

Figure 29 shows the dendrogram for the Low Equity category. In this instance, the green cluster and the red cluster are about the same size, by number of portfolios. It is therefore difficult to say what the difference is between them, but the red cluster joins the cyan cluster, which is much smaller, before they all join together in the single cluster. We would need to dig deeper into portfolios in each of these clusters to understand their differences.

As discussed under the PCA section, one way to get insights into what may make these clusters similar or different is to include some markers that are very well defined and may represent a dimension of importance along which portfolios in the category are invested. Sometimes these can be very useful, and sometimes they may be so different from the portfolios as to provide no insights or help at all.

We could include a number of asset class indices into the analysis alongside the portfolios e.g. indices for local equities, property, nominal bonds, inflation-linked bonds, cash, and global equities, bonds and cash. We could even break some of those indices up into sub-indices e.g. sector indices for local equities, or duration buckets for nominal bonds. We could also include the category average to see which cluster would include it.

We could also look at the average asset allocation (if available) for each of the clusters, and see whether any numbers stand out as being significantly different. Ideally, these averages should be over the same time period as the returns being considered.

4.3.4 SOUTH AFRICAN – MULTI ASSET – INCOME

Figure 30 shows the dendrogram for the Income category. This dendrogram shows some truly fascinating results, and is one of the reasons that they can be so useful. Although you still have three coloured clusters, one cluster has a single solitary portfolio.

As the name of this portfolio implies, this is an inflation-linked bond portfolio, and it is perhaps surprising that it has chosen this category instead of the Interest Bearing Variable Term category, like some of the other inflation-linked bond portfolios. Perhaps this is intentional, because the portfolio plans on investing in other asset classes allowed by the category, or perhaps it is because it felt that the Variable Term category was not appropriate as we have previously argued.

The red cluster appears to have a sub cluster of five portfolios that seem to be doing something completely differently to the rest of the cluster. Some have Flexible in their name, while the others may be predominantly bond portfolios. The green cluster appears fairly well defined, although it is clearly very different from all of the other portfolios that eventually join together before joining the green cluster at the furthest distance. These portfolios are clearly therefore very different and we would need to dig a little deeper to

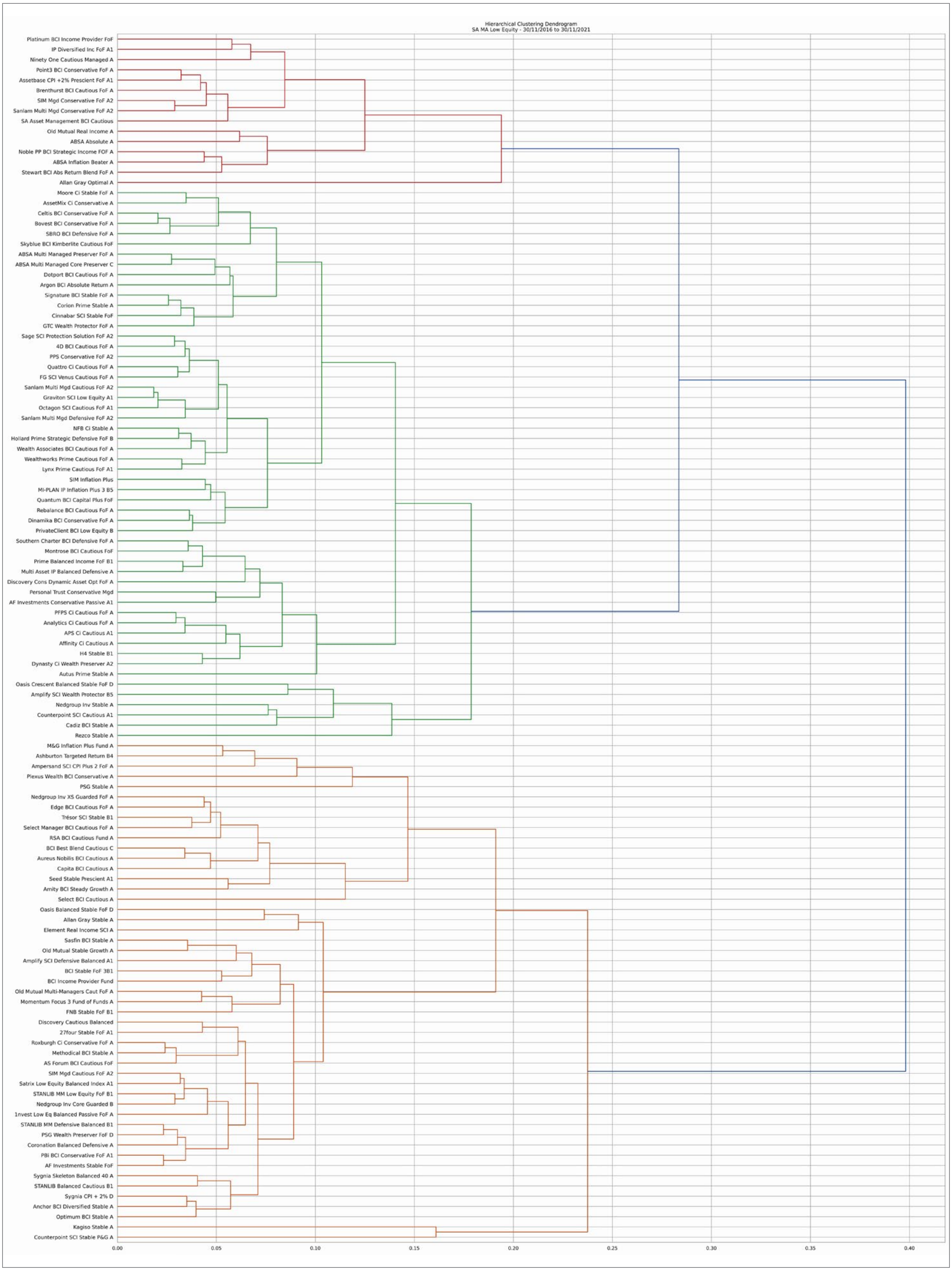


Figure 29 Dendrogram – multi asset – low equity

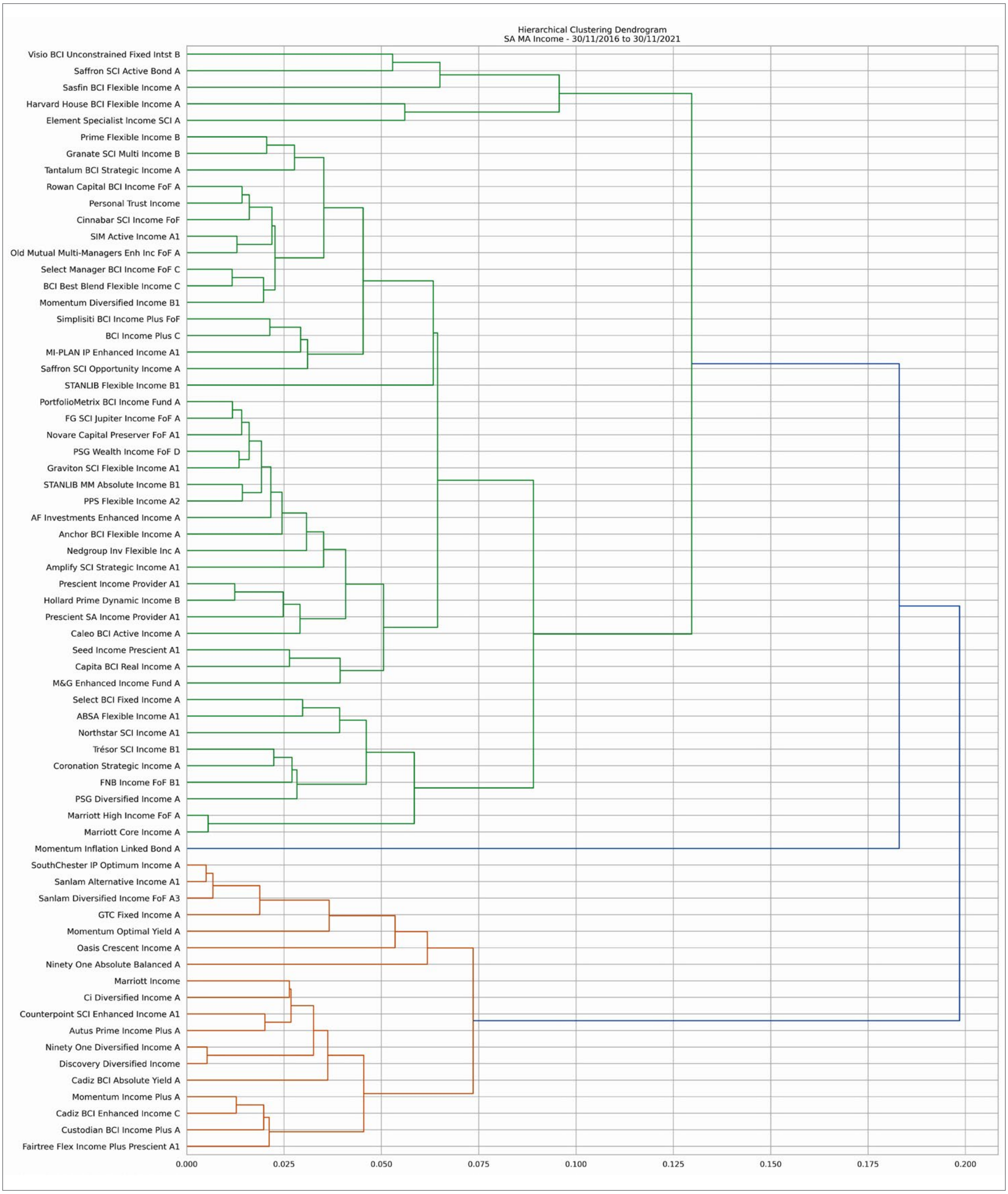


Figure 30 Dendrogram – multi asset – income

uncover the reasons for these differences, perhaps using some of the techniques discussed above.

4.3.5 SOUTH AFRICAN – MULTI ASSET – FLEXIBLE

Figure 31 shows the dendrogram for the Flexible category. For this category, you would expect to see some bigger differences as this category is so heterogeneous as we have seen from the risk return scatterplots. You can certainly see some clustering of property-heavy portfolios, and separately the clustering of more income-type portfolios. You can also see the clustering of the high equity portfolios, if you know which they are.

4.3.6 SOUTH AFRICAN – EQUITY – GENERAL

Figure 32 shows the dendrogram for the General category. Here the clustering could be along many different lines as the underlying philosophies and processes of managers

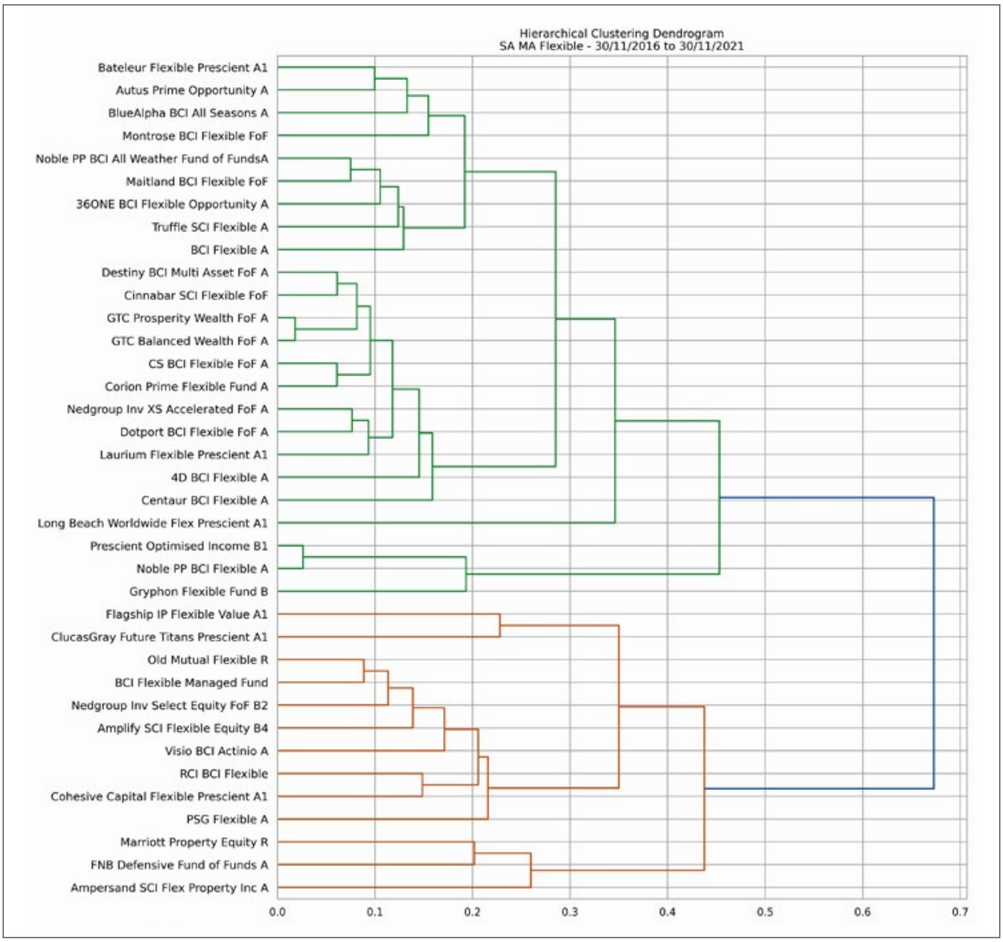


FIGURE 31 Dendrogram – multi asset – flexible

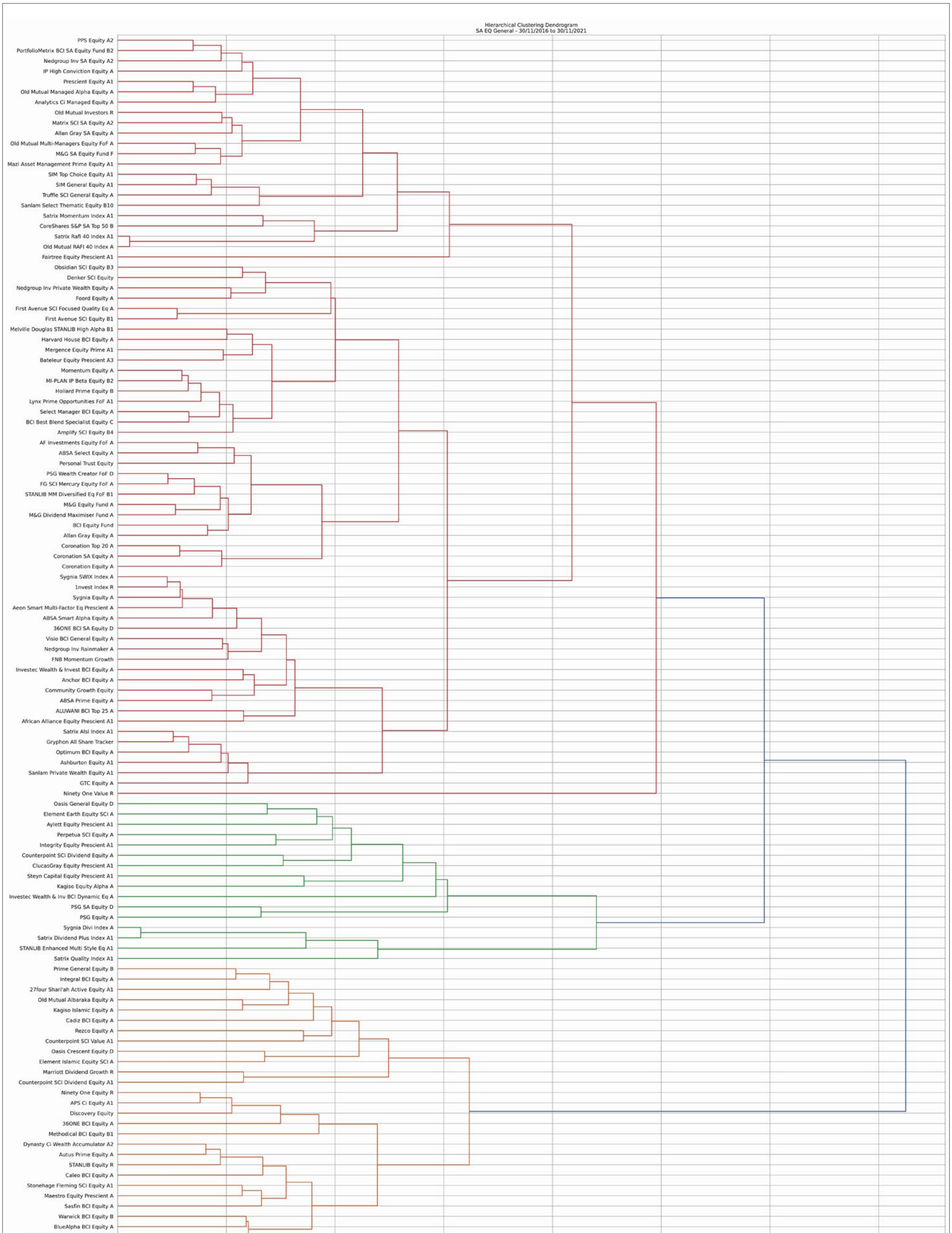


Figure 32 Dendrogram – equity – general

can vary widely. You can certainly pick up on some style clusters, as well as some sector clusters (specifically resources heavy). This perhaps becomes even more apparent when we combine all of the Equity categories as they include sectors.

4.3.7 SOUTH AFRICAN – EQUITY – LARGE CAP

Figure 33 shows the dendrogram for the Large Cap category. As previously suspected this seems to be made up of two primary clusters. The largest (green) is made up of all the index tracking portfolios, specifically the market cap weighted Top 40 index. The smaller cluster (red) are the active portfolios attempting something different. We refer to the Satrix Equally Weighted Top 40 Index portfolio as active, despite it tracking an index, because the weighting methodology chosen is an active decision (as are all index tracking decisions, which is why we refer to these portfolios as index trackers, and not passive portfolios).

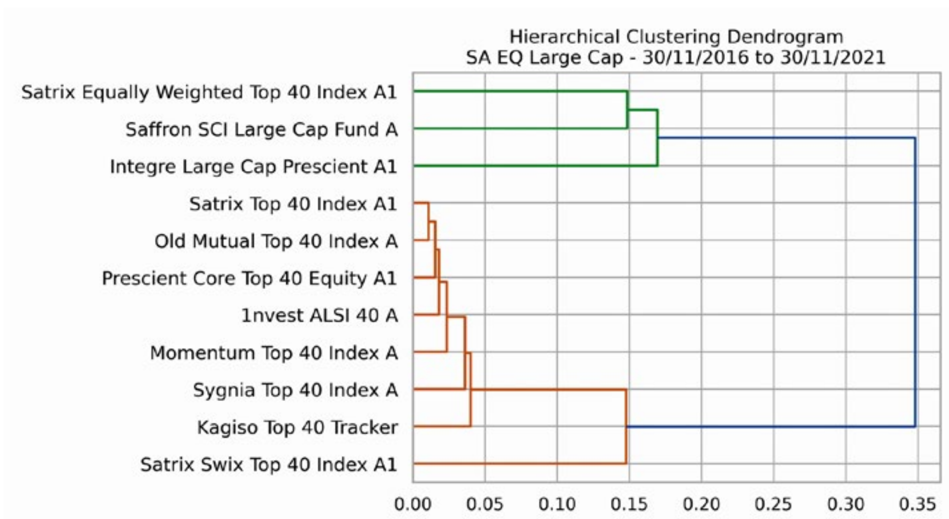


FIGURE 33 Dendrogram – equity – large cap

4.3.8 SOUTH AFRICAN – EQUITY – ALL OTHER CATEGORIES

The rest of the category dendrograms are presented in Figures 34, 35, 36, and 37, without any further commentary.

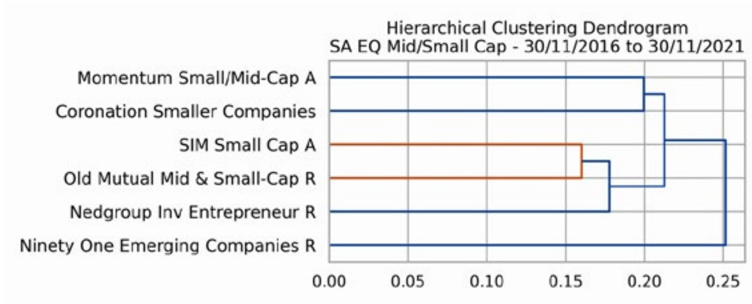


Figure 34 Dendrogram – equity – mid / small cap

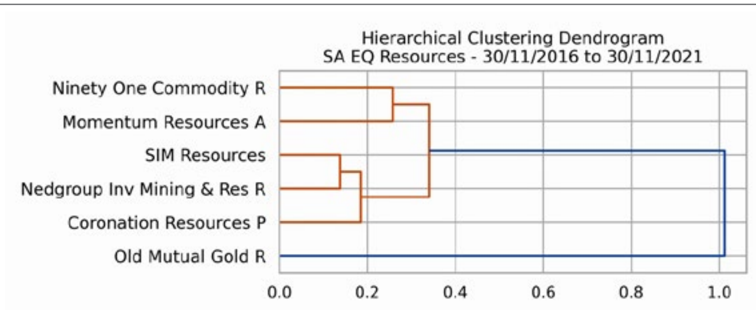


Figure 35 Dendrogram – equity – resources

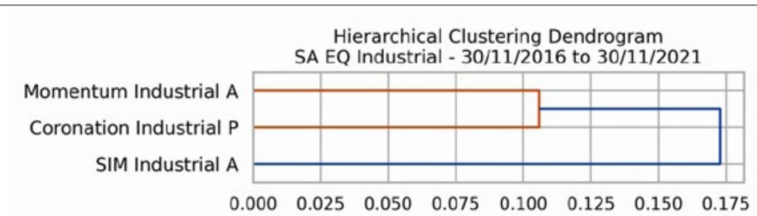


Figure 36 Dendrogram – equity – industrial

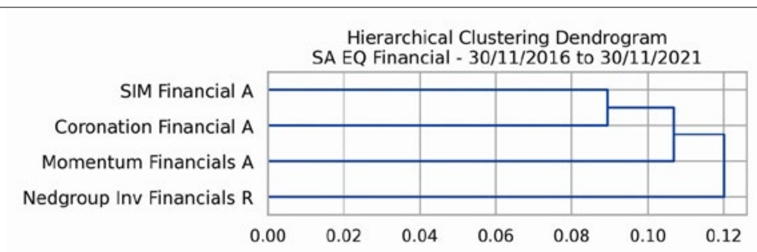


Figure 37 Dendrogram – equity – financial

4.3.9 SOUTH AFRICAN – INTEREST BEARING – MONEY MARKET

Figure 38 shows the dendrogram for the Money Market category.

Despite the category being fairly homogeneous, you can still see the clear distinction between clusters. Note however how much smaller the range on the x-axis is versus the equity category.

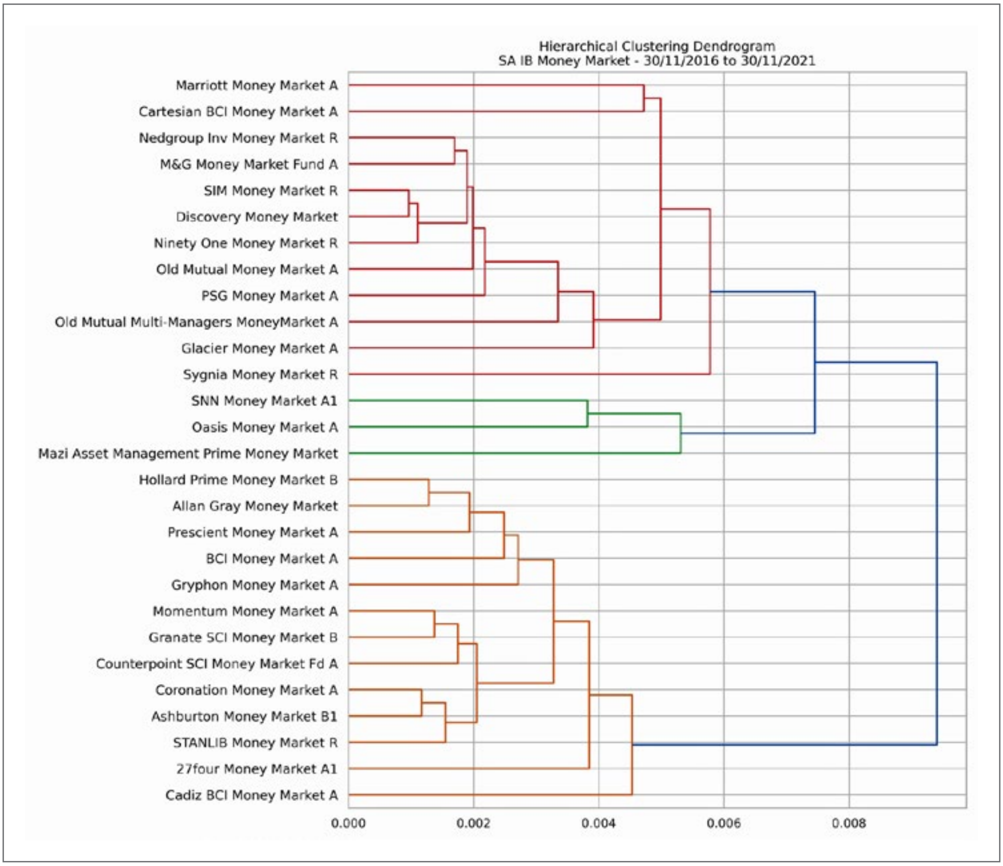


Figure 38 Dendrogram – interest bearing – money market

4.3.10 SOUTH AFRICAN – INTEREST BEARING – SHORT TERM

Figure 39 shows the dendrogram for the Short Term category. We had previously noted that some portfolios in this category were very conservatively managed, somewhat like Money Market portfolios, and you can see those in the second (bottom) of the two red clusters. The first (top) of the red clusters is managed a little less conservatively, and the last two clusters (blue and green) are managed most aggressively.

It is interesting to see the results when the Money Market and the Short Term categories are combined, how the clusters form and which portfolios in the Short Term category fall into the same cluster as the Money Market portfolios. We shall see this below.

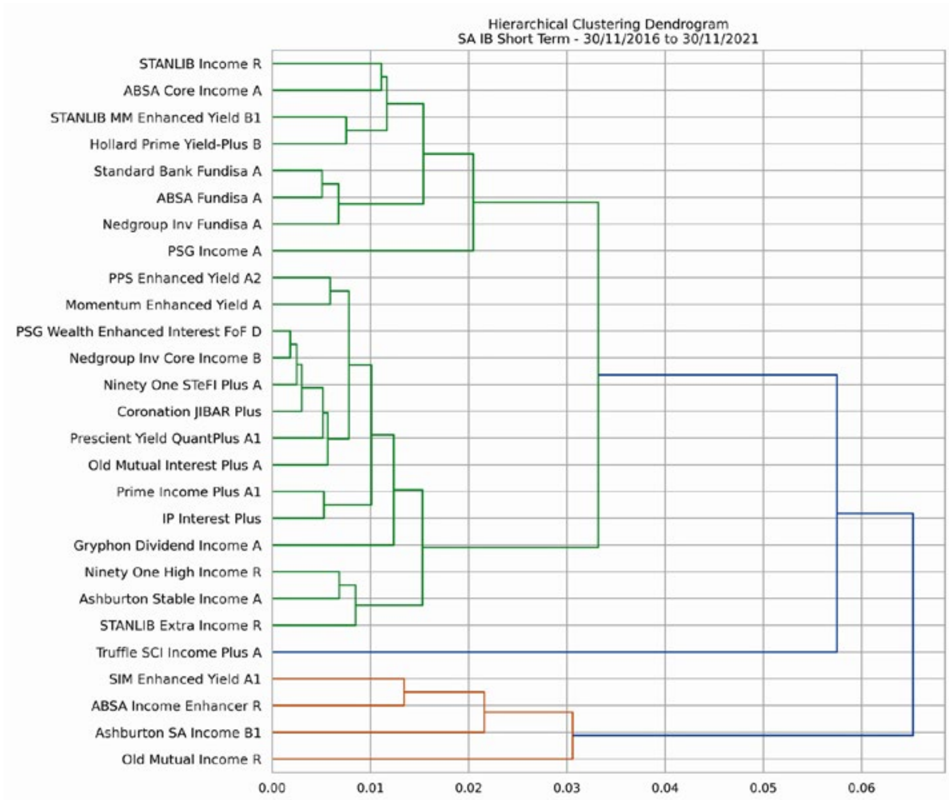


FIGURE 39 Dendrogram – interest bearing – short term

4.3.11 SOUTH AFRICAN – INTEREST BEARING – VARIABLE TERM

Figure 40 shows the dendrogram for the Variable Term category, and we should expect to see some distinct clustering here based on the work done previously. The dendrogram does not disappoint, and again provides a wonderful example of the technique in uncovering the structure of the underlying portfolios.

At the bottom of the dendrogram we have the Colourfield BCI Income portfolio that is clearly very different to all others. Colourfield is known to manage Liability Driven Investment (LDI) solutions, and this portfolio could represent just how distinct that strategy is to a typical bond portfolio managed to the All Bond Index (ALBI). In the red cluster, we have the inflation-linked bond portfolios, along with the flexible or short-duration bond portfolios, again very distinct from the very long duration of bond portfolios managed against the ALBI.

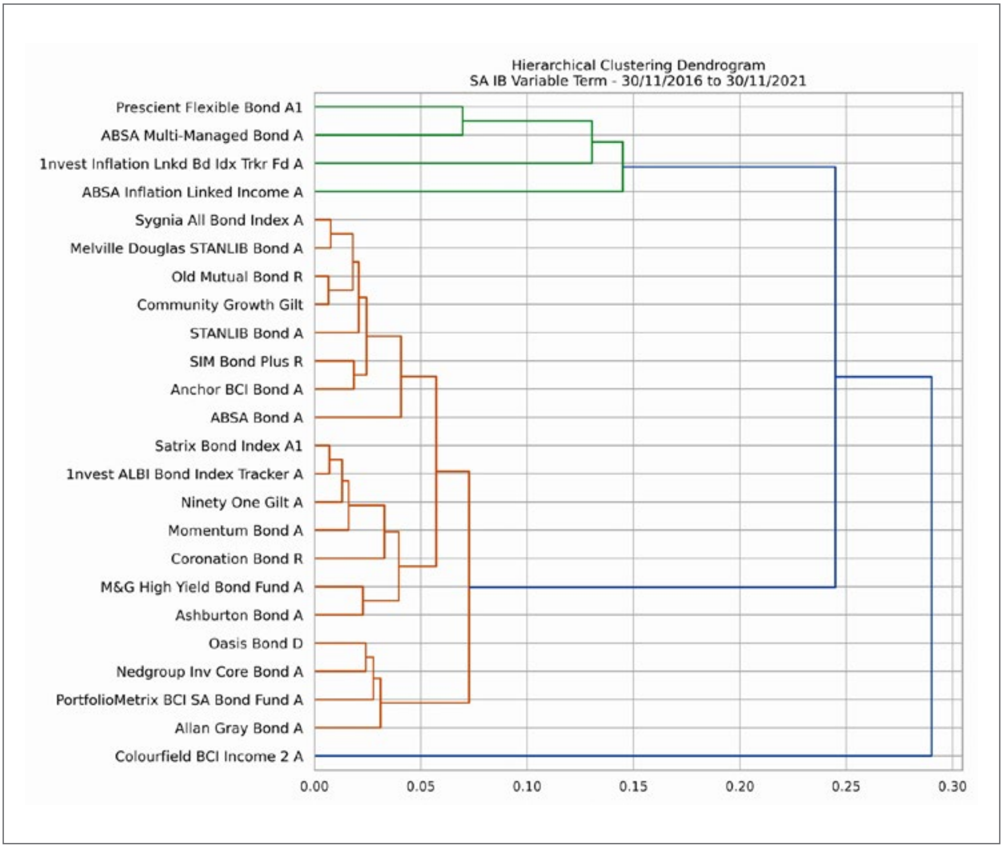


FIGURE 40 Dendrogram – interest bearing – variable term

4.3.12 SOUTH AFRICAN – REAL ESTATE – GENERAL

Figure 41 shows the dendrogram for the Real Estate category, and here we have two clusters with two portfolios each, that appear to be distinct. Even the two portfolios in the green cluster appear to be fairly different from each other. The cyan cluster also appears to have two moderately distinct clusters.

4.4 Hierarchical clustering and dendrograms for multiple categories

To get a better appreciation for the clustering of all the portfolios within a tier, we can apply the clustering techniques developed in the previous section to a combination of categories within the tier.

4.4.1 SOUTH AFRICAN – MULTI ASSET

We begin with the Multi Asset tier – because of its very large size in terms of the number of portfolios, and the complexity involved in visualising such large data sets, we begin by combing just two categories at a time. We hope that building our knowledge in this

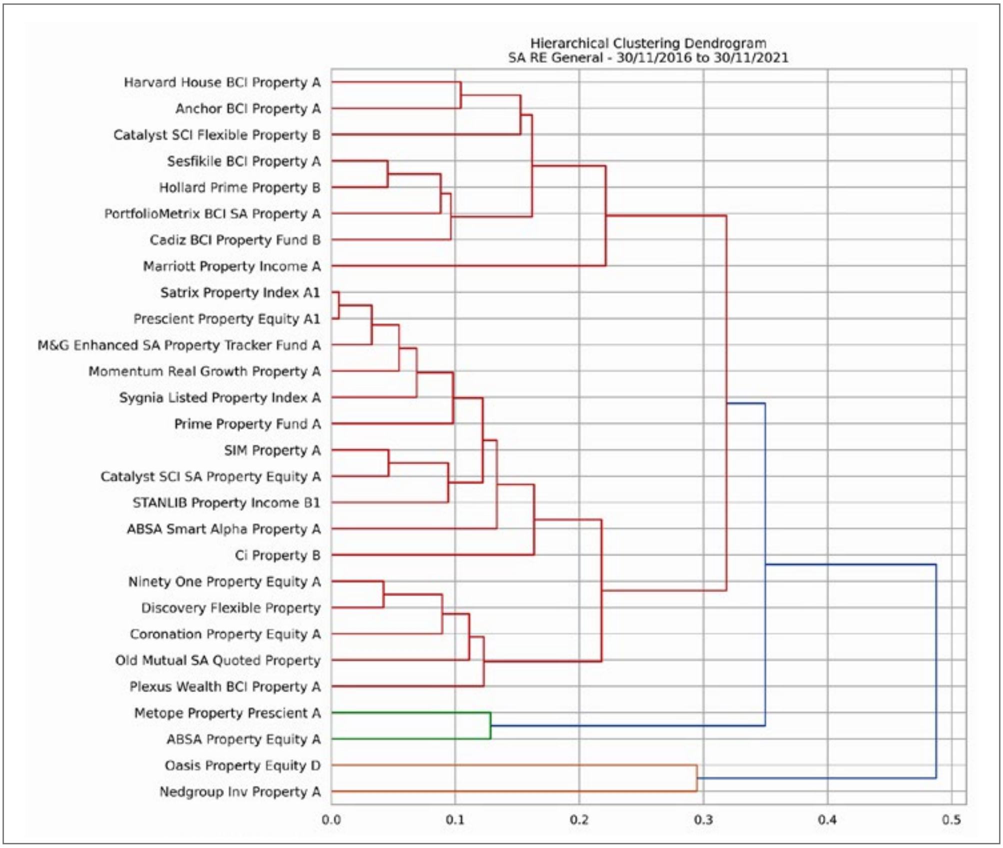


Figure 41 Dendrogram – real estate – general

way will make the comprehension easier when we are looking at the final structure of the portfolios, and the categories that represent them.

Figure 42 shows the dendrogram for the High Equity and Medium Equity categories combined. If we had no knowledge of the portfolios in these two categories, except the belief that the categories were well-defined homogeneous groups of portfolios, we may expect the two primary clusters to reflect the two categories, one with all of the High Equity portfolios, and the other with all the Medium equity portfolios.

However, having looked at this data before, we have noted the level of overlap in risk and return, and the definitions of these categories, so we may expect the two primary clusters to reflect something a little different. We may, for example, expect the two primary clusters to reflect the level of risk within the collection of portfolios, clustering higher risk portfolios together and lower risk portfolios together. We already know that certain High Equity portfolios have fairly low volatility (risk), even if we haven't unpacked why this is the case, and similarly, we know that certain Medium Equity portfolios have higher risk, again without necessarily knowing why.

The primary clusters seem to reflect exactly this, with the green cluster being dominated by High Equity portfolios, interspersed with some Medium Equity portfolios, and the red cluster dominated by Medium Equity portfolios, although it has many High Equity portfolios within it as well.

It is perhaps a good time to mention another limitation of this technique. While these dendrograms are very powerful and useful, the amount of data and information they contain can become overwhelming. In this dendrogram we are considering 224 portfolios, and how those all relate to each other.

Looking at this data through the lens of many smaller clusters would give us insights into the similarity of individual portfolios. This could be important for many different exercises including getting an understanding of how portfolios are managed by understanding how they behave in relation to other portfolios.

Zooming out to the level of two or three clusters, however, is more likely to provide us with some high level structure of the overall population being considered. Here, we are focused on the latter as we are trying to understand how homogeneous or heterogeneous the population and the categories to which they belong are, and how they relate to each other.

Figure 43 shows the dendrogram for the Medium Equity and Low Equity categories combined. We again note some results similar to what we observed above. Although the two primary clusters represent the two categories under consideration, each is interspersed with portfolios from the other category.

In this case however the one cluster (green) is by far dominated by Low Equity portfolios, whereas the other (red) is well represented by both categories. You can however see the two sub-clusters within this red cluster that seem to do a much better job of separating the Medium Equity portfolios from the Low Equity portfolios.

Figure 44 shows the dendrogram for the Low Equity and Income categories combined. This seems like the cleanest dendrogram seen so far, with the red cluster consisting of Low Equity portfolios only (not a single Income portfolio). Although the green cluster is predominantly made up of Income category portfolios, it does contain some Low Equity category portfolios, potentially highlighting that some of these may be more conservatively managed.

Figure 45 shows the dendrogram for the High Equity, Medium Equity and Low Equity categories combined. Although we have made the page larger to make the dendrogram more legible, it may still be a little difficult to discern the patterns in this data which is now really large with 342 portfolios combined.

You should however be able to see that the red cluster is dominated by High Equity portfolios, and the green cluster by Low Equity portfolios. You will however still see some High Equity portfolios in the Low Equity cluster and, vice versa, some Low Equity portfolios in the High Equity Cluster. The Medium Equity portfolios are interspersed throughout both clusters, but you can sometimes pick up sub-clusters that appear to have groups of Medium Equity category portfolios.

The most useful result from analysing such a large dendrogram is being able to see the portfolios that are potentially out of place, and to then be able to do further analysis to figure out whether these portfolios are possibly misclassified.

Figure 46 shows the dendrogram for the High Equity, Medium Equity, Low Equity, and Income categories combined. As per above, we have to begin talking more about the sub-clusters here, as the two primary clusters are no longer homogeneous with so many portfolios – we are now up to 409 portfolios.

If we consider the first/top cluster (red) we can see that this has two very distinct sub-clusters, one with all of the Income portfolios, and very few Low Equity portfolios, and the other with the Low Equity portfolios, although this cluster includes both Medium Equity and High Equity portfolios.

The bottom/second cluster (green) is also sub-divided into many different clusters, but mainly holds all of the High Equity portfolios. The Medium Equity portfolios are mainly interspersed throughout this cluster, although they also tend to group together around much smaller sub-clusters within yet other sub clusters.

The best way to appreciate this complexity and the finer detail is to print this entire dendrogram on a big piece of paper and to analyse it at a very high level for the broad structure, before considering the various branches of the tree to see what portfolios each of the clusters contain. There is unfortunately just too much information here for us to dig into the detail on everything.

Figure 47 shows the dendrogram including the Flexible category. Figure 48 shows the dendrogram for all the Multi Asset categories. Figure 49 shows the dendrogram for all the Multi Asset categories, along with the Equity General category.

We leave the analysis of these to the reader, but the results should not be too surprising, with a continuation of the overlap of the portfolios in many of the categories supporting our contention that the categories could perhaps be better defined to make them more homogeneous, and these dendrograms provide the supporting evidence. Please take some time to go through these in detail as they contain useful information.

4.4.2 SOUTH AFRICAN – EQUITY – ALL CATEGORIES

Figure 50 shows the dendrogram for the combined Equity tier. As previously suspected, the Resources category portfolios form their own cluster (green) and are very different from the rest. The primary cluster (cyan) holds most of the General category portfolios, along with the portfolios from the Large Cap category, and the Industrial category, although in both cases these are part of smaller clusters. A smaller third cluster (red) holds portfolios from the Mid/Small Cap category, and the Financial category, along with portfolios from the General category, although the portfolios in the former two categories also form smaller clusters and group together.

4.4.3 SOUTH AFRICAN – INTEREST BEARING – ALL CATEGORIES

Figure 51 shows the dendrogram for the combined Interest Bearing tier. Again no major surprises here with the bottom cluster (green) containing all of the Money Market and Short Term category portfolios, and being very distinct to the Variable Term category. One portfolio from the Variable Term category does make it into this cluster which confirms our previous analysis and suspicions.

The structure of the Variable Term category looks very similar to the structure of the dendrogram when this category was considered in isolation. It not only has much more variability than the other cluster, but is so different to the other cluster (green) that the structure of this green cluster is actually difficult to see. It is perhaps therefore useful to look at just the Money Market and Short Term categories combined, so that we can better appreciate how these portfolios compare.

Figure 52 shows the dendrogram for the Money Market and Short Term categories combined, making it much easier to see the structure of these portfolios. Clearly the top two clusters (red and blue) are doing something completely differently to all the other portfolios. The green cluster is however very interesting because it can be seen to include two sub-clusters that are also quite different, with the top cluster consisting of just portfolios from the Short Term category. The bottom cluster however contains portfolios from both categories, suggesting that many of the portfolios in the Short Term category are being managed much more conservatively, and are sometimes perhaps indistinguishable from money market portfolios.

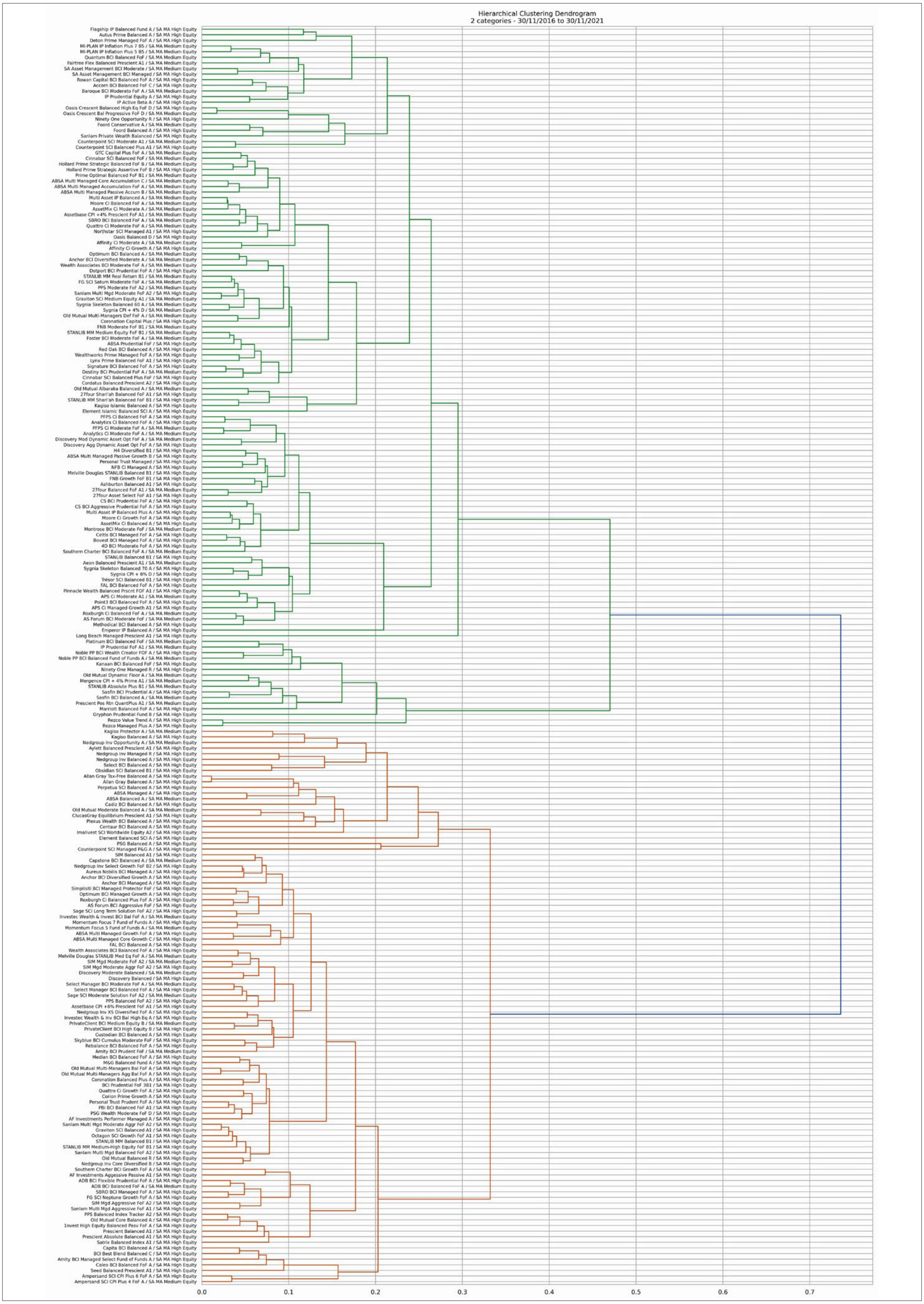


FIGURE 42 Dendrogram – multi asset – high, & medium equity

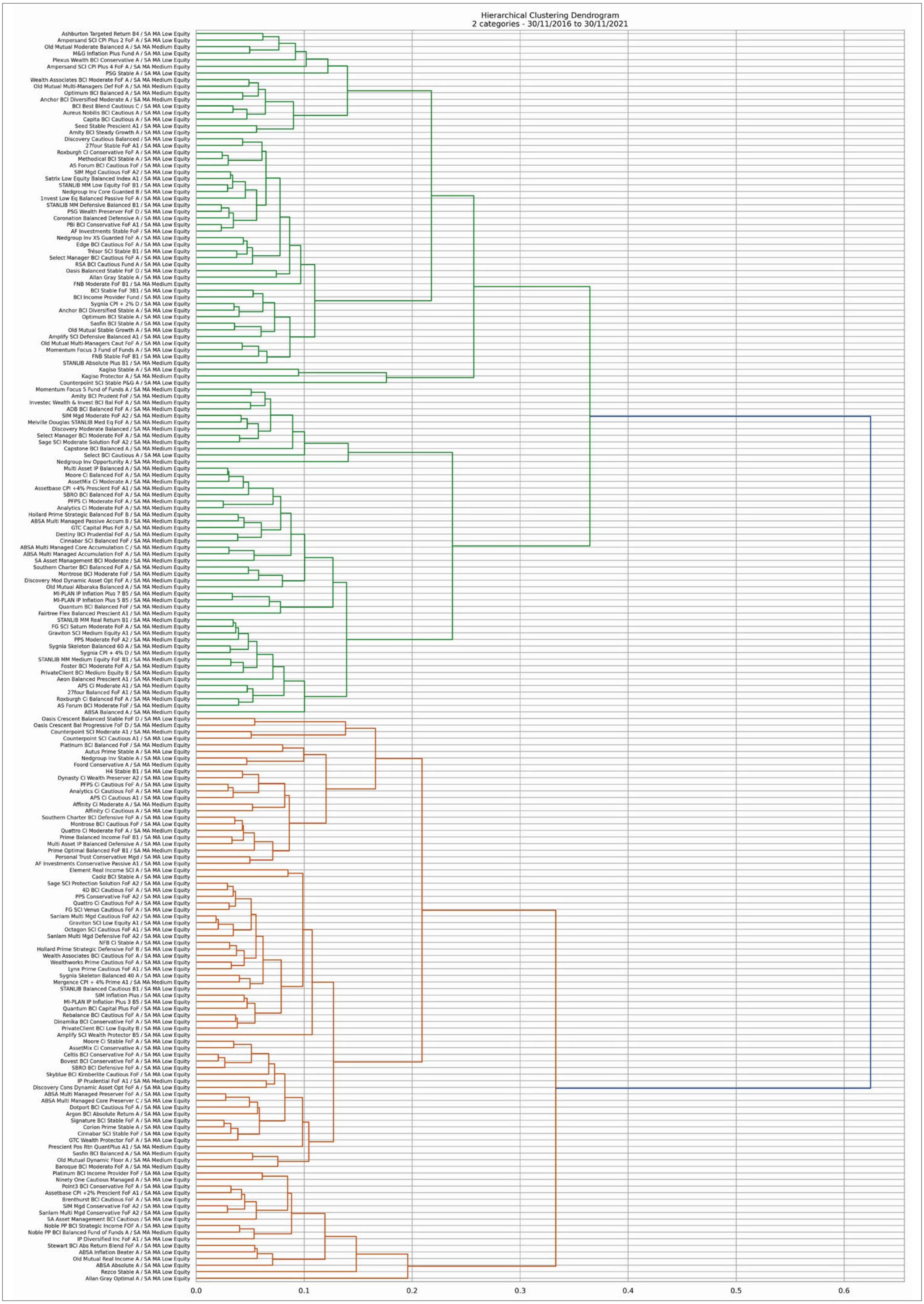


FIGURE 43 Dendrogram – multi asset – medium & low equity

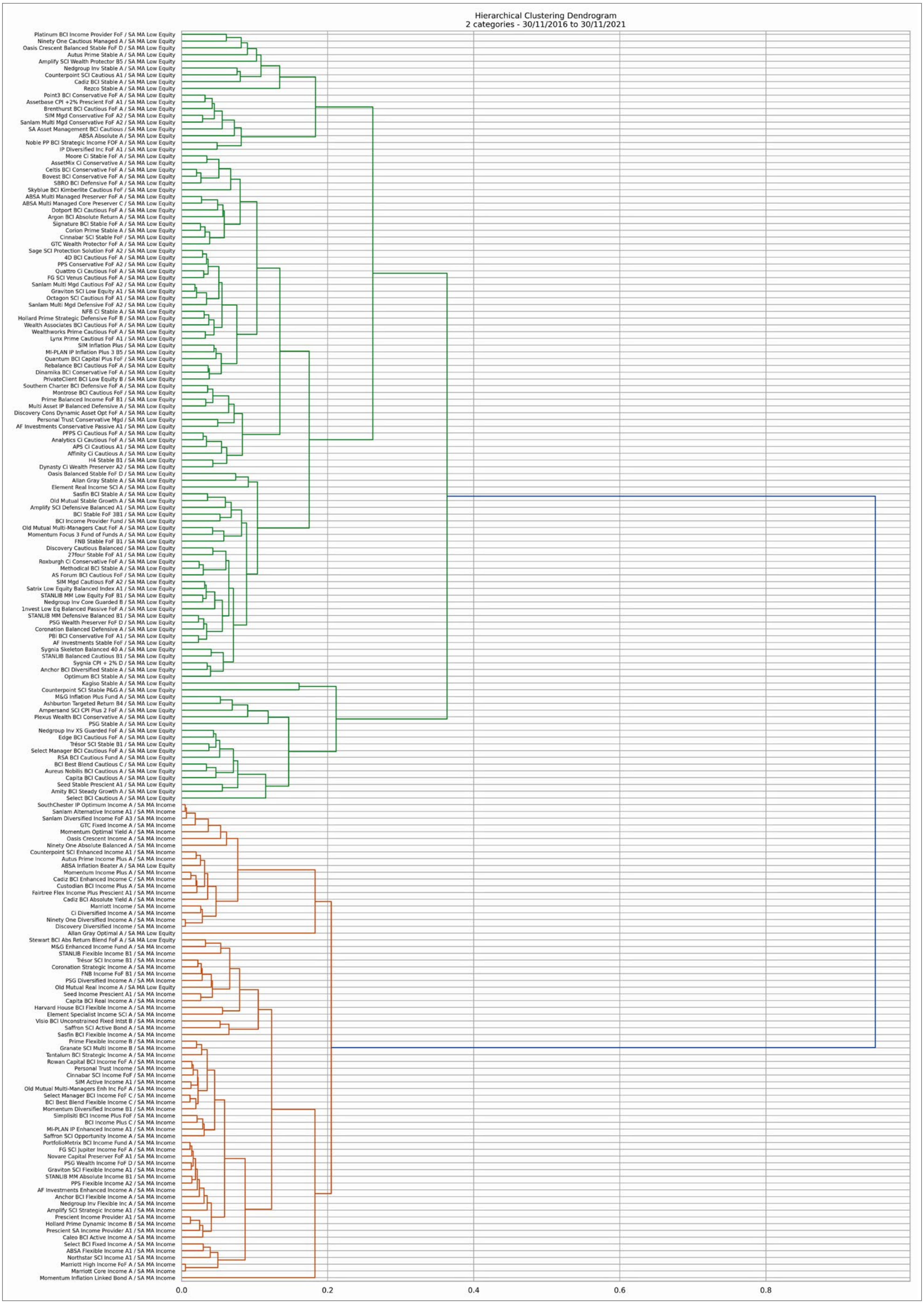


FIGURE 44 Dendrogram – multi asset – low equity & income

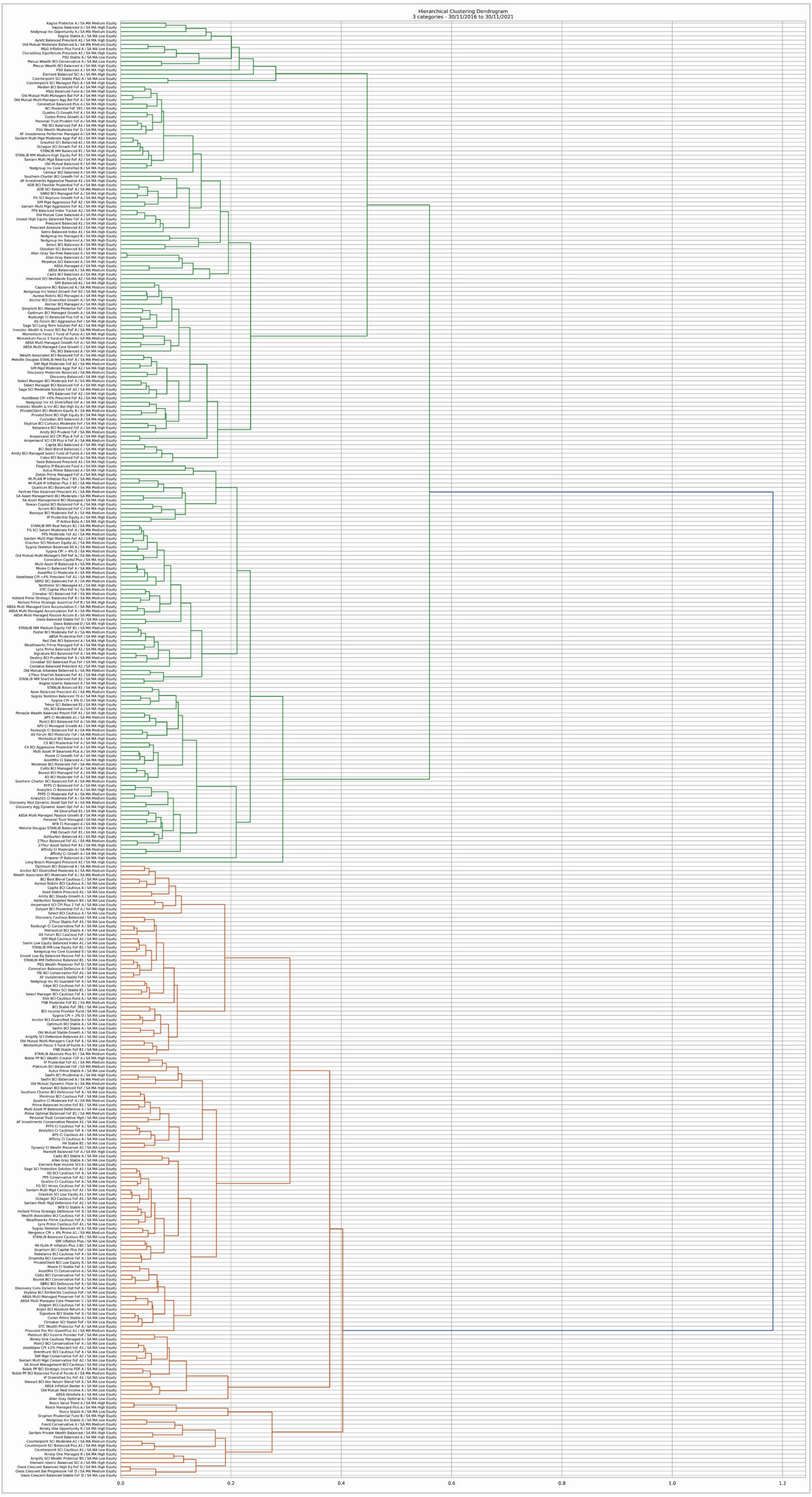


FIGURE 45 Dendrogram – multi asset – high, medium & low equity

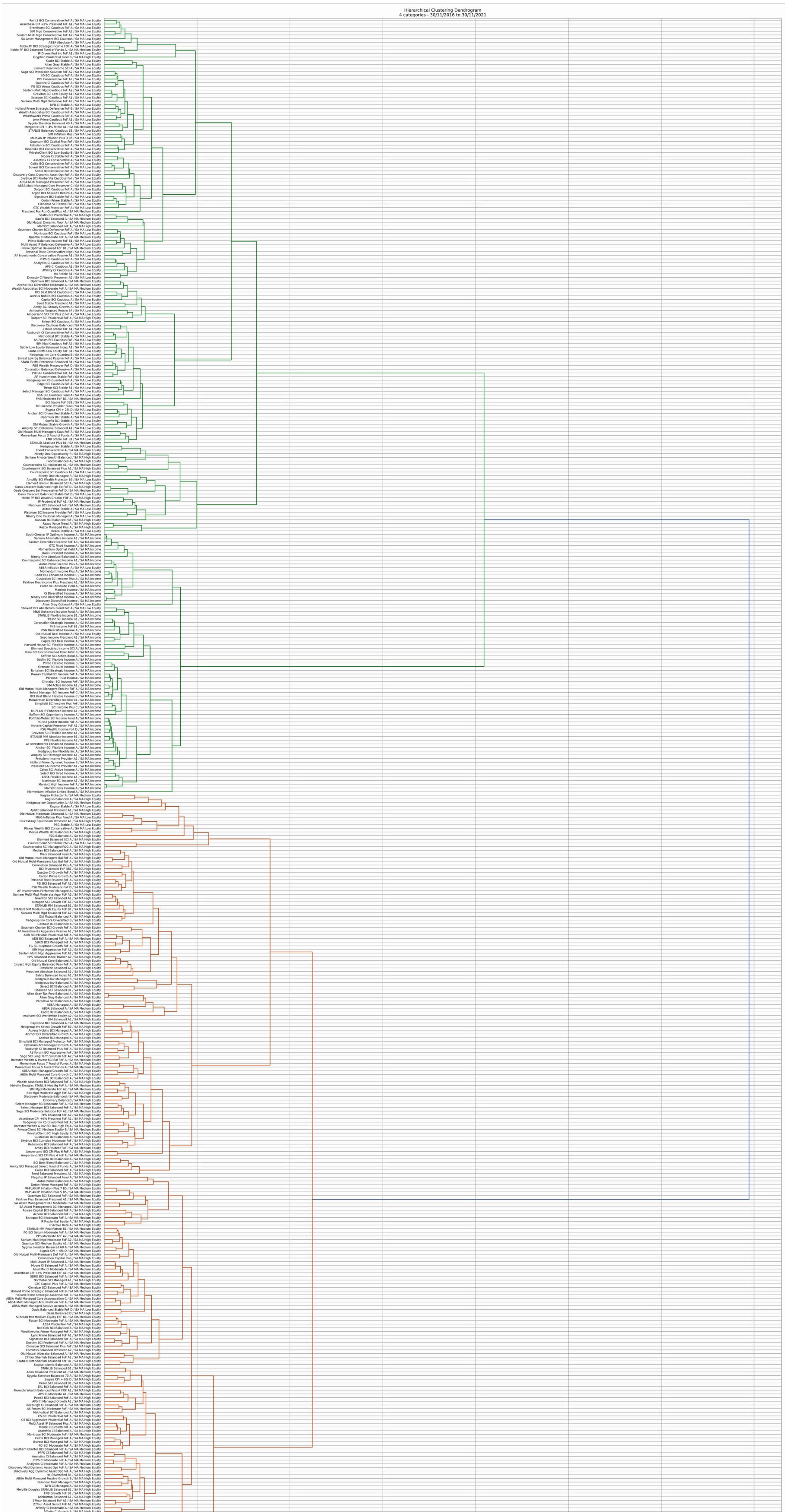


FIGURE 46 Dendrogram – multi asset – high, medium, low equity & income

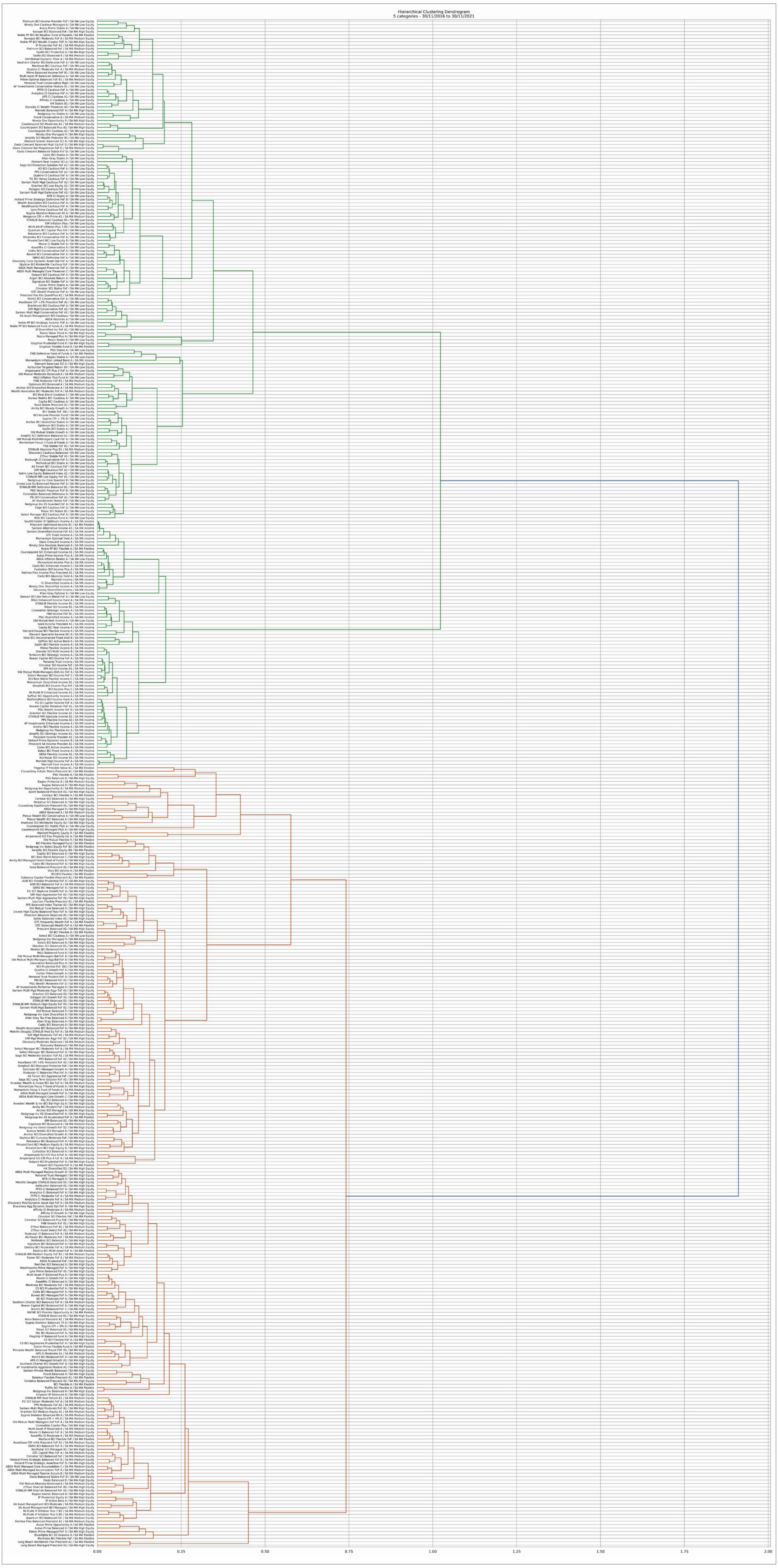


FIGURE 47 Dendrogram – multi asset – all categories (excl. target date)

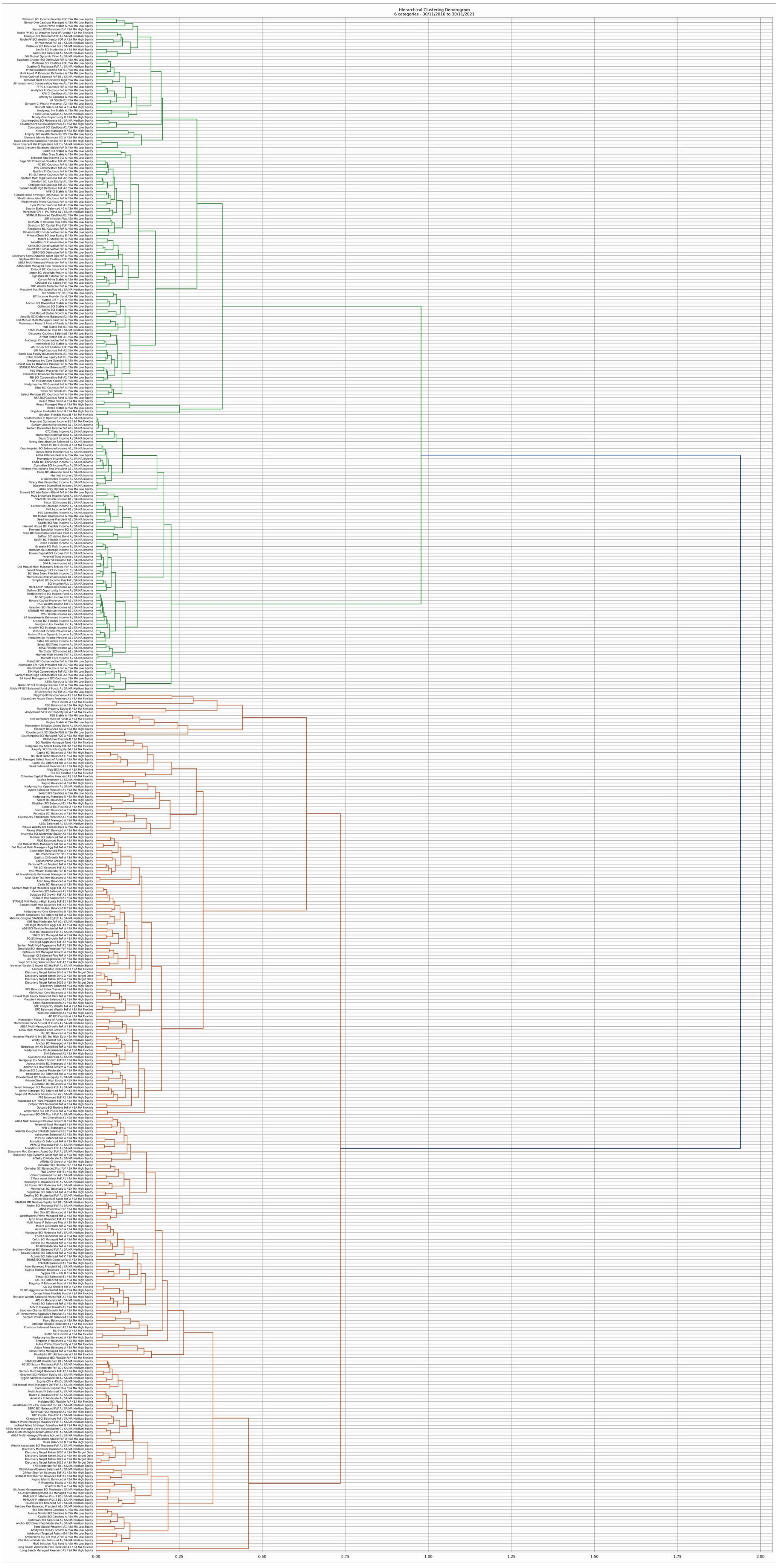


FIGURE 48 Dendrogram – multi asset – all categories

Figure 49 Dendrogram – multi asset – all categories, & equity – general

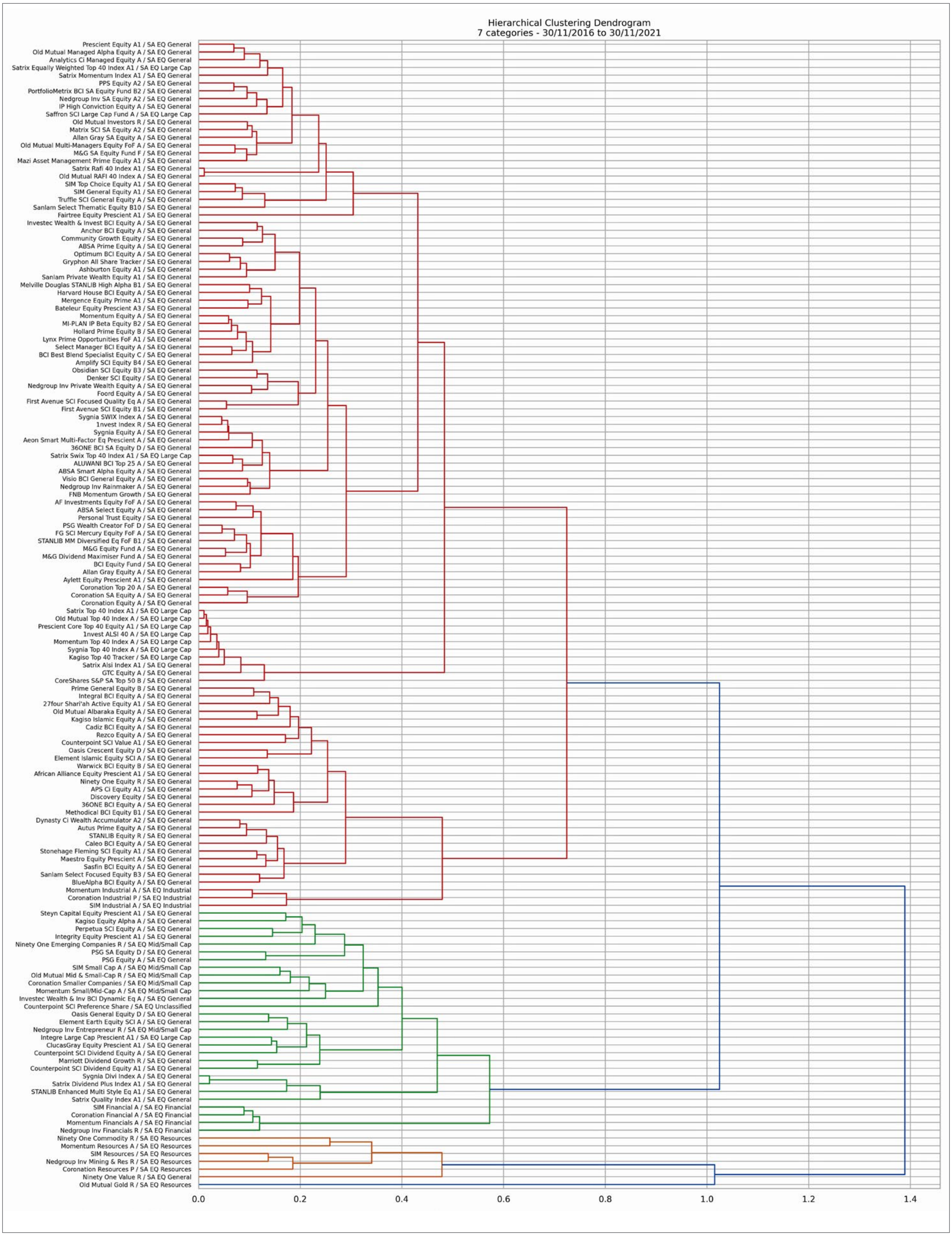


Figure 50 Dendrogram – equity – all categories

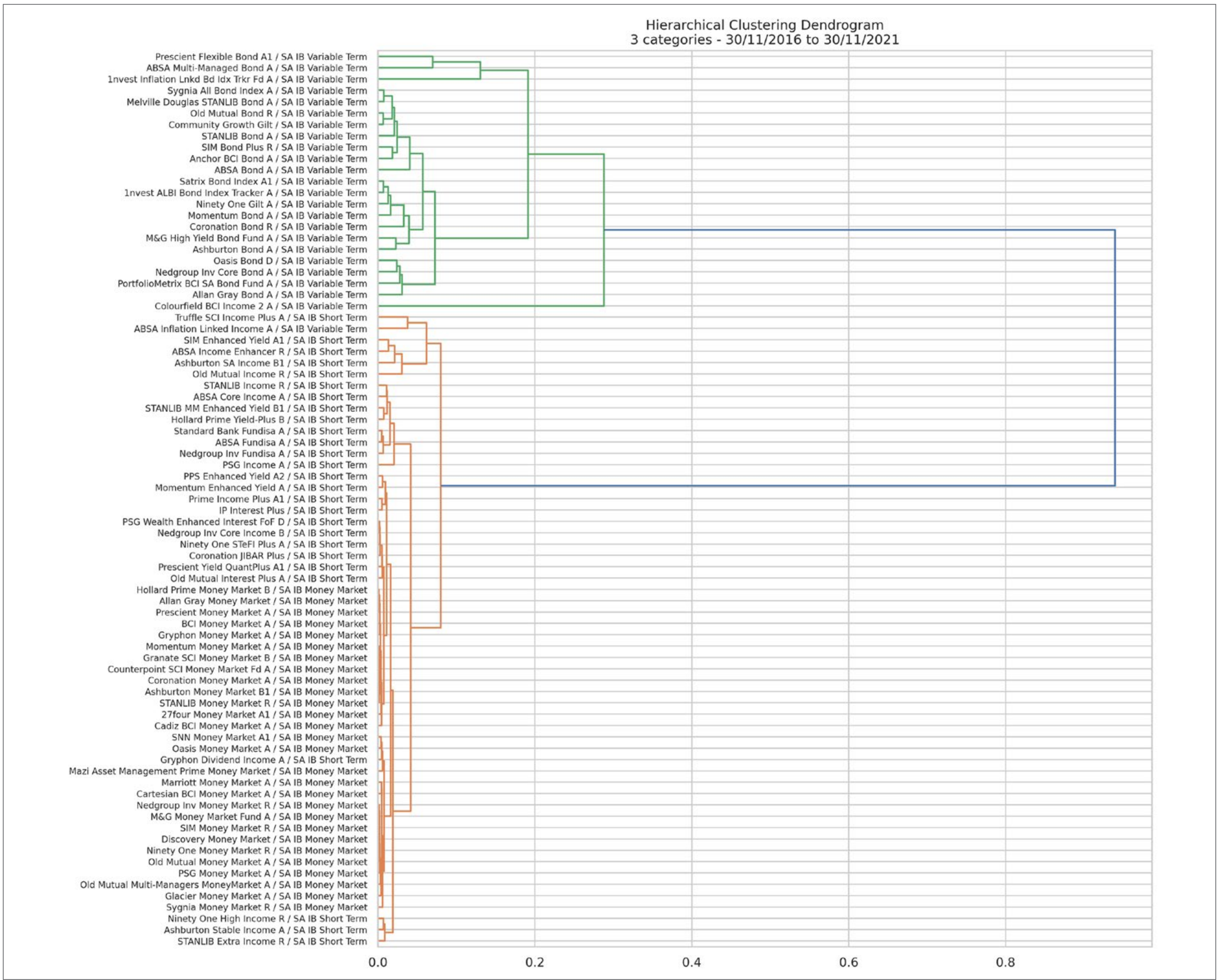


Figure 51 Dendrogram – interest bearing – all categories

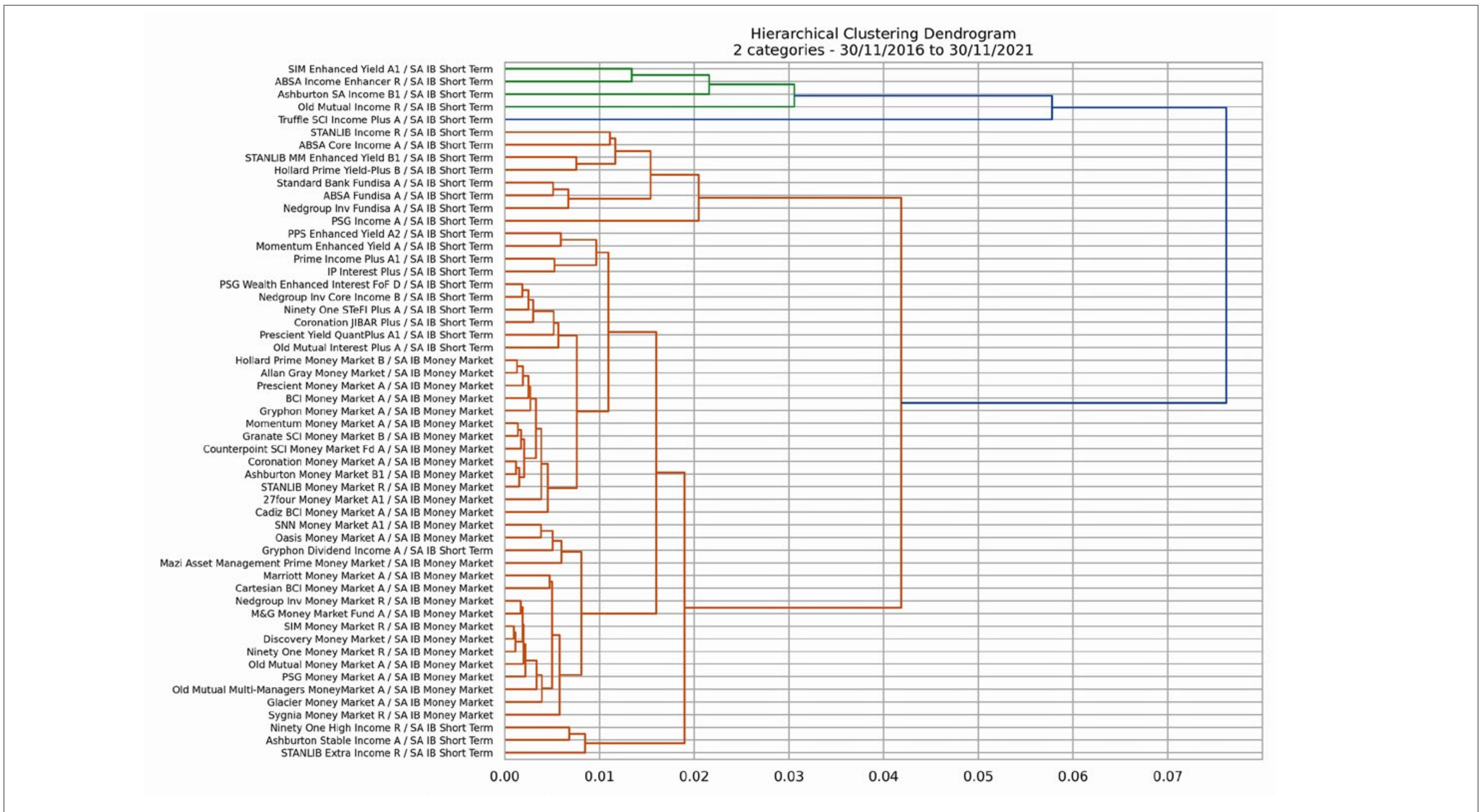


Figure 52 Dendrogram – interest bearing – money market & short term

4.5 Correlations, correlation matrices, heat maps, and cluster maps

Let us take a quick detour back to one of the other traditional methods that are typically used when comparing portfolios, namely correlations and correlation matrices. Correlation measures the goodness of fit of the linear relationship between two variables, in our case the monthly returns of two portfolios. It does not measure the beta (slope/sensitivity) of the relationship, but rather the variability explained by the relationship (technically the square-root of the explained variability).

Within a category of n portfolios, there will be n*(n-1)/2 such pairwise correlations, so the numbers can become difficult to understand in totality when the number of portfolios is high, although still useful individually. Two portfolios only have a single correlation, but three portfolios have three pairwise correlations, and four portfolios have double the amount at six. You can imagine how intractable the problem becomes when you consider the High Equity category with 152 portfolios and 11 476 pairwise correlations.

Figure 53 shows the correlation matrix, which is the correlation between each pair of portfolios, for a relatively small category, namely Equity – Large Cap. Points worth noting include: the matrix is the same above and below the diagonal, signifying that the correlation between portfolio A and portfolio B is the same as the correlation between portfolio B and portfolio A; and the correlations along the diagonal are all 1, as each portfolio is perfectly correlated with itself. This is why you get n*(n-1)/2 pairwise correlations instead of n^2.

Note how much information this matrix includes, essentially 55 pairwise correlations. That is a lot of numbers to process for a very small category with only 11 portfolios. So while individual numbers in the matrix are useful, and even the comparison of some provide some insights into the structure of this category, overall it is very difficult to make complete sense of this data. The problem grows exponentially with the number of portfolios, as we see below, so we need to see if we can do better than this simple visualisation.

One method to help with the visualisation of this amount of data is to produce a heatmap of the correlation numbers in the correlation matrix. Although the colours can be configured to the user's preference, a typical colour set may range from red for higher numbers to black for lower numbers, as per Figure 54. How much easier is it to read the same correlation numbers in this format, versus the way the same information is presented in Figure 53. We are immediately attracted to the dark row and column (portfolio), which is clearly the most different from every other portfolio. Many other pairwise correlations also round to 1, indicating that many of these portfolios are index trackers, all tracking the Top 40 index.

We can however go one step better in terms of making this information more easily digestible, and that is to reorder the portfolios based on the hierarchical clustering we have seen in the previous section. Figure 55 shows the same data in a cluster map, combing the heatmap we have just seen, along with the dendrogram introduced in the previous section. The portfolios are now ordered so that closely related portfolios lie near each other in the matrix, but you can still see the correlation between any two pairs of portfolios that are

FundName	1nvest ALSI 40 A	Integre Large Cap Prescient A1	Kagiso Top 40 Tracker	Momentum Top 40 Index A	Old Mutual Top 40 Index A	Prescient Core Top 40 Equity A1	Saffron SCI Large Cap Fund A	Satrix Equally Weighted Top 40 Index A1	Satrix Swix Top 40 Index A1	Satrix Top 40 Index A1	Sygnia Top 40 Index A
1nvest ALSI 40 A	1.000	0.781	0.996	0.998	0.999	0.999	0.903	0.876	0.950	0.999	0.997
Integre Large Cap Prescient A1	0.781	1.000	0.775	0.783	0.785	0.781	0.909	0.917	0.831	0.786	0.801
Kagiso Top 40 Tracker	0.996	0.775	1.000	0.995	0.996	0.996	0.890	0.873	0.947	0.996	0.994
Momentum Top 40 Index A	0.998	0.783	0.995	1.000	0.998	0.998	0.905	0.877	0.946	0.998	0.995
Old Mutual Top 40 Index A	0.999	0.785	0.996	0.998	1.000	0.999	0.902	0.875	0.947	1.000	0.997
Prescient Core Top 40 Equity A1	0.999	0.781	0.996	0.998	0.999	1.000	0.902	0.874	0.947	0.999	0.996
Saffron SCI Large Cap Fund A	0.903	0.909	0.890	0.905	0.902	0.902	1.000	0.931	0.930	0.905	0.916
Satrix Equally Weighted Top 40 Index A1	0.876	0.917	0.873	0.877	0.875	0.874	0.931	1.000	0.897	0.877	0.889
Satrix Swix Top 40 Index A1	0.950	0.831	0.947	0.946	0.947	0.947	0.930	0.897	1.000	0.950	0.958
Satrix Top 40 Index A1	0.999	0.786	0.996	0.998	1.000	0.999	0.905	0.877	0.950	1.000	0.998
Sygnia Top 40 Index A	0.997	0.801	0.994	0.995	0.997	0.996	0.916	0.889	0.958	0.998	1.000

Figure 53 Correlation matrix – equity – large cap

not closely related, something you could not do with the dendrogram from the previous section.

Unfortunately, this becomes unwieldy very quickly as the number of portfolios increases as can be seen in Figure 56. We therefore do not follow this process any further nor do we show the cluster heat maps for each of the categories covered above, although some of them are small enough to visualise. We are also not able to show the combination of categories as we have done above. As the number of portfolios grows ever larger, any potential understanding of the resulting charts and tables can quickly become overwhelming.

We instead turn to our next unsupervised machine learning technique that allows us to continue to build on the knowledge we have gained above. Next up is K-Means Clustering.

4.6 K-Means clustering

4.6.1 PORTFOLIO CLUSTERS WITHIN CATEGORIES

Let us begin by considering K-means clustering within each of the categories we have considered above. Perhaps it is important to know some of the outputs from this technique. The model will provide the cluster that each portfolio belongs to, within the number of clusters specified by the user. It will also provide the centroid of each cluster, along each dimension (feature) provided, hence a vector with 60 elements, one for each of the months that returns were provided for.

Visualising which portfolios fall within which cluster is rather trivial, and doesn't

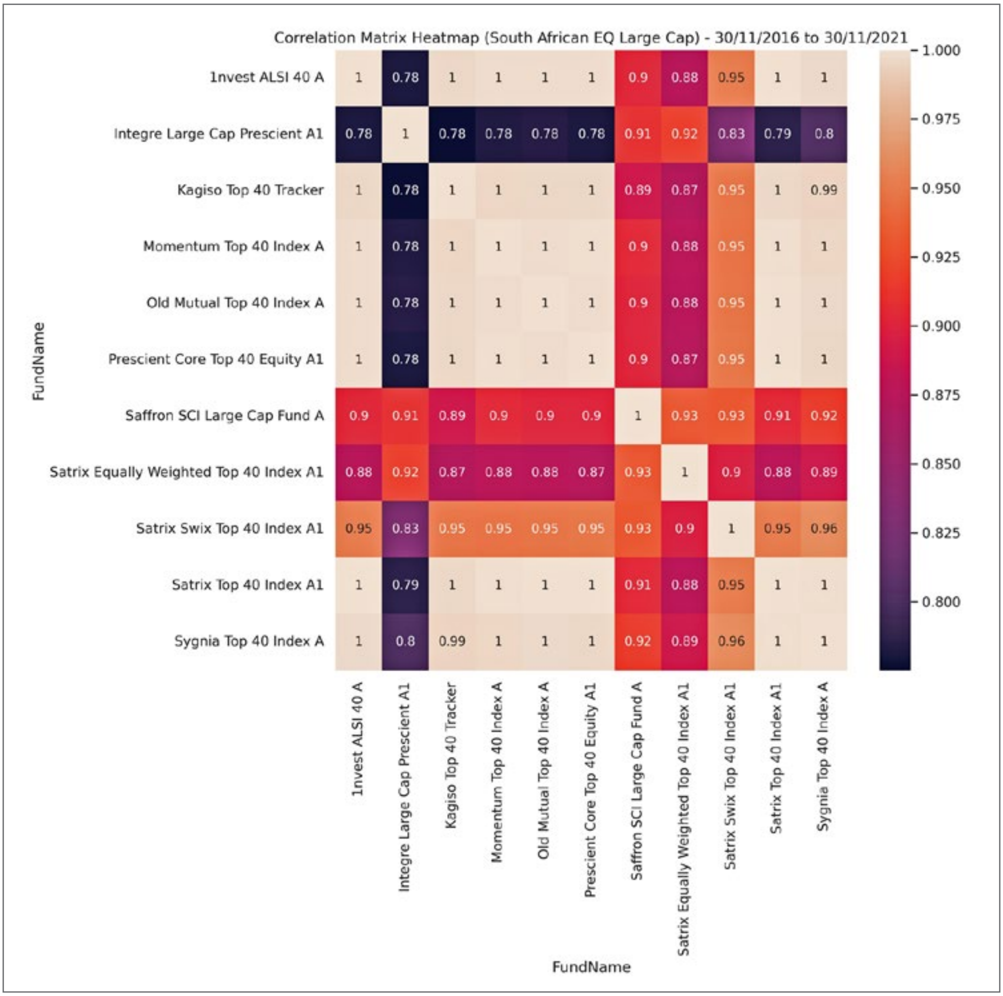
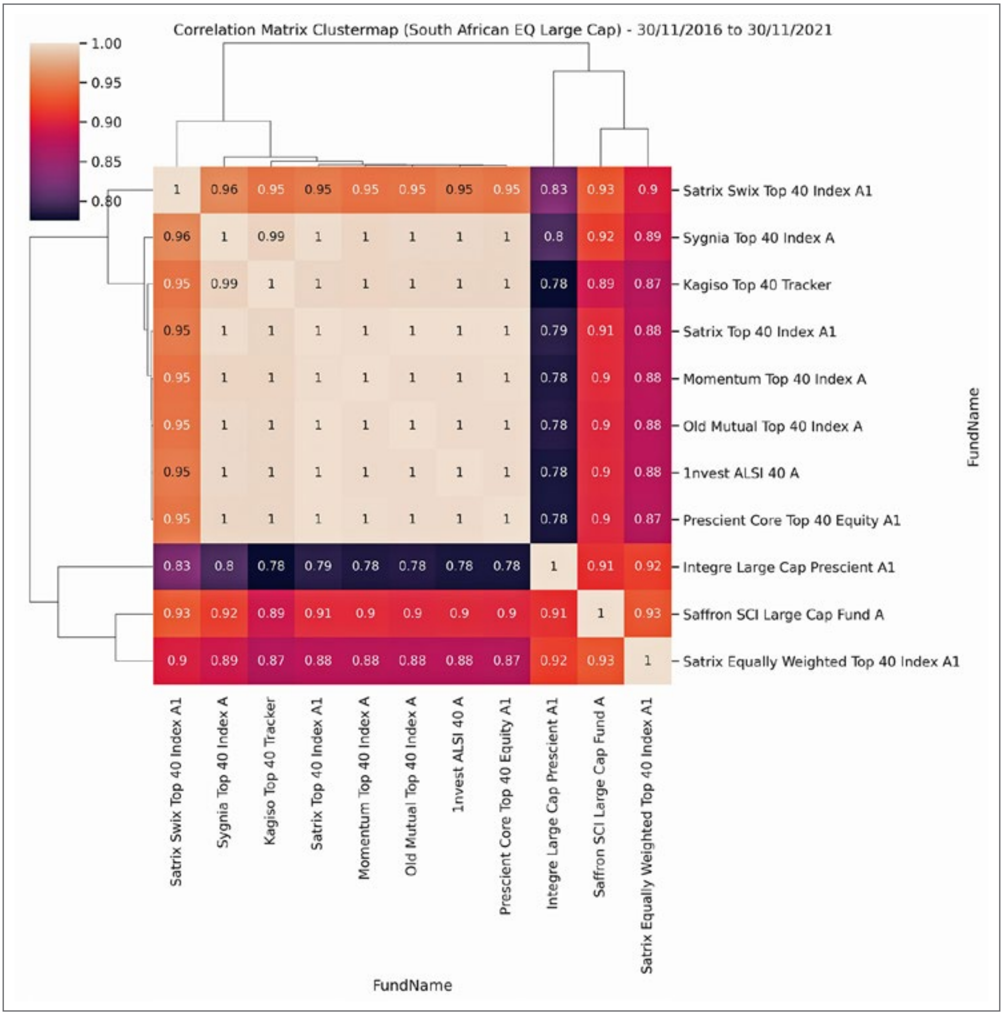


Figure 54 Correlation matrix heatmap – equity – large cap

provide any information about the relative position of the portfolios within each cluster, or the relative position of the clusters. If we had just two dimensions, we could plot the portfolios along these two dimensions, along with the cluster centroids to appreciate all of these relative positions. We could also do this with three dimensions, although this becomes a little more difficult to visualise on a two-dimensional surface such as a piece of paper or computer screen. Visualising 60 dimensions is however impossible, so we need to consider alternatives.

Although some readers may want to see the lists of portfolios in each cluster using this technique, we think the value of this is very limited and we therefore do not provide this. It is however readily available from the algorithm that provides the cluster number for each portfolio.



	Satrix Swix Top 40 Index A1	Sygnia Top 40 Index A	Kagiso Top 40 Tracker	Satrix Top 40 Index A1	Momentum Top 40 Index A	Old Mutual Top 40 Index A	1nvest ALSI 40 A	Prescient Core Top 40 Equity A1	Integre Large Cap Prescient A1	Saffron SCI Large Cap Fund A	Satrix Equally Weighted Top 40 Index A1
Satrix Swix Top 40 Index A1	1	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.83	0.93	0.9
Sygnia Top 40 Index A	0.96	1	0.99	1	1	1	1	1	0.8	0.92	0.89
Kagiso Top 40 Tracker	0.95	0.99	1	1	1	1	1	1	0.78	0.89	0.87
Satrix Top 40 Index A1	0.95	1	1	1	1	1	1	1	0.79	0.91	0.88
Momentum Top 40 Index A	0.95	1	1	1	1	1	1	1	0.78	0.9	0.88
Old Mutual Top 40 Index A	0.95	1	1	1	1	1	1	1	0.78	0.9	0.88
1nvest ALSI 40 A	0.95	1	1	1	1	1	1	1	0.78	0.9	0.88
Prescient Core Top 40 Equity A1	0.95	1	1	1	1	1	1	1	0.78	0.9	0.87
Integre Large Cap Prescient A1	0.83	0.8	0.78	0.79	0.78	0.78	0.78	0.78	1	0.91	0.92
Saffron SCI Large Cap Fund A	0.93	0.92	0.89	0.91	0.9	0.9	0.9	0.9	0.91	1	0.93
Satrix Equally Weighted Top 40 Index A1	0.9	0.89	0.87	0.88	0.88	0.88	0.88	0.87	0.92	0.93	1

Figure 55 Correlation matrix cluster map – equity – large cap

We instead go back to our risk versus return scatterplots, which were so useful in understanding the overlap between portfolios in different categories, or the ranges within single categories, and use the cluster to colour the portfolios (dots), to see if this adds additional colour (bad pun intended) to the analysis.

4.6.1.1 SOUTH AFRICAN – MULTI ASSET

Figure 57 shows the same risk versus return scatterplot for the High Equity category that we saw before, with the addition of colours to the dots (portfolios) based on their cluster from the K-means clustering algorithm. While we could plot the centroids of each of these colours based on these two dimensions of risk and return, it is important to note that this is not how the clustering was done. To re-iterate, the clustering was done on the

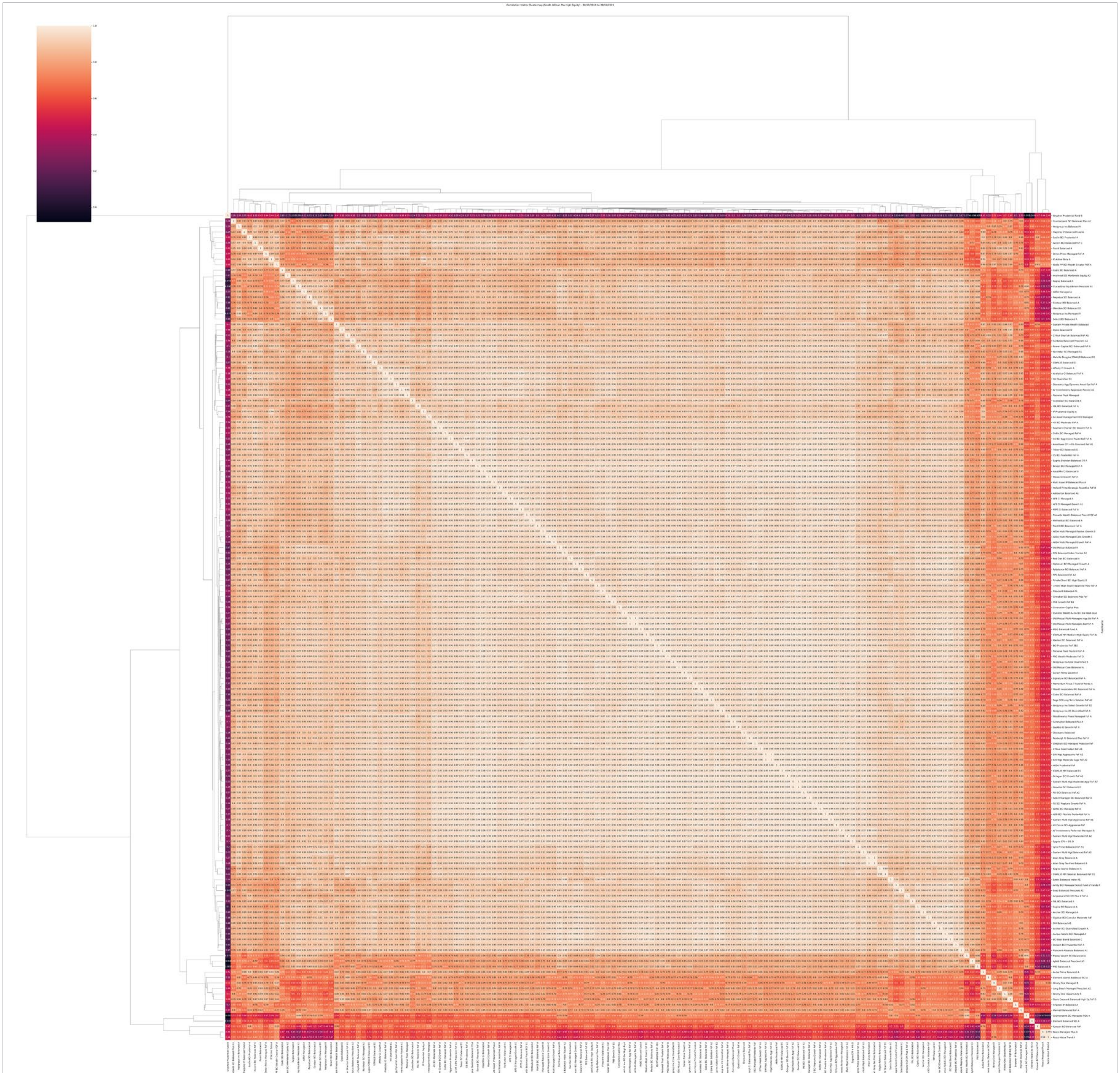


FIGURE 55 Correlation matrix cluster map – multi asset – high equity

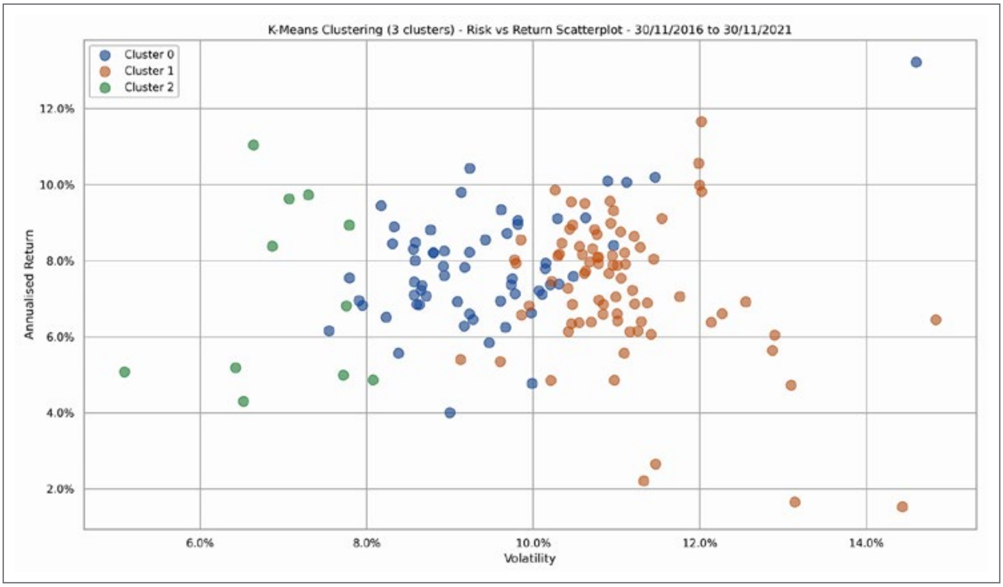


FIGURE 57 K-means clustering (three clusters) – multi asset – high equity

60 monthly returns provided for each portfolio, so the centroid for each cluster exists in that 60-dimensional space which we cannot visualise. We are merely looking at the clusters created in risk versus return space instead.

The choice of three clusters is a hyperparameter of the K-means clustering algorithm, and we chose three clusters for this category because we have already seen that the portfolios appear to be managed on the basis of low, medium, and high risk. Note that there is no guarantee, or perhaps even expectation, that portfolios would cluster on this basis, if we had not seen the data before, or the analysis above.

You will see that although there is still a degree of overlap between the clusters, they do seem to separate into a low risk, medium risk, and high risk cluster, with some medium risk cluster portfolios clearly having very high risk. We have alluded to this above and come back to it below as promised.

As a quick aside, we could easily perform a K-means clustering using the annualised return and risk as the two features of the data, or even a single dimension of risk if we wanted, and then use the same risk versus return scatterplot to visualise the data along with the centroid for each cluster, which would now be the actual centroid as the dimensions are identical. For risk alone, this would be a single dimension. We look at this briefly below just to get a sense for the information that this could provide, which again may provide some useful insights and spark additional ideas on how many different ways there are to analyse and visualise this data and how portfolios are structured.

Figure 58 shows the risk versus return scatterplot for the Medium Equity category, but with colours representing the clustering. We have again chosen three clusters given our

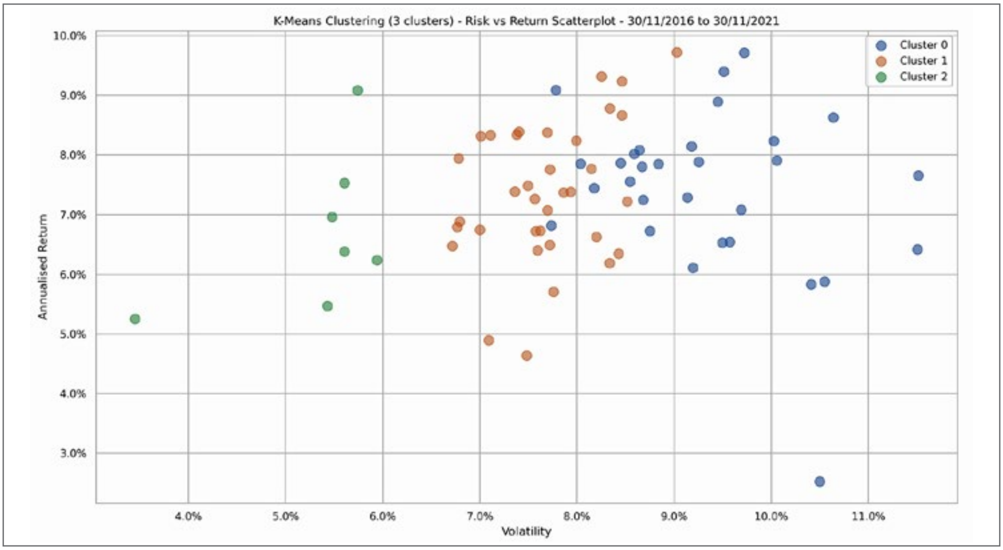


FIGURE 58 K-means clustering (three clusters) – multi asset – medium equity

knowledge of how these portfolios are managed, and see a similar pattern to the High Equity category, with some portfolios being managed on a low-risk basis and clustering together, and some on a high-risk basis also clustering together, although with some overlap with the portfolios that are perhaps managed on a medium-risk basis.

Figure 59 shows the risk versus return scatterplot for the Low Equity category, but with colours representing the clustering. We have again chosen three clusters given our

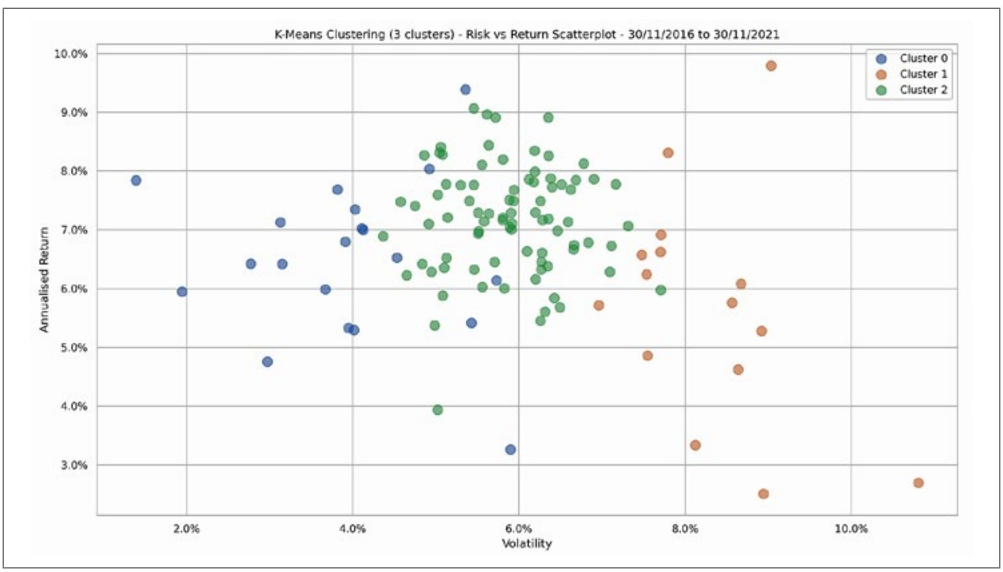


FIGURE 59 K-means clustering (three clusters) – multi asset – low equity

knowledge of how these portfolios are managed, and see a similar pattern to the High Equity and Medium Equity categories, with portfolios being managed on a lower-risk, medium-risk, and higher-risk basis (relatively speaking, as the risk is lower than for the higher equity categories). The clusters are fairly distinct with little overlap.

Figure 60 shows the same risk versus return scatterplot that we saw before for the Income category, but with coloured clusters. Although we could have easily just chosen two clusters in this case, choosing three shows us the solitary portfolio on the bottom right corner of the chart in its own cluster. This portfolio is probably the inflation-linked bond portfolio that sits in this category, and was previously shown to be completely different to the rest of the category. The other two clusters are fairly well demarcated along the risk spectrum as we have seen in the other categories.

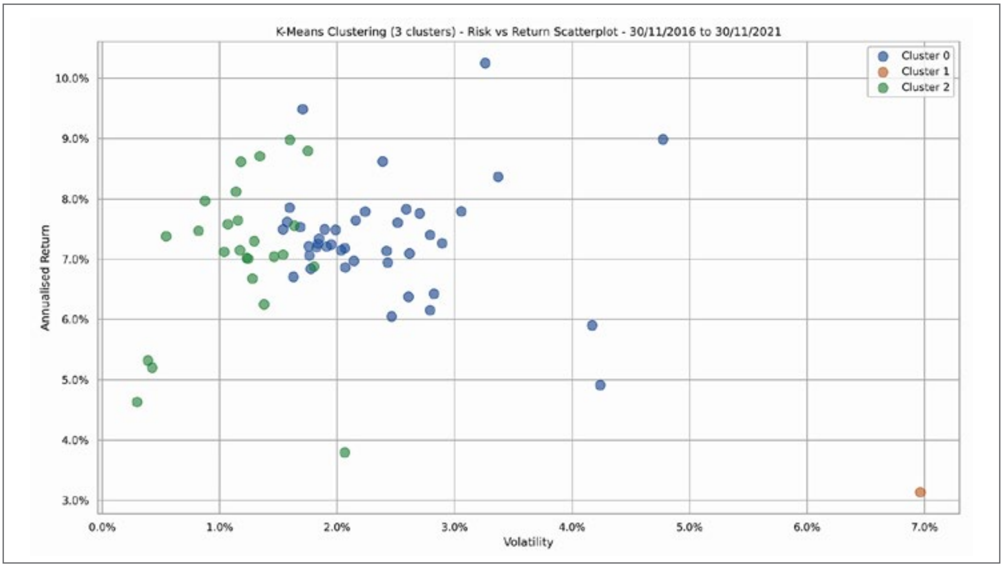


Figure 60 K-Means clustering (three clusters) – multi asset – income

Figure 61 shows the risk versus return scatterplot for the Flexible category with coloured clusters. As this category is such a mess in terms of heterogeneous portfolios, along many different dimensions, we tried various values for the number of clusters before settling on five.

On the basis of these five clusters, we had: one with two flexible property or property-dominated portfolios: one with the two very distinct equity portfolios with little global exposure; one with the three more conservative, income-type portfolios; one with high equity portfolios with less global exposure; and the final one also with high equity exposure but with much more global exposure. Please refer to the Hierarchical Clustering Dendrogram for this category to see how the clusters looked in that chart.

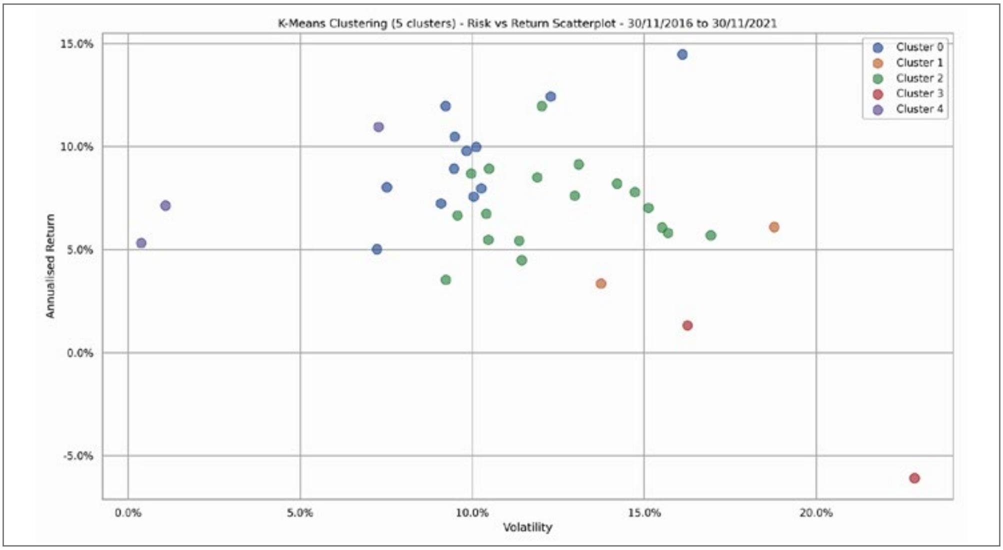


Figure 61 K-Means clustering (three clusters) – multi asset – flexible

4.6.1.2 SOUTH AFRICAN – EQUITY

Figure 62 shows the risk versus return scatterplot for the Equity General category with two coloured clusters. As previously argued, it is difficult to classify a group of portfolios with volatility ranging from 10% p.a. to 22% p.a. as fairly homogeneous, but it is also difficult to create a demarcation based solely on volatility.

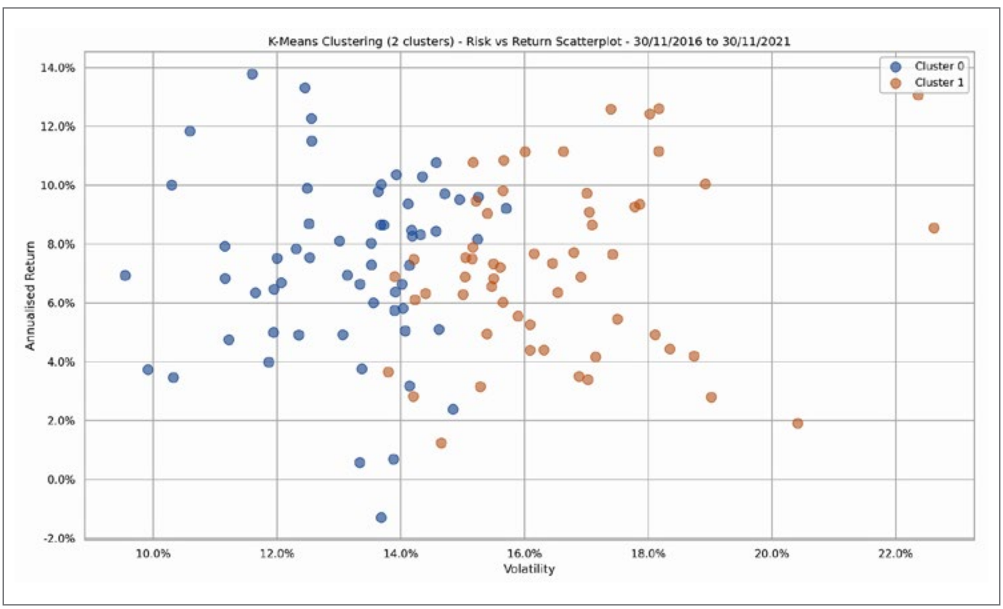


Figure 62 K-Means clustering (two clusters) – equity – general

Using only two clusters we do get a fairly decent demarcation along the volatility axis, but again we remind the reader that the clustering was not done on this basis, so the little overlap that exists implies a really good result. Could this be used in any way to possibly sub-divide this category? We consider one possible sub-division below, but it may be completely unrelated to the clustering presented here.

4.6.1.3 SOUTH AFRICAN – INTEREST BEARING

Figure 63 shows the risk versus return scatterplot for the Variable Term category, along with three coloured clusters.

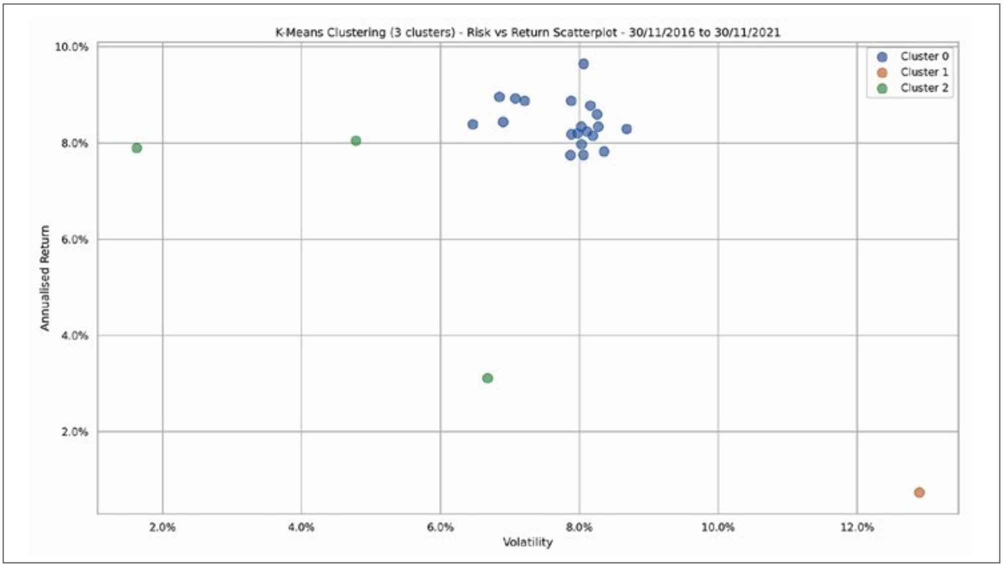


FIGURE 63 K-Means clustering (three clusters) – interest bearing – variable term

Again, consistent with the dendrogram that we saw before for this category, one portfolio sits in a cluster all alone, while another cluster contains the inflation-linked bond and flexible duration portfolios, with the final cluster containing the rest of the portfolios being managed against the All Bond Index.

Figure 64 shows the risk versus return scatterplot for the Short Term category, along with three coloured clusters. These clusters are again fairly distinct with the cluster on the left representing the very conservatively managed portfolios, and the cluster on the right representing the portfolios managed more aggressively. The solitary portfolio on the right is clearly doing something very different. Dropping the analysis to just two clusters actually forces this portfolio on the top right to cluster with the more conservatively managed portfolios which is rather bizarre and would require further investigation; we suspect it is the lack of duration in this portfolio that creates this clustering, i.e. the extra risk does not come from interest rate duration.

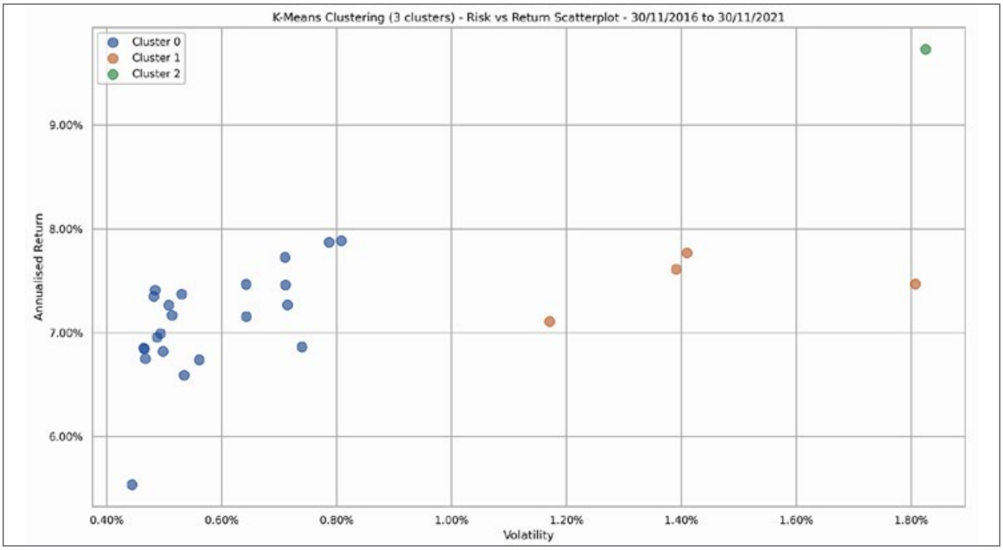


FIGURE 64 K-Means clustering (three clusters) – interest bearing – short term

4.6.1.4 SOUTH AFRICAN – REAL ESTATE

The Real Estate category does not yield any interesting results along the risk versus return dimensions, so we do not include it here. When we looked at the Hierarchical Clustering and the resultant dendrogram, there are very distinct clusters within this category, which were confirmed with the K-means clustering technique, but it does not translate nicely into the risk and return dimensions, which is an important lesson. We made the point before that two very different portfolios, in terms of how they are managed and the shares they hold, could produce relatively similar risk and return numbers, purely by chance.

4.6.2 PORTFOLIO CLUSTERS USING RISK, OR RISK AND RETURN AS DIMENSIONS

As mentioned above, we could apply a clustering algorithm like K-means clustering to just the volatility numbers of portfolios, or the risk and returns of the portfolios, instead of applying it to the monthly returns. This may be useful if risk was a good indicator of the similarity and differences of portfolios but, as we have argued, it is not necessarily so e.g. two portfolios with the same risk number could be very different. It is hopefully more obvious that two portfolios that have very different risk numbers are unlikely to be similar.

Figure 65 shows the result of clustering on the basis of the risk (volatility) dimension only. Now we have very distinct clusters with no overlap as every portfolio belongs to the cluster of which it is closest to the centroid. Looking at the categories being considered in this analysis, we could arbitrarily label the cluster below from left to right as Income, Low Equity, Medium Equity, High Equity, and Equity. If we analysed whether the portfolios in those clusters actually were in those categories, we would possibly find many mismatches. We explore this in the next section.

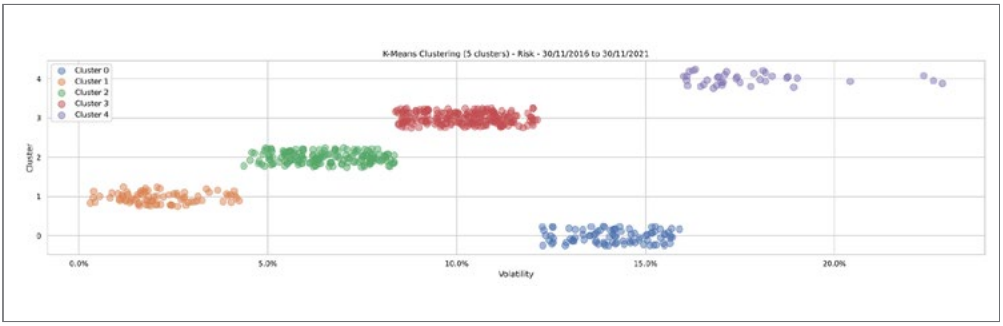


FIGURE 65 K-Means clustering (five clusters) – risk – multi asset (excl. target date) & equity general

Now if we use both dimensions of risk and return to apply the K-Means clustering algorithm, we get the results in Figure 66. See how distinct every cluster is because the same dimensions are used to perform the clustering and to visualise the results.

FIGURE 66 K-Means clustering (five clusters) – risk vs return –multi asset (excl. target date) & equity general

How meaningful is the clustering though, in terms of grouping homogeneous portfolios? You have to appreciate the demarcation problem when you essentially have variables that sit on a continuum, such as risk. Two portfolios that sit next to each other but are from different clusters are more alike than they are with other portfolios from the same cluster that are further away.

4.6.3 PORTFOLIO CLUSTERS USING CROSS TABULATION

4.6.3.1 SOUTH AFRICAN – MULTI ASSET – HIGH, AND MEDIUM EQUITY (TWO CLUSTERS)

There are many different ways that we could think about using K-means clustering to analyse the returns from the portfolios in all the different categories that we have considered so far. We could use the knowledge that we have gained from some of the other methods, both traditional and machine learning, to inform the parameters that we use to analyse the portfolios further.

In Figure 67, we consider the portfolios in the Multi Asset High Equity and Medium Equity categories, and ask the algorithm to provide us with just two clusters for the data. Instead of looking at which portfolios fall into each of those clusters, we want to see what category those portfolios fall into for each cluster, to see whether we can establish some kind of pattern or relationship.

```
K-Means Clustering Crosstab - 30/11/2016 to 30/11/2021

Cluster                          0    1
Category
South African MA High Equity    117   35
South African MA Medium Equity   27   45

Cluster       0   1
Volatility 10.5 7.8
```

Figure 67 K-Means clustering (two clusters) – multi asset – high & medium equity

Let us try to make some sense of these results. Cluster 1 seems to be dominated by portfolios from the Medium Equity category (45 versus 35 from the High Equity category). Looking at it another way, there are more Medium Equity portfolios in cluster 1 than there are in cluster 0 (45 versus 27). Cluster 0, on the other hand, seems to be dominated by portfolios from the High Equity category (117 versus 27 from the Medium Equity category). Again looking at it another way, there are more High Equity portfolios in cluster 0 than in cluster 1 (117 versus 35).

No matter how we look at the data, cluster 1 appears to be more representative of the Medium Equity category, and cluster 0 more representative the High Equity category. This also appears to be confirmed by the volatility numbers supplied at the bottom of the table, where cluster 1 has portfolios with an average volatility of 7.8% p.a., and for cluster 0 the average volatility of the portfolios is 10.5% p.a.

This can also be seen in Figure 68, a risk versus return scatterplot for the portfolios in these two categories, but with a difference, as we have distinguished between each of the combinations of category and cluster. Cluster 0 is in red, and cluster 1 is in green, and although there is some overlap, there also appears to be a fairly good demarcation. We also then distinguish between portfolios in the Medium Equity portfolios, marked with a filled 'o', and portfolios in the High Equity category, marked with a 'x'.

We could use the clustering derived from this method to understand what some of the other differences are between the two clusters, subject to the availability of data. For example, if we had asset allocations over the five-year period, say averages, we may find that one cluster does in fact have more equity, local and/or total, than the other, or perhaps more offshore assets, or even more property. If you know the underlying portfolios well, you will in fact find this to be the case.

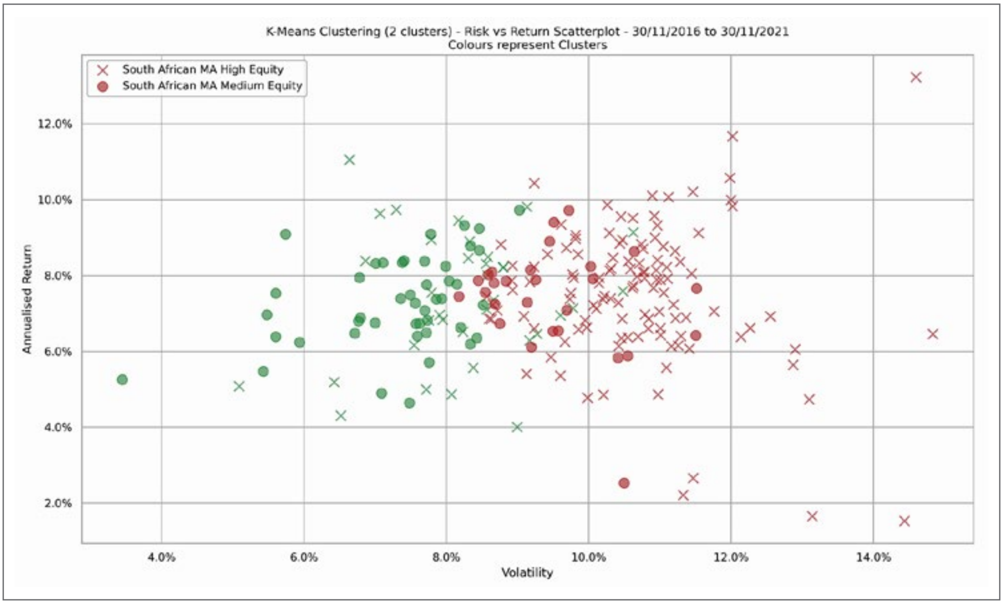


Figure 68 K-Means clustering (2 clusters) – risk vs return –multi asset – high & medium equity

This is by itself not that interesting. What is more interesting is that many portfolios are potentially misclassified, if we were to use the clusters as a guide to an appropriate classification of portfolios, using K-Means Clustering of returns. Specifically, 35 portfolios that are classified in the High Equity category would probably be better classified in the Medium Equity category. Similarly, 25 portfolios classified as Medium Equity portfolios may be better classified as High Equity. Obviously you could have a potential issue classifying High Equity portfolios as Medium Equity, if they exceed the Medium Equity total equity exposure limit. It would also not make sense for Medium Equity portfolios, that do not breach the total equity exposure limit, to be classified as High Equity.

It is also worth pointing out that we are using an unsupervised machine learning technique, in that we have not trained a model by providing the labels for each portfolio (their category). We have merely provided returns with no output label, and asked the algorithm for the structure of the data. We could have instead applied a supervised machine learning technique where we train the model on what a High Equity portfolio looks like, versus a Medium Equity portfolio, and then asked the model to classify each portfolio based on this trained model. Any portfolios that are misclassified could then be seen as problems with the data (portfolios' classifications), instead of a problem with the model (error in classifying correctly).

4.6.3.2 SOUTH AFRICAN – MULTI ASSET – HIGH AND MEDIUM EQUITY (THREE CLUSTERS)

Given that we know that some High Equity portfolios and some Medium Equity portfolios are lower risk than the typical portfolios in their categories, or the averages of the categories, we may want to ask the algorithm to instead give us three clusters for these two categories, hoping that one cluster catches the lower risk portfolios, thus allowing the other two clusters to capture the higher risk portfolios (medium and high).

In Figure 69, we can see that a new cluster (numbered 0) has taken some portfolios from each of the two categories, more so from the Medium Equity category, possibly implying that it does represent the lowest risk cluster.

Looking at the volatility numbers at the bottom of the table, we can confirm that cluster 0 is in fact the lowest volatility cluster at a volatility of 6.7% p.a. This has also allowed both cluster 1 and cluster 2 to include portfolios with higher volatilities on average.

We hope that it is not lost on the reader that this method (K-means clustering), is providing some validity to using risk as a basis to decide how similar or different portfolios are, something that may have been intuitively obvious from the beginning, but without further support until now.

Evidence for this can be seen in Figure 70 which shows the risk versus return scatterplot for all of the portfolios in these two categories, colour coded by the cluster in which they fall, with red representing the new lowest risk cluster.

```
K-Means Clustering Crosstab - 30/11/2016 to 30/11/2021

Cluster                         0   1   2
Category
South African MA High Equity   11  57  84
South African MA Medium Equity 14  48  10

Cluster      0   1    2
Volatility 6.7 8.9 10.9
```

FIGURE 69 K-Means clustering (three clusters)– multi asset – high & medium equity

FIGURE 70 K-Means clustering (three clusters) – risk vs return – multi asset – high & medium equity

4.6.3.3 SOUTH AFRICAN – MULTI ASSET – HIGH, MEDIUM AND LOW EQUITY (THREE CLUSTERS)

If we add the Low Equity category into the mix while keeping the number of clusters at three, we should see the High Equity and Medium Equity portfolios with lower volatility cluster with the new Low Equity portfolios introduced. Figure 71 summarises these results. Note that the cluster numbers by themselves have no meaning.

```
K-Means Clustering Crosstab - 30/11/2016 to 30/11/2021

Cluster                          0   1   2
ASISACategory
South African MA High Equity     9  87  56
South African MA Low Equity    107   7   4
South African MA Medium Equity  10  15  47

Cluster       0    1   2
Volatility 5.7 10.7 8.8
```

FIGURE 71 K-Means clustering (three clusters) – multi asset – high, & medium & low equity

Let us first understand what the three clusters represent as before. Cluster 0 seems the easiest to understand as it has 107 portfolios from the Low Equity category, and very few by comparison from the other two categories, with nine from High Equity, and ten from Medium Equity. The average volatility of this category is also the lowest at 5.7% p.a. Cluster 0 therefore seems to represent the Low Equity cluster.

Cluster 1 not only has the highest volatility at 10.7% p.a., but also has the most portfolios from the High Equity category at 87. It is perhaps surprising that some portfolios from the other two categories would fall into this cluster, given the categories' total equity exposure limitations, but we have previously seen this happen, and we explain this below. Cluster 1 therefore appears to represent the High Equity category.

Finally, most portfolios from the Medium Equity category fall into cluster 2, with 47 in total. Although this cluster has even more portfolios from the High Equity category at 56, we know that the High Equity category can have much lower risk portfolios with less total equity exposure. This cluster also has the second highest volatility at 8.8% p.a. Cluster 2 therefore best represents the Medium Equity category.

The interactions between the categories and the clusters can be seen in the risk versus return scatterplot in Figure 72.

Figure 72 K-Means clustering (three clusters) – risk vs return – multi asset – high, medium & low equity

4.6.3.4 *SOUTH AFRICAN – MULTI ASSET – ALL CATEGORIES (EXCL. FLEXIBLE & TARGET DATE) (FOUR CLUSTERS)*

Figure 73 includes the Income category in the mix with the other three categories, and increases the number of clusters to four, to account for low risk Income category portfolios.

Cluster 0 now appears to represent the Income category portfolios, while including

12 Low Equity category portfolios, and one each from the other two categories, confirming just how conservatively managed these portfolios are.

Cluster 2 definitely represents the Low Equity category with no Income category portfolios, but numerous portfolios from the other two categories, namely 12 from High Equity and 24 from Medium Equity.

Cluster 1 appears to represent the Medium Equity category, but perhaps the riskier of the portfolios in this category, thus attracting many portfolios from the High Equity category that are managed slightly less aggressively.

Finally, cluster 3 appears to represent the High Equity category, although again it includes 12 portfolios from the Medium Equity category, and seven from the Low Equity category.

Note how close the volatilities of cluster 1 and cluster 3 are, suggesting that there could be some overlap between these two clusters. How this all plays out can be seen in Figure 74, the risk versus return scatterplot, where the overlap between cluster 1 (green) and cluster 3 (cyan) is more apparent.

K-Means Clustering Crosstab - 30/11/2016 to 30/11/2021

Cluster	0	1	2	3
Category				
South African MA High Equity	1	65	12	74
South African MA Income	67	0	0	0
South African MA Low Equity	12	0	99	7
South African MA Medium Equity	1	35	24	12

Cluster	0	1	2	3
Volatility	2.3	9.2	6.3	10.8

FIGURE 73 K-Means clustering (four clusters) – multi asset – all categories (excl. flexible & target date)

4.6.3.5 SOUTH AFRICAN – MULTI ASSET – ALL CATEGORIES (FIVE CLUSTERS)

Figure 75 performs the clustering on all the Multi Asset categories, including Flexible and Target Date, and asks the algorithm to produce just five clusters from all of these portfolios. Clusters 1 through 4 actually resemble the clusters from the previous section quite closely i.e. Low Equity, Medium Equity, High Equity, and Income respectively. Cluster 0 seems to focus on the Flexible category portfolios that don't fit neatly elsewhere, and probably includes the portfolios with the highest property exposure. Also note the average volatilities at the bottom of the table to confirm the analysis provided above.

Figure 76 shows these relationships in the risk versus return scatterplot.

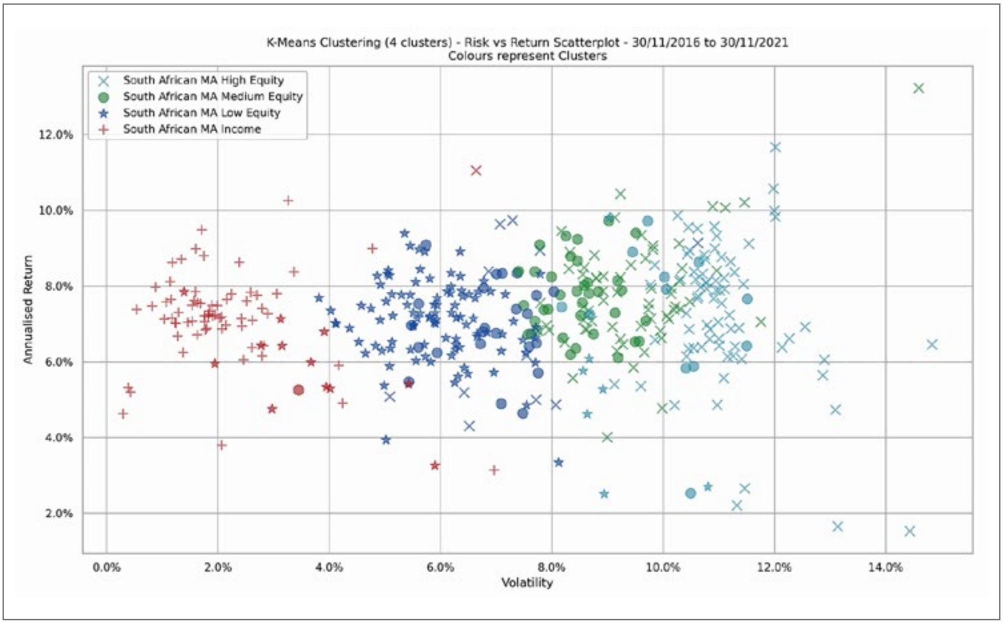


FIGURE 74 K-Means clustering (four clusters) – risk vs return – multi asset – all excl. flexible & target date

K-Means Clustering Crosstab - 30/11/2016 to 30/11/2021

Cluster	0	1	2	3	4
Category					
South African MA Flexible	5	2	12	15	3
South African MA High Equity	3	9	58	81	1
South African MA Income	0	0	0	0	67
South African MA Low Equity	1	103	0	2	12
South African MA Medium Equity	0	22	37	12	1
South African MA Target Date	0	1	3	4	0

Cluster	0	1	2	3	4
Volatility	15.0	6.3	9.1	11.0	2.3

FIGURE 75 K-Means clustering (five clusters) – multi asset – all categories

4.6.3.6 SOUTH AFRICAN – EQUITY

It is interesting to try out different parameters for the number of clusters when combining all the categories in the Equity tier. We begin by trying out the clustering first with just two clusters, and then based on the results we observe, we try the clustering again but with three clusters.

In Figure 77 the model calls for two clusters, and the demarcation is unambiguous. Cluster 1 includes all of the portfolios in the Resources category, and cluster 0 includes

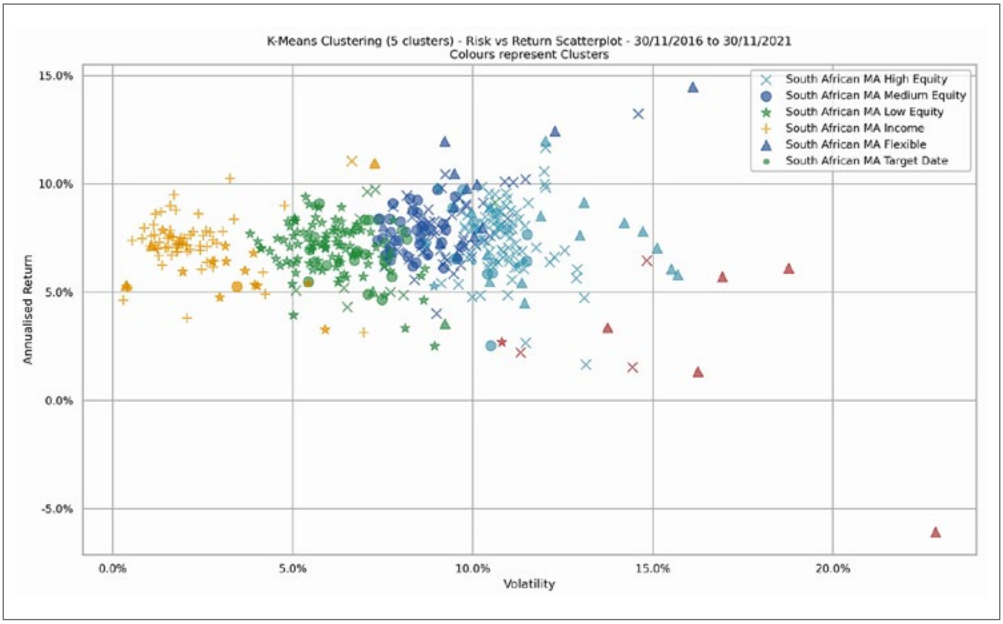


Figure 76 K-Means clustering (five clusters) – risk vs return – multi asset – all categories

everything else. That is how different Resources shares, and portfolios with large exposure to these, have behaved. The potential problem with these results is that cluster 0 could be interpreted to be homogeneous, something for which that this technique does not provide any further clarity. We therefore try the algorithm with three clusters.

In Figure 78 the model calls for three clusters instead of just two, and the results change substantially. Cluster 1, which still includes all of the portfolios from the Resources category, now includes two portfolios from the General category; clearly these portfolios

K-Means Clustering Crosstab - 30/11/2016 to 30/11/2021

Cluster	0	1
Category		
South African EQ Financial	4	0
South African EQ General	117	0
South African EQ Industrial	3	0
South African EQ Large Cap	11	0
South African EQ Mid/Small Cap	6	0
South African EQ Resources	0	6
South African EQ Unclassified	1	0

Cluster	0	1
Volatility	15.1	29.2

Figure 77 K-Means clustering (two clusters) – equity – all categories

must be resources heavy. While cluster 0 still contains most of the rest of the portfolios, cluster 2 has taken some portfolios of the other categories, including all the Mid/Small Cap category, and all the Financial category portfolios. Note the average volatility of cluster 1.

K-Means Clustering Crosstab – 30/11/2016 to 30/11/2021

Cluster	0	1	2
Category			
South African EQ Financial	0	0	4
South African EQ General	88	2	27
South African EQ Industrial	3	0	0
South African EQ Large Cap	8	0	3
South African EQ Mid/Small Cap	0	0	6
South African EQ Resources	0	6	0
South African EQ Unclassified	0	0	1

Cluster	0	1	2
Volatility	14.3	27.6	16.7

FIGURE 78 K-Means clustering (three clusters) – equity – all categories

The risk versus return scatterplot of the three cluster analysis can be seen in Figure 79. The separate clustering of the Resources category portfolios is perhaps now more obvious, as they have had much higher returns and volatility than all other portfolios, on average.

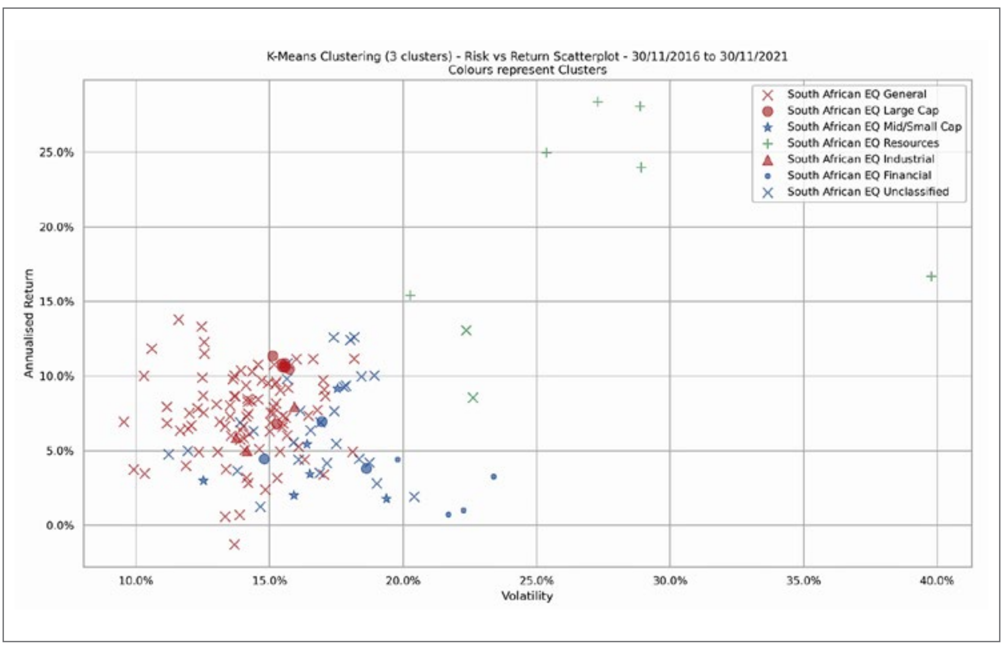


FIGURE 79 K-Means clustering (three clusters) – risk vs return – equity – all categories

4.6.3.7 SOUTH AFRICAN – INTEREST BEARING

Finally, let us consider the Interest Bearing tier, by again combing the three underlying categories to see what the clustering algorithm makes of these portfolios and categories. In Figure 80 the model calls for four clusters, but no matter how many clusters we ask for, we cannot separate out the Money Market category and Short Term category portfolios from the bulk of the Variable Term category portfolios.

Cluster 0 includes every single portfolio in both the Money Market and Short Term categories, as these are so distinct from the Variable Term category, with the exception of one portfolio in that category that gets included in this cluster. That portfolio should really not belong in the Variable Term category as we have previously argued. The average volatility of this cluster is really low at less than 1% p.a.

Cluster 1 then includes all of the portfolios in the Variable Term category that actually belong there because they are managed against the All Bond Index, and hence have similar duration to that benchmark.

Cluster 2 includes a solitary portfolio and we can guess which one this is based on our previous analysis. This portfolio is clearly distinct from all others and also does not belong in the Variable Term category.

Finally, cluster three includes the inflation-linked bond portfolio, the flexible bond portfolio, and the short-duration bond portfolio, and we have seen this pattern before.

```
K-Means Clustering Crosstab - 30/11/2016 to 30/11/2021

Cluster                          0   1  2  3
Category
South African IB Money Market   28   0  0  0
South African IB Short Term     27   0  0  0
South African IB Variable Term   1  19  1  3

Cluster      0   1    2   3
Volatility 0.6 7.8 12.7 6.1
```

FIGURE 80 K-Means clustering (four clusters) – interest bearing – all categories

Also note that although many results seem to repeat using the different methodologies, they need not do so. Many of these machine learning algorithms may use the exact same data, and even share similar methodologies to establish similarities and differences, such as distances, but the algorithms can differ substantially in how they decide to cluster portfolios. These changes can even be made through the hyperparameters discussed before within a single technique.

Figure 81 shows the analysis in the usual risk versus return scatterplot, including the coloured clusters. This again demonstrates how having access to 60 data points (five years of monthly returns) can provide so much more information than just two data points (risk and return). Clearly the blue star amongst the green stars is different when looking at monthly returns, even though it looks similar when you look at just risk and return. In future, you therefore should not expect this portfolio necessarily to behave like the others being managed against the ALBI.

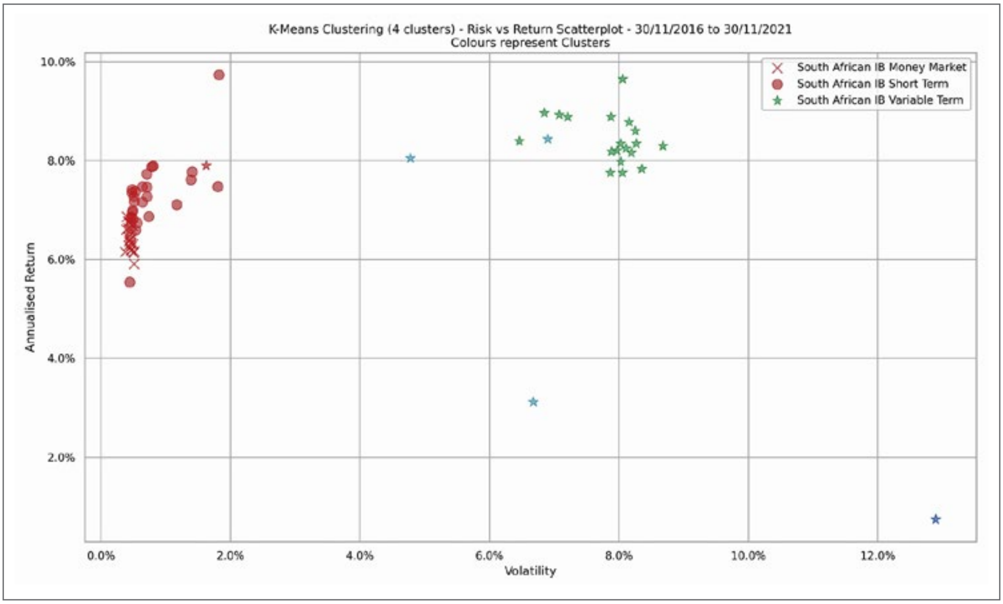


FIGURE 81 K-Means clustering (four clusters) – risk vs return – interest bearing – all categories

4.7 Principal component analysis

We have explored techniques that cluster portfolios based on historical returns, but without a visually appealing way of demonstrating these relationships except for the dendrogram which provided for increasingly distant relationships. We tried to visualise the clustering using the traditional risk versus return scatterplot, but these dimensions do not reflect the same dimensions as the dimensions in the analysis, thus sometimes providing a poor representation of why the portfolios are clustered as they are.

What if instead we could get the best of both methods, i.e. the analysis done on all 60 dimensions representing the historical returns, and a way of plotting these in a way consistent with the dimensions. This is what a PCA allows us to do, by either plotting three dimensions at a time, or two, usually the most important dimensions, i.e., those that describe the most variability in the data. Instead of having to choose the months that do this, the PCA provides the principal components that are a combination of all the months.

Because the principal components are all orthogonal to each other, you can use the traditional scatterplot to visualise this, again either in a 3D or a 2D scatterplot. A 3D scatterplot is of limited use on paper, as it has to be flattened and the 3D structure is often hard to see and appreciate. On a computer screen using an appropriate application, it may be better if you are allowed to rotate the 3D plot to look at it from different angles. You can even use a smart phone to project the 3D plot onto the real world using augmented reality, in which case you can literally walk amongst the data and 'see' the distance between the portfolios, representing their similarity or difference.

Unfortunately, this technique is only powerful if the first two or three principal components explain most of the variability in the data, as we cannot visualise more than three dimensions simultaneously. Fortunately, in most cases three dimensions is more than enough.

We therefore provide the results from the PCAs on each of the tiers, with their categories combined, in three different charts. The first is a bar plot of the variability explained by each of the first five principal components. The second is a 3D scatterplot of the data, colour-coded by category. Only a single view of this is shown making the chart of limited use, but at least the reader is able to see some of the groupings. The final chart shows 2D pairwise scatterplots of the first three principal components.

4.7.1 SOUTH AFRICAN – MULTI ASSET

For the Multi Asset tier, Figure 82 shows the amount of variability explained by each of the first five principal components graphically.

The first principal component explains over 56% of the variability. We do not unpack the linear combination that makes up this principal component, but this would typically be part of the analysis. Without too much imagination though, we can confidently say that it probably measures equity exposure, or growth asset exposure more generally, and conceptually, risk.

The second principal component explains 14% of the variability, taking the combined total for the first two principal components to 70%. Appreciate the power of this technique to be able to summarise most of the variability in 60 dimensional data with just two dimensions. The third principal component only explains 3.7% of the variability, and each subsequent component even less, so is not really worth looking at any further.

If it is easy to deduce what the first principal component represents, the second is a lot harder. It could be global exposure, or property exposure. Always first look at the most volatile asset classes (in multi-asset class portfolios) to dominate the variability in the returns observed, especially when looking at risky portfolios.

Figure 83 shows the 3D scatterplot for the Multi Asset categories, where the categories are colour coded, and the axes are the first three principal components. You can see the tight grouping of the Income category portfolios whereas the High Equity category portfolios seem to be fairly spread out. It is unfortunately hard to view exactly where each point sits in 3D space, so on paper this chart is of limited use, but remember the tool for when you have the opportunity to manipulate a 3D plot on a computer or a phone.

To add to our understanding of hierarchical clustering techniques, imagine that these first

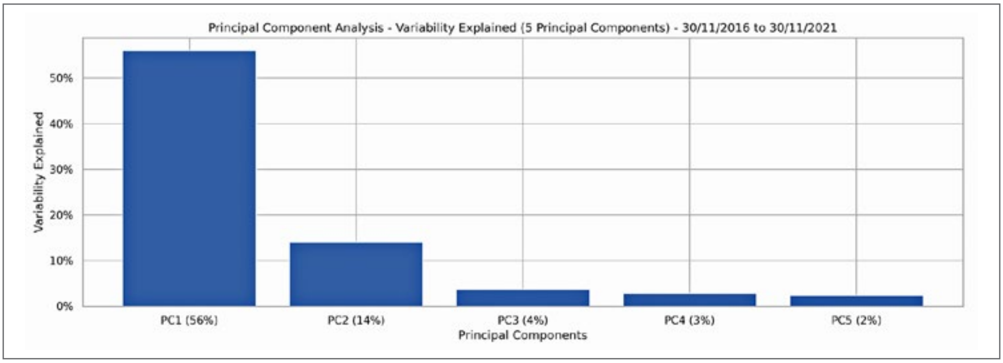


Figure 82 PCA – variability explained – multi asset

three principal components explain all of the variabilities in the returns. The clustering algorithm simply looks to add the two closest points into a single cluster, with the distance between them as the measure of similarity.

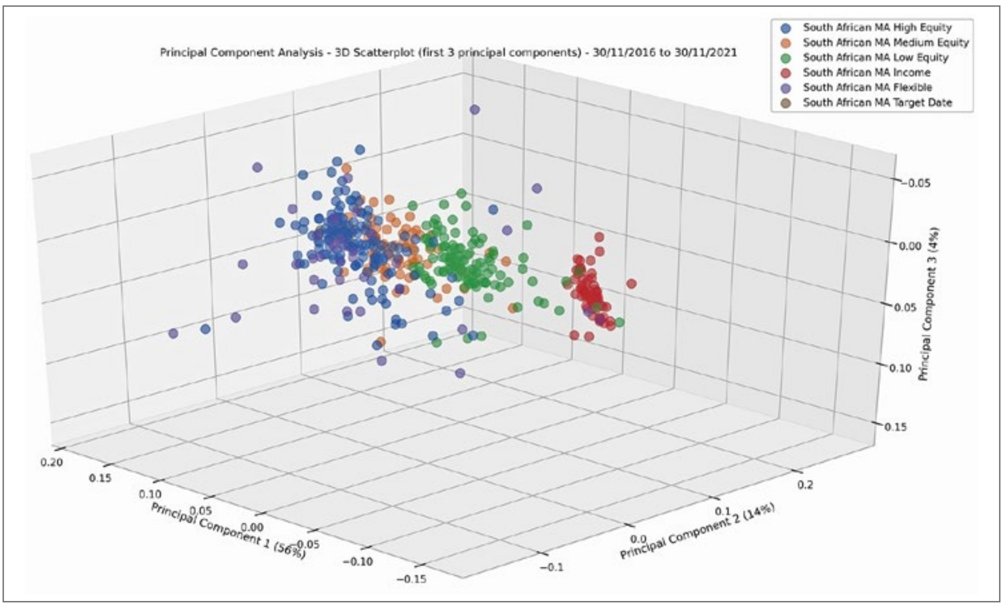


Figure 83 PCA – 3D scatterplot – multi asset

Figure 84 breaks down the 3D plot by looking at just two axes at the same time as a more commonly used 2D scatterplot. The first column of this pair plot compares the first principal component (on the x-axis) to the second principal component in row two, and the third in row three, using a scatterplot, with each dot representing a portfolio.

The diagonal of the matrix represents a histogram for each principal component, and as the first principal component explains 56% of the variability of the returns, it is quite informative. The Income category portfolio is grouped on the far left (in purple), and the Flexible category portfolios are spread out all the way to the far right (in pink).

Although the charts in row 2, column 1, and row 1, column 2, are essentially the same, with the axes transposed, we prefer looking at the chart with the primary principal component on the x-axis, which is the chart in row 2, column 1. This chart resembles the risk versus return chart in many ways, but mainly along the x-axis.

One of the best ways to understand what the second principal component could represent, would be to consider what the two portfolios at each of the extremes of this principal component looked like, perhaps by looking at the asset allocation or portfolio holdings. A quick peek at this data reveals that one of the portfolios has no offshore exposure, while the other sits at a maximum exposure of 30%.

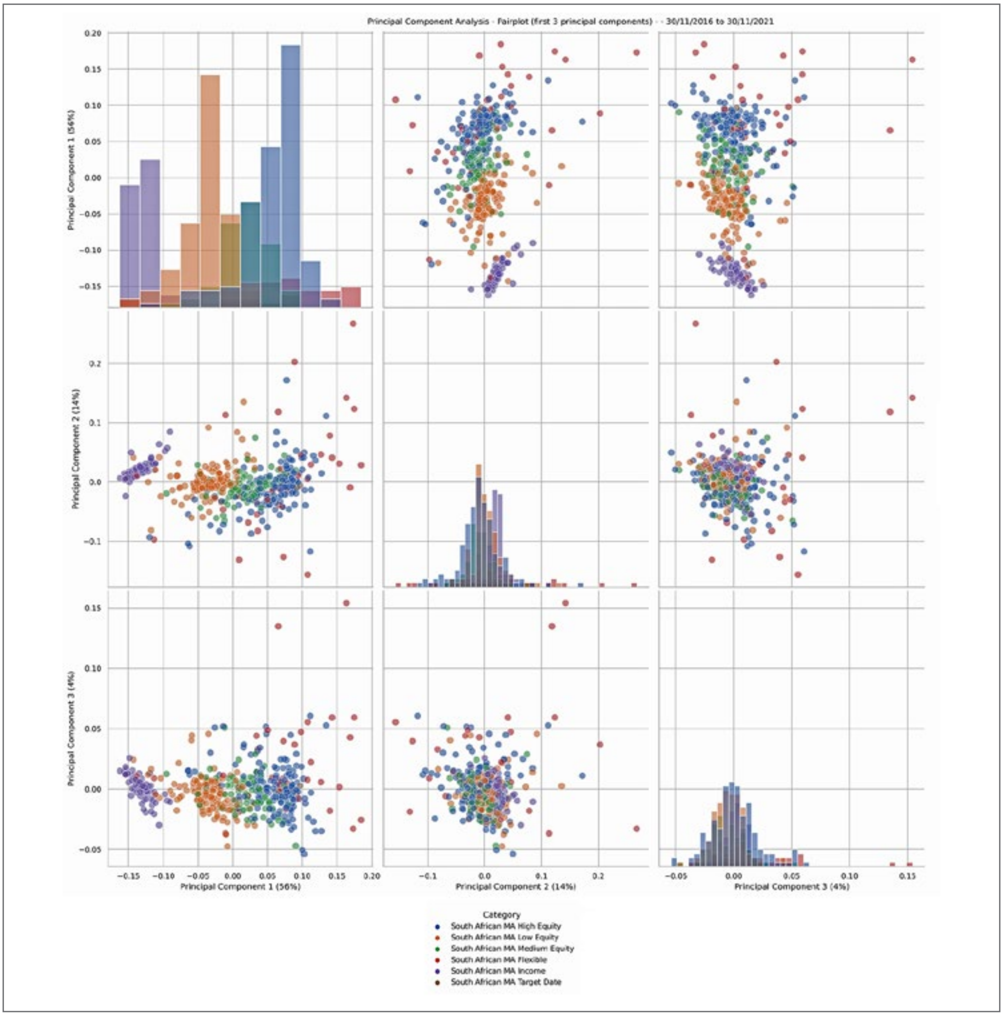


Figure 84 PCA – 2D pair plot – multi asset

4.7.2 SOUTH AFRICAN – EQUITY

For the Equity tier, Figure 85 shows the amount of variability explained by each of the first five principal components. In this case, the first principal component explains only 32% of the variability, which we can make sense of by considering that we are essentially comparing equity portfolios, which are going to be very similar in that they all invest in equities. The second principal component explains 20%, and the third an additional 10%, for a total of 62% for the first three principal components. This is not as much as we had for the Multi Asset tier, but still not a bad number.

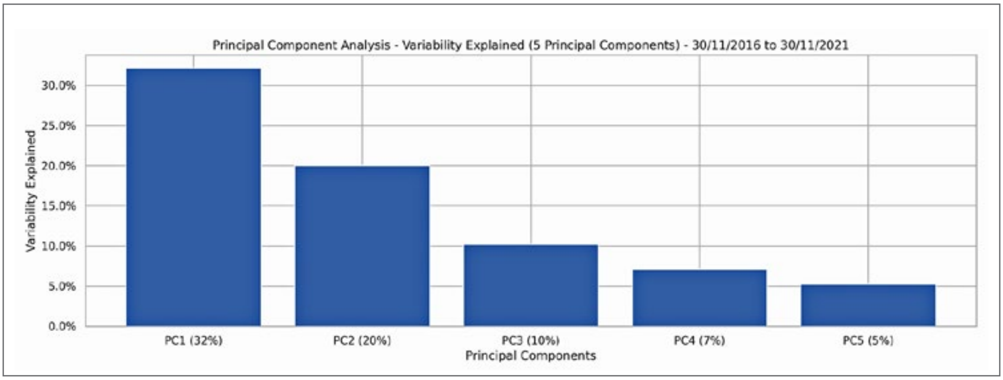


FIGURE 85 PCA – variability explained – equity

Figure 86 shows the 3D scatterplot for the Equity tier, using the first three principal components. Again, it is impossible to judge the relative position of the portfolios on this 2D representation of the 3D plot.

Figure 87 shows the pair plot for the Equity tier, using the first three principal components, and this provides more insights into the possible interpretation of each of these principal components.

Focusing first on the chart in row 2 of column 1, we can see some lonely points far off to the right along the first principal component. These portfolios (purple) appear to be from the Resources category, so the first principal component probably represents resources exposure. Having looked at these charts for many years, this result is somewhat surprising, as the first principal component is typically total equity exposure. Note that this category is allowed a minimum of 80% total equity exposure. However, given the moves in resources shares over the last five years this result is not completely surprising.

As additional evidence for our contention, note how the Financial category portfolios (green) sit at the opposite extreme of the first principal component, with no resources exposure. The General category portfolios (orange) sit between the extremes as expected, as they have some resources exposure. Note how two of the orange dots sit further to the right, and turn back to the K-Means clustering section where we found two of the General category portfolios group with the Resources category portfolios into a single cluster.

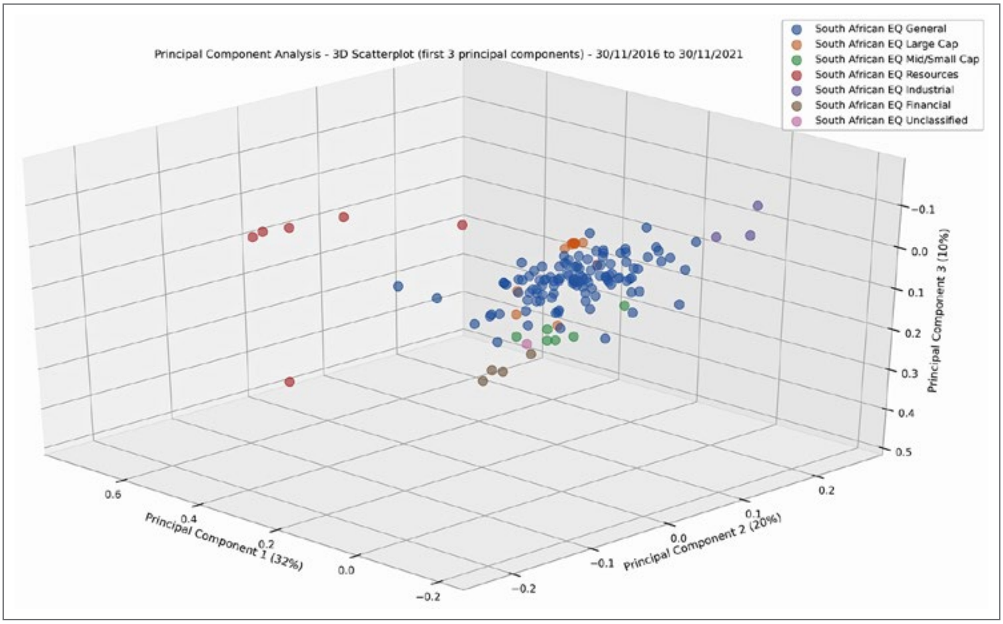


Figure 86 PCA – 3D scatterplot – equity

This then appears to downgrade the total equity exposure to the second principal component, and it probably explains why this principal component explains 20% of the variability, which is still a substantial number.

For the third principal component, we need to consider the portfolios sitting at the extremes of this dimension, which appears to include the Resources category portfolios. You may find that the relationships appear because of a third variable e.g. offshore exposure. It may be that Resources companies often act like rand hedges, because commodities are generally priced in US dollars, and that these portfolios are reacting like others that have more foreign exposure. This is just a conjecture that would need to be investigated.

4.7.3 SOUTH AFRICAN – INTEREST BEARING

For the Interest Bearing tier, Figure 88 shows the amount of variability explained by each of the first five principal components. In this case, the first principal component explains a whopping 84% of the variability, which may be obvious from the previous analysis that clearly showed duration as the dominant difference between portfolios in the categories across this tier. The second principal component explains 10%, and the third an additional 1%, for a total of 95% for the first three principal components. You are unlikely to ever see higher numbers than these.

Figure 89 shows the 3D scatterplot, with the Money Market category portfolios all on top of each other. Figure 90 shows the pair plot for the Interest Bearing tier, again using the first three principal components.

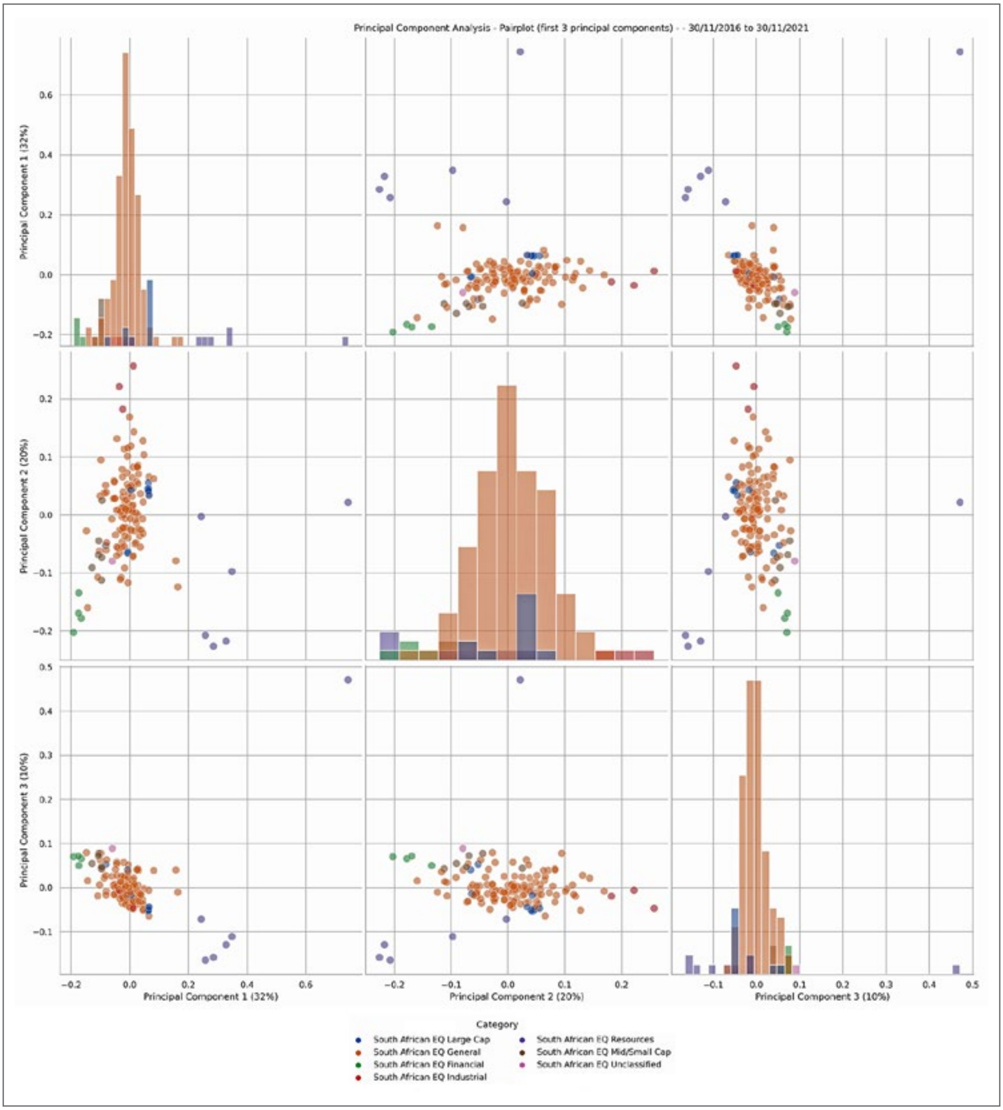


Figure 87 PCA – 2D pair plot – equity

The first principal component undoubtedly represents duration as we know that Money Market category portfolios contain essentially none, and the Short Term category portfolios contain very little, whereas the Variable Term category portfolios contain lots, along with their All Bond Index.

Once you know which portfolio sits at the extreme of the second principal component, you are more likely to recognise what this component represents, and in this case the portfolio is the Colourfield BCI Income portfolio, so the second principal component represents exposure to inflation-linked bonds. This is confirmed by the point sitting mid-

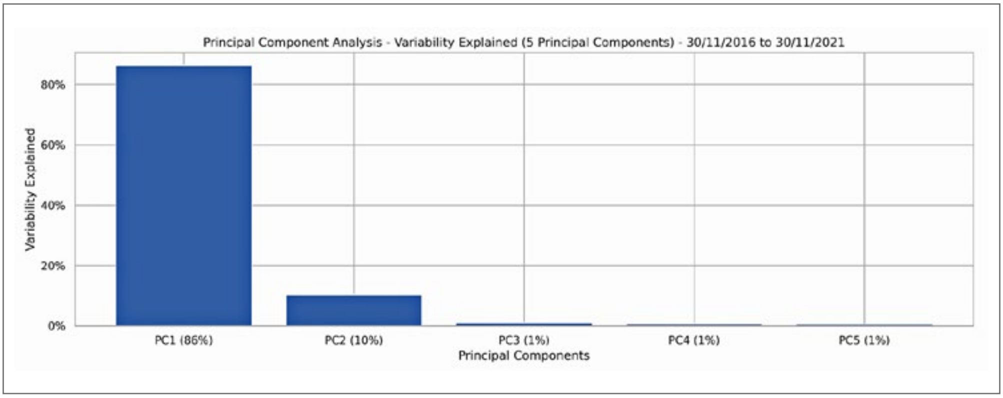


FIGURE 88 PCA – variability explained – interest bearing

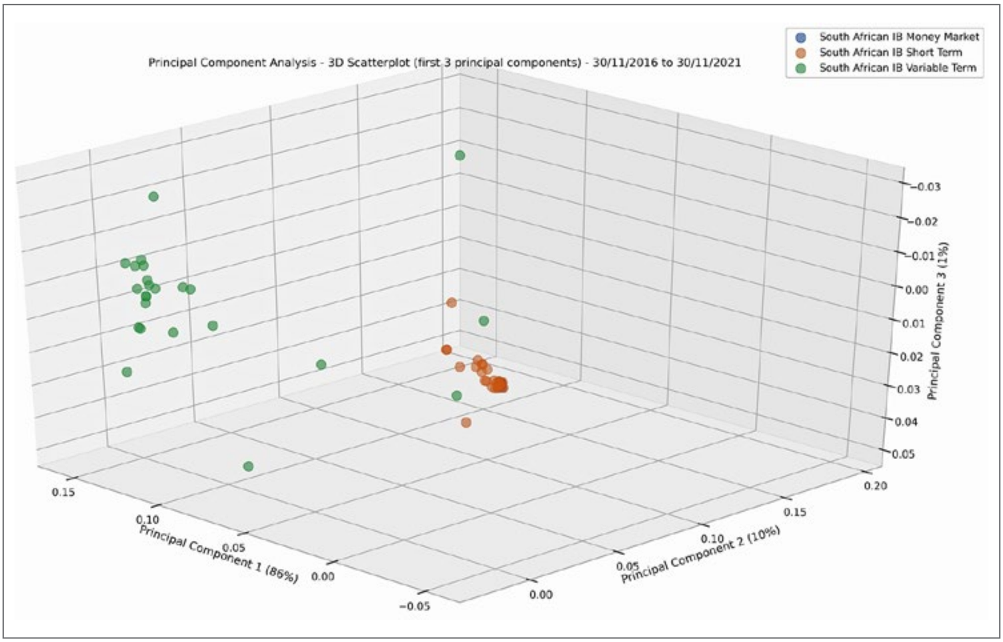


FIGURE 89 PCA – 3D scatterplot – interest bearing

point along this dimension, which is an inflation-linked bond tracker portfolio.

We are using certain literal descriptors of what the principal components represent, but these representations could be more abstract or conceptual. For example, the first principal component could actually represent interest rate risk, which is often proxied by duration. The second principal component may not represent exposure to inflation-linked bonds more generally, but rather real interest rate risk, again often proxied by inflation-linked bonds. The third principal component, although not relatively important, may represent credit risk.

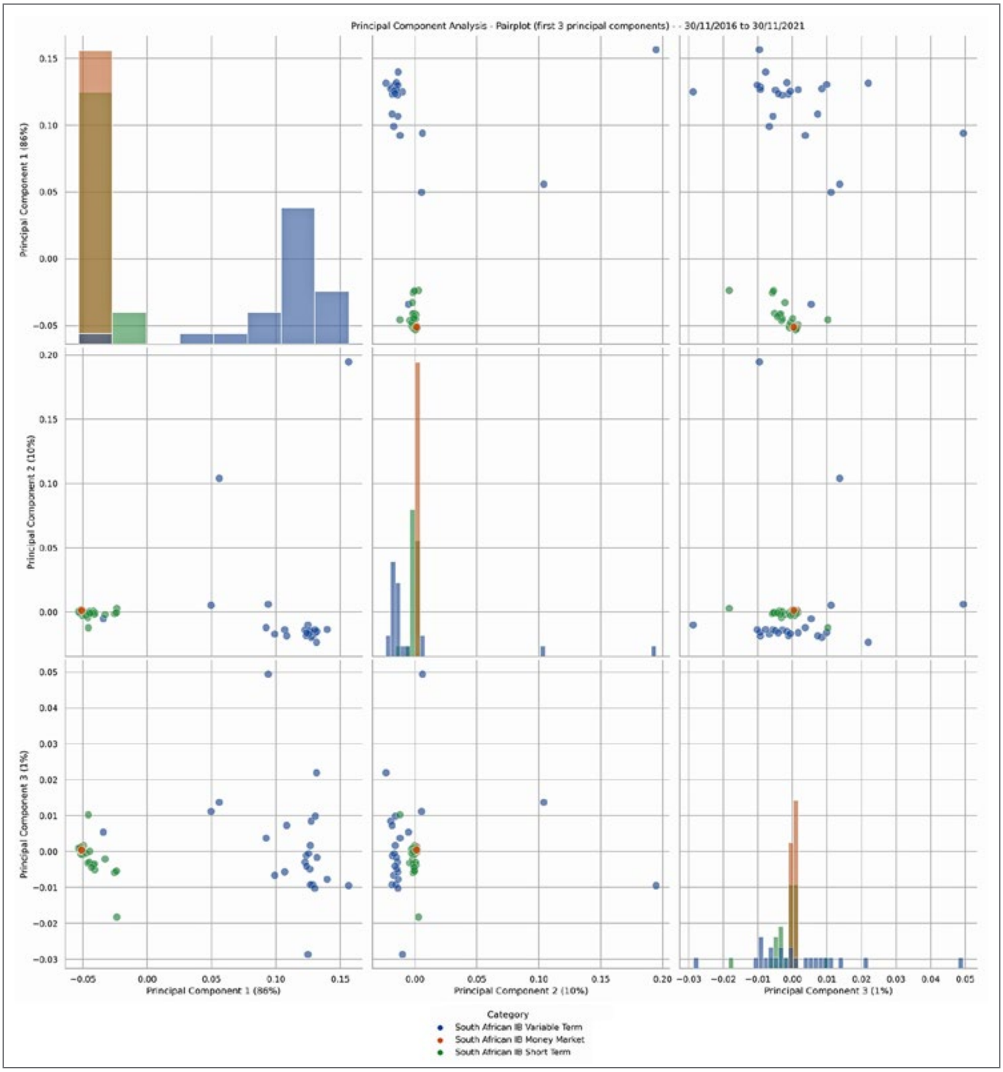


FIGURE 90 PCA – 2D pair plot – interest bearing

4.7.4 SOUTH AFRICAN – REAL ESTATE

For the Real Estate tier, Figure 91 shows the amount of variability explained by each of the first five principal components. In this case, the first principal component explains 43% of the variability, which is going to be really interesting to understand, as unlike other tiers, this one should be fairly homogeneous. The second principal component explains 18%, and the third an additional 12%, for a total of 73% for the first three principal components. In this case, the third principal component is almost as important as the second, so we should probably dig a little deeper into both of these.

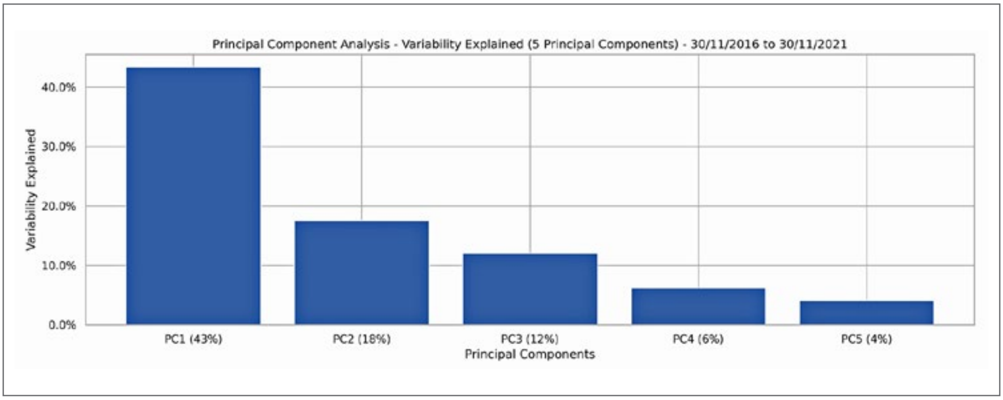


Figure 91 PCA – variability explained – real estate

Figure 92 shows the 3D scatterplot, with no obvious pattern with the exception of a few points at extremes.

Figure 93 shows the pair plot for the Real Estate tier, using the first three principal components.

The first principal component probably represents the exposure to property, with portfolios on the far left including some of the property index trackers, and the ones on the right doing something very different as this category includes some very active managers.

The second principal component appears to represent offshore exposure; portfolios at

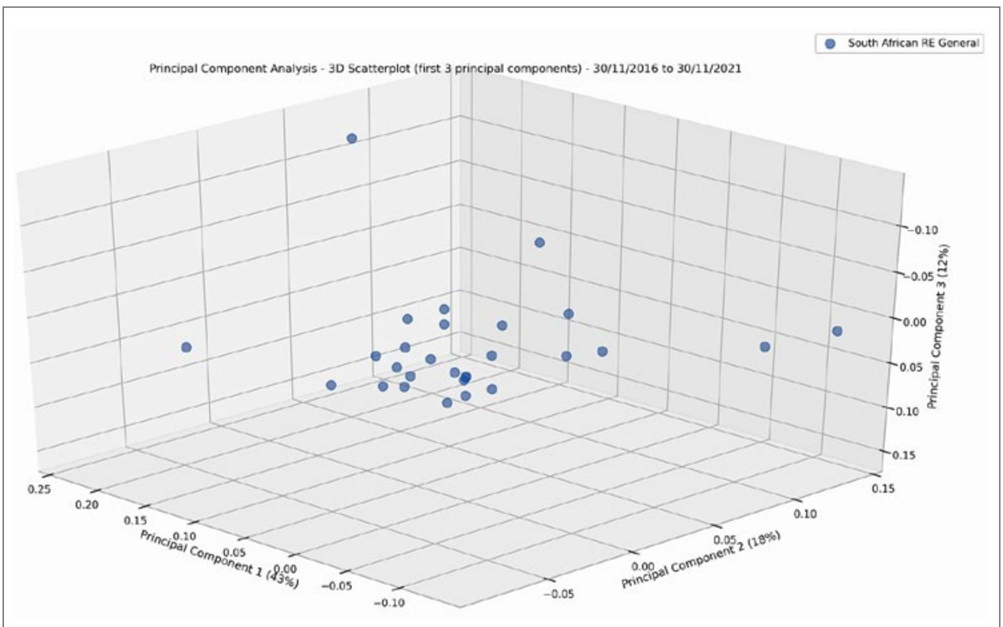


Figure 92 PCA – 3D scatterplot – real estate

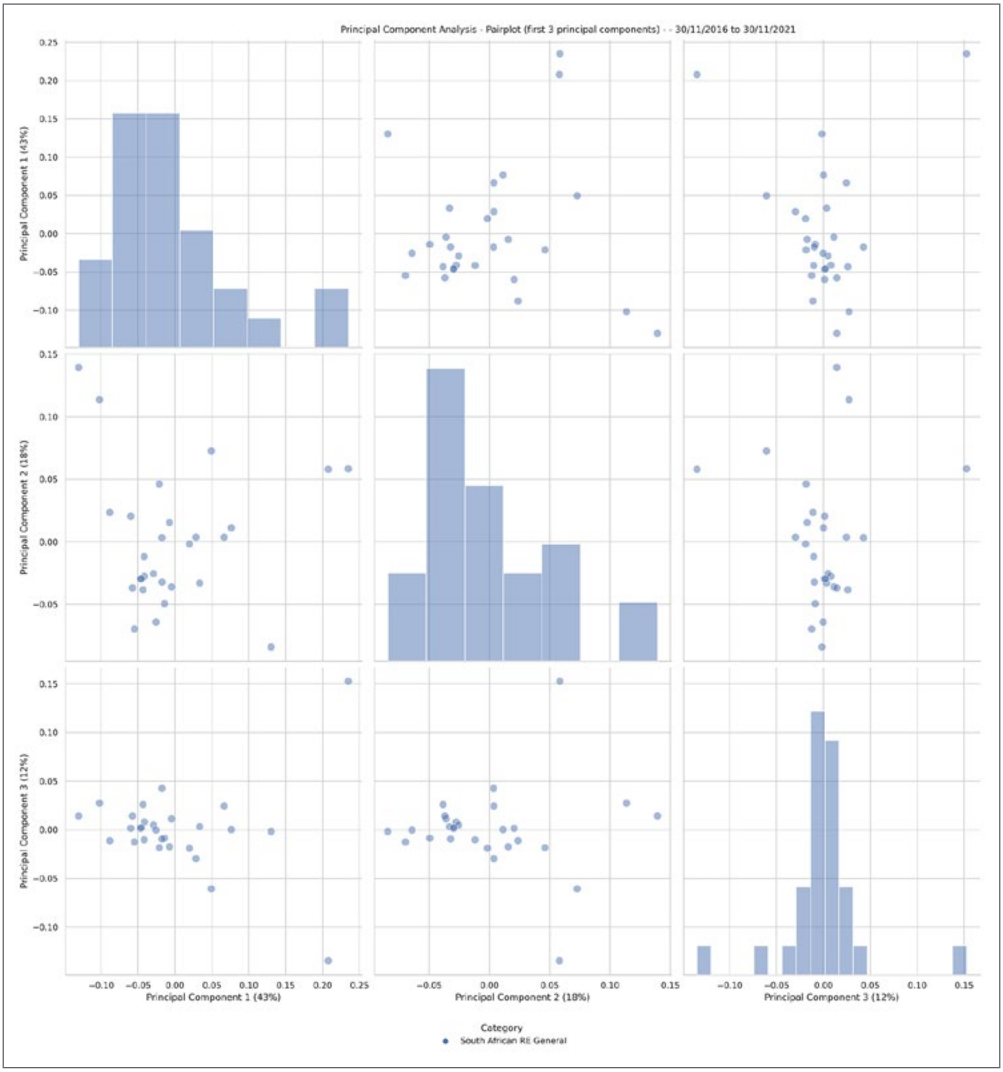


FIGURE 93 PCA – 2D pair plot – real estate

one extreme have exposure while at the other extreme portfolios have none. It is perhaps not obvious that many property shares may be classified as local, but they have some or a lot of exposure to offshore property.

The third principal component could represent exposure to the smaller cap part of the market, as the property sector has some really large property shares that dominate the benchmarks.

This analysis should be taken as provisional thoughts, which would need deeper analysis to confirm.

4.8 K nearest neighbours

There are three sections to each figure in the analysis which follows. We focus on the first two.

The first in the upper left corner of the image is a square matrix with as many rows and columns as there are categories of portfolios being analysed by the K nearest neighbours (KNN) model. This is called a confusion matrix. The order of both the rows and columns is the same, and the same as the rows in the next section, which has the category names. The number in each cell of the confusion matrix corresponds to the predictions made by the model.

In Figure 94, the number in the first cell for example (row one, column one), says that the model predicts that two portfolios in the Flexible category are predicted to fall into the Flexible category. This is of all 37 portfolios that are actually in that category. The rest of the portfolios in the Flexible category are predicted to fall into the other categories used in the analysis.

The model therefore predicts that 29 portfolios should actually be in the High Equity category, two in the Income category, two in the Low Equity category, two in the Medium Equity category, and zero in the Target Date category.

The next two sections are the classification report, and we focus only on some of the numbers in this report, as the rest are not really that important for this exercise. Precision measures the proportion of portfolios that are predicted to fall into a specific category, that are actually classified in that category. Recall, on the other hand, measures how many portfolios in a specific category are predicted to fall into that category.

So, for example, the Flexible category has 100% precision, because both portfolios that are predicted to be in the Flexible category, are classified as being in the Flexible category, and not from some other category. The Flexible category however has a very low recall because only two of the 37 portfolios (5%) are predicted to fall into the category with which they are classified.

4.8.1 SOUTH AFRICAN – MULTI ASSET

For the South African Multi Asset tier, we have included all of the categories, of which there are six, and chosen a value of 15 for the number of neighbours parameter (k=15). This is a somewhat arbitrary number that would typically be chosen by running the training and testing the model on out-of-sample data to see what provided the best fit for the model. We are however not trying to provide the best fit to the data here, but rather using the model to look for potentially misclassified portfolios, or portfolios that may otherwise be better classified in another category.

Although the results can change when changing this hyperparameter (this is what k is called), the results will be fairly robust to changes. If the value is too low (say five), then the Target Date portfolios will be classified in the same category that they are in, which is not the result that we are looking for. This is known as overfitting the data.

Figure 94 shows the results of applying the KNN classifier to the portfolios above, and shows some interesting results that are consistent to what we have seen before. Focusing on the second row, the High Equity category, we can see that although most portfolios in this category (125 out of 152) are predicted to fall into this category, you have some that are predicted to belong in other categories as we have seen before, specifically nine in the Low Equity category, and 18 in the Medium Equity category.

The Income category has 100% recall with all of the portfolios predicted to fall into the Income category. The Target Date portfolios are predicted to fall into the other categories, making this category obsolete (four in the High Equity category, three in the Medium Equity category, and one in the Low Equity category). You would probably find this consistent with the asset allocation of those portfolios.

The Medium Equity category provides some useful insights about both the portfolios and the classification standard, but also about the limitations of the modelling. Recognise that a portfolio in the High Equity category can be managed the same as a portfolio in the Medium Equity category, because although the category has a higher total equity exposure limit, this is not a requirement to have more total equity exposure. A portfolio in the Medium Equity category can however not be managed the same as a portfolio in the High Equity category, if that portfolio has a higher total equity exposure than the 60% limit that applies to the Medium Equity category.

The model therefore 'learns' that many less risky High Equity portfolios, perhaps with total equity exposure less than 60%, are indeed High Equity portfolios, and the classification will 'learn' to classify them as such. This now creates a problem for Medium Equity portfolios, as they may sit a lot closer to many of these High Equity portfolios with lower total equity exposure, and hence be predicted to fall into the High Equity category.

```
KNN Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[  2  29   2   2   2   0]
 [  0 125   0   9  18   0]
 [  0   0  67   0   0   0]
 [  0   0   6 111   1   0]
 [  0  16   0  15  41   0]
 [  0   4   0   1   3   0]]
                               precision    recall  f1-score   support

     South African MA Flexible       1.00      0.05      0.10        37
  South African MA High Equity       0.72      0.82      0.77       152
       South African MA Income       0.89      1.00      0.94        67
   South African MA Low Equity       0.80      0.94      0.87       118
South African MA Medium Equity       0.63      0.57      0.60        72
  South African MA Target Date       0.00      0.00      0.00         8

                      accuracy                           0.76       454
                     macro avg       0.67      0.56      0.55       454
                  weighted avg       0.76      0.76      0.72       454
```

FIGURE 94 KNN classifier (k=15) – multi asset

This is probably what we are observing in the confusion matrix with the Medium Equity category having 16 portfolios predicted to fall into the High Equity category, as the High Equity category probably has many lower total equity exposure portfolios. Although this is one possible reason for the results, there are others, including exposure to property, and drift above the 60% limit which is allowed. These are covered in more detail below.

4.8.2 SOUTH AFRICAN – EQUITY

The results for the South African Equity tier are also fascinating as can be seen in Figure 95. The selection of k in this analysis is more important because so many of the categories have such low numbers of portfolios in them. A high k of 15, as we have here could lead to all equity portfolios being predicted to fall into the General category that includes 117 portfolios, and we see that happen with the exception of five portfolios (three in Large Cap and two in Resources).

```
KNN Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[  0   4   0   0   0   0   0]
 [  0 117   0   0   0   0   0]
 [  0   3   0   0   0   0   0]
 [  0   8   0   3   0   0   0]
 [  0   6   0   0   0   0   0]
 [  0   4   0   2   0   0   0]
 [  0   1   0   0   0   0   0]]
                               precision    recall  f1-score   support

    South African EQ Financial       0.00      0.00      0.00         4
      South African EQ General       0.82      1.00      0.90       117
   South African EQ Industrial       0.00      0.00      0.00         3
     South African EQ Large Cap       0.60      0.27      0.37        11
South African EQ Mid/Small Cap       0.00      0.00      0.00         6
     South African EQ Resources       0.00      0.00      0.00         6
 South African EQ Unclassified       0.00      0.00      0.00         1

                      accuracy                           0.81       148
                     macro avg       0.20      0.18      0.18       148
                  weighted avg       0.69      0.81      0.74       148
```

FIGURE 95 KNN classifier (k=15) – equity

4.8.3 SOUTH AFRICAN – INTEREST BEARING

The results for the South African Interest Bearing tier are also interesting as can be seen in Figure 96. The selection of k in this example is less important as each of the categories has about an equal number of portfolios.

The precision of 100% for the Variable Term category should not be surprising, as portfolios in the other two categories cannot hold long duration assets. Neither is the 100% recall for the Money Market portfolios because they are unique in terms of their duration and maturity limits.

We should also not be surprised with the eight portfolios in the Short Term category that are predicted to fall into the Money Market category, as we have made this point previously using other techniques, or that one portfolio in the Variable Term category is predicted to fall into the Short Term category, although we may have expected more than one.

The inflation-linked bond portfolios unfortunately have no other category to go into, and are therefore predicted to belong to the Variable Term category. Ideally, these should be moved into their own category, if there were enough of them, or rather into an Unclassified category, given that there are so few of them.

```
KNN Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[28  0  0]
 [ 8 19  0]
 [ 0  1 23]]
                                precision    recall  f1-score   support

 South African IB Money Market       0.78      1.00      0.88        28
   South African IB Short Term       0.95      0.70      0.81        27
South African IB Variable Term       1.00      0.96      0.98        24

                      accuracy                           0.89        79
                     macro avg       0.91      0.89      0.89        79
                  weighted avg       0.90      0.89      0.88        79
```

FIGURE 96 KNN classifier (k=10) – interest bearing

4.8.4 MIXING TIERS

It may be apparent to some readers that some of the analysis in this paper indicates that the South African Multi Asset Income category may have some overlap with the Money Market and Short Term categories in the Interest Bearing tier. This would have been seen, for example, in the low volatility numbers for some of these portfolios.

We have therefore applied the KNN modelling to these three categories, and the results can be seen in Figure 97. As suspected, six portfolios in the Income category are predicted to fall into the Money Market category, and nine into the Short Term category. This is not surprising given the overlap between these categories; although the Income category allows equity and property exposure, there is no requirement to hold these assets.

What is perhaps more surprising is that three portfolios in the Short Term category are predicted to fall into the Income category, where the limits of the Short Term category would not allow this. This merely highlights the same issue raised before around the overlap with the portfolios in the Income category, causing the category to include many portfolios that behave like portfolios in the Short Term category.

We could also include the Equity General category into the Multi Asset tier, because we suspect that some portfolios in the Flexible category are managed like equity portfolios. In fact, we have previously made the point that the equity category is in essence also a multi asset category as only 80% is required to be invested in equity.

```
KNN Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[28  0  0]
 [12 12  3]
 [ 6  9 52]]
                           precision    recall  f1-score   support

South African IB Money Market       0.61      1.00      0.76        28
  South African IB Short Term       0.57      0.44      0.50        27
      South African MA Income       0.95      0.78      0.85        67

                     accuracy                           0.75       122
                    macro avg       0.71      0.74      0.70       122
                 weighted avg       0.79      0.75      0.75       122
```

FIGURE 97 KNN classifier (k=15) – interest bearing – money market, short term & multi asset – income

Figure 98 shows the results of this analysis. As suspected seven portfolios in the Flexible category are predicted to fall into the Equity General category. Surprisingly, 26 portfolios in the Equity General category are predicted to fall into the Multi Asset High Equity category; we remind the reader that there is no overlap between these categories, except possibly through drift. The reasons for this are beyond the scope of this paper, but perhaps we can hint at a possible explanation by considering that certain shares are sensitive to interest rates (specifically banks and property companies).

For completeness, we include all of the South African categories into a single model to see the results – see in Figure 99. Two standouts would be the 100% recall for the Real Estate

```
KNN Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[ 91   0  26   0   0   0   0]
 [  7   0  24   2   2   2   0]
 [  1   0 124   0   9  18   0]
 [  0   0   0  67   0   0   0]
 [  0   0   0   6 111   1   0]
 [  0   0  16   0  15  41   0]
 [  0   0   4   0   1   3   0]]
                            precision    recall  f1-score   support

      South African EQ General       0.92      0.78      0.84       117
     South African MA Flexible       0.00      0.00      0.00        37
  South African MA High Equity       0.64      0.82      0.72       152
       South African MA Income       0.89      1.00      0.94        67
   South African MA Low Equity       0.80      0.94      0.87       118
South African MA Medium Equity       0.63      0.57      0.60        72
  South African MA Target Date       0.00      0.00      0.00         8

                      accuracy                           0.76       571
                     macro avg       0.56      0.59      0.57       571
                  weighted avg       0.71      0.76      0.73       571
```

FIGURE 98 KNN classifier (k=15) – multi asset, & equity – general

```
KNN Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[  0   4   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0  92   0   0   0   0   0   0   0   0   0  25   0   0   0   0   0]
 [  0   3   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   8   0   3   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   6   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   4   0   2   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0]
 [  0   0   0   0   0   0   0  28   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0  12  12   0   0   0   3   0   0   0   0]
 [  0   0   0   0   0   0   0   0   1  21   0   0   2   0   0   0   0]
 [  0   6   0   0   0   0   0   1   1   0   0  23   0   2   2   0   2]
 [  0   1   0   0   0   0   0   0   0   0   0 124   0   9  18   0   0]
 [  0   0   0   0   0   0   0   6   9   0   0   0  52   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   6 111   1   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0  16   0  15  41   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   4   0   1   3   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  28]]
                                  precision    recall  f1-score   support

     South African EQ Financial       0.00      0.00      0.00         4
       South African EQ General       0.74      0.79      0.76       117
    South African EQ Industrial       0.00      0.00      0.00         3
     South African EQ Large Cap       0.60      0.27      0.37        11
South African EQ Mid/Small Cap       0.00      0.00      0.00         6
     South African EQ Resources       0.00      0.00      0.00         6
  South African EQ Unclassified       0.00      0.00      0.00         1
  South African IB Money Market       0.60      1.00      0.75        28
    South African IB Short Term       0.52      0.44      0.48        27
 South African IB Variable Term       1.00      0.88      0.93        24
      South African MA Flexible       0.00      0.00      0.00        37
   South African MA High Equity       0.64      0.82      0.72       152
        South African MA Income       0.83      0.78      0.80        67
    South African MA Low Equity       0.80      0.94      0.87       118
 South African MA Medium Equity       0.63      0.57      0.60        72
   South African MA Target Date       0.00      0.00      0.00         8
       South African RE General       0.93      1.00      0.97        28

                       accuracy                           0.72       709
                      macro avg       0.43      0.44      0.43       709
                   weighted avg       0.66      0.72      0.68       709
```

Figure 99 KNN classifier (k=15) – all tiers & categories

and the Money Market categories, as these portfolios typically do not fit into any other category, unless those categories have many portfolios that fit into them, which luckily is not the case. The highest precision at 93% is the Real Estate category with two portfolios from the Flexible category predicted to fall into this category. The reason is simply that these are two flexible property portfolios.

4.9 Decision trees

The results for the Decision Trees Classifier machine learning technique are identical to those from the KNN classifier, the confusion matrix and the classification report, so we do not go through an explanation of this again (refer to Section 4.8, KNN, for an explanation).

4.9.1 SOUTH AFRICAN – MULTI ASSET

For the South African Multi Asset tier, the results are fairly similar to what we saw with the KNN classifier, and even identical in certain instances, for example, the recall for the Income category. There are some notable differences though, and the Flexible category stands out. The results can be seen in Figure 100.

With KNN, most of the Flexible category portfolios were predicted to fall into the High Equity category, and none of the High Equity portfolios were predicted to fall into the Flexible category. With the Decision Tree classifier, most of the Flexible category portfolios are predicted to remain where they are, and 12 of the High Equity category portfolios are predicted to fall into the Flexible category.

```
Decision Tree Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[ 22   9   2   1   3   0]
 [  9 135   0   2   6   0]
 [  0   0  67   0   0   0]
 [  0   1   4 107   6   0]
 [  0  13   0   4  55   0]
 [  0   6   0   0   2   0]]
                                precision    recall  f1-score   support

     South African MA Flexible       0.71      0.59      0.65        37
  South African MA High Equity       0.82      0.89      0.85       152
       South African MA Income       0.92      1.00      0.96        67
   South African MA Low Equity       0.94      0.91      0.92       118
South African MA Medium Equity       0.76      0.76      0.76        72
  South African MA Target Date       0.00      0.00      0.00         8

                      accuracy                           0.85       454
                     macro avg       0.69      0.69      0.69       454
                  weighted avg       0.83      0.85      0.84       454
```

FIGURE 100 Decision tree classifier (min sample leaf = 10) – multi asset

Many of the portfolios in the Resources category also are predicted to fall into the Large Cap category, which is perhaps not too surprising, although we previously noted just how different some of these portfolios are from all other equity portfolios. This is surely just a function of the hyperparameter chosen; choosing a different value for this parameter or another hyperparameter would likely change the results substantially.

Figure 101 shows the actual decision tree for the classification of the Multi Asset tier. This is explained in more detail below, but it shows the month and value on which decisions are made at each node in the tree.

4.9.2 SOUTH AFRICAN – EQUITY

For the South African Equity tier, the results look different to the KNN classifier, as can be seen in Figure 102. The most surprising result is that the Financial category gets many predictions, whereas it had none using the KNN classifier, partly because it had so few portfolios (four). It includes four portfolios from the General category (not surprising),

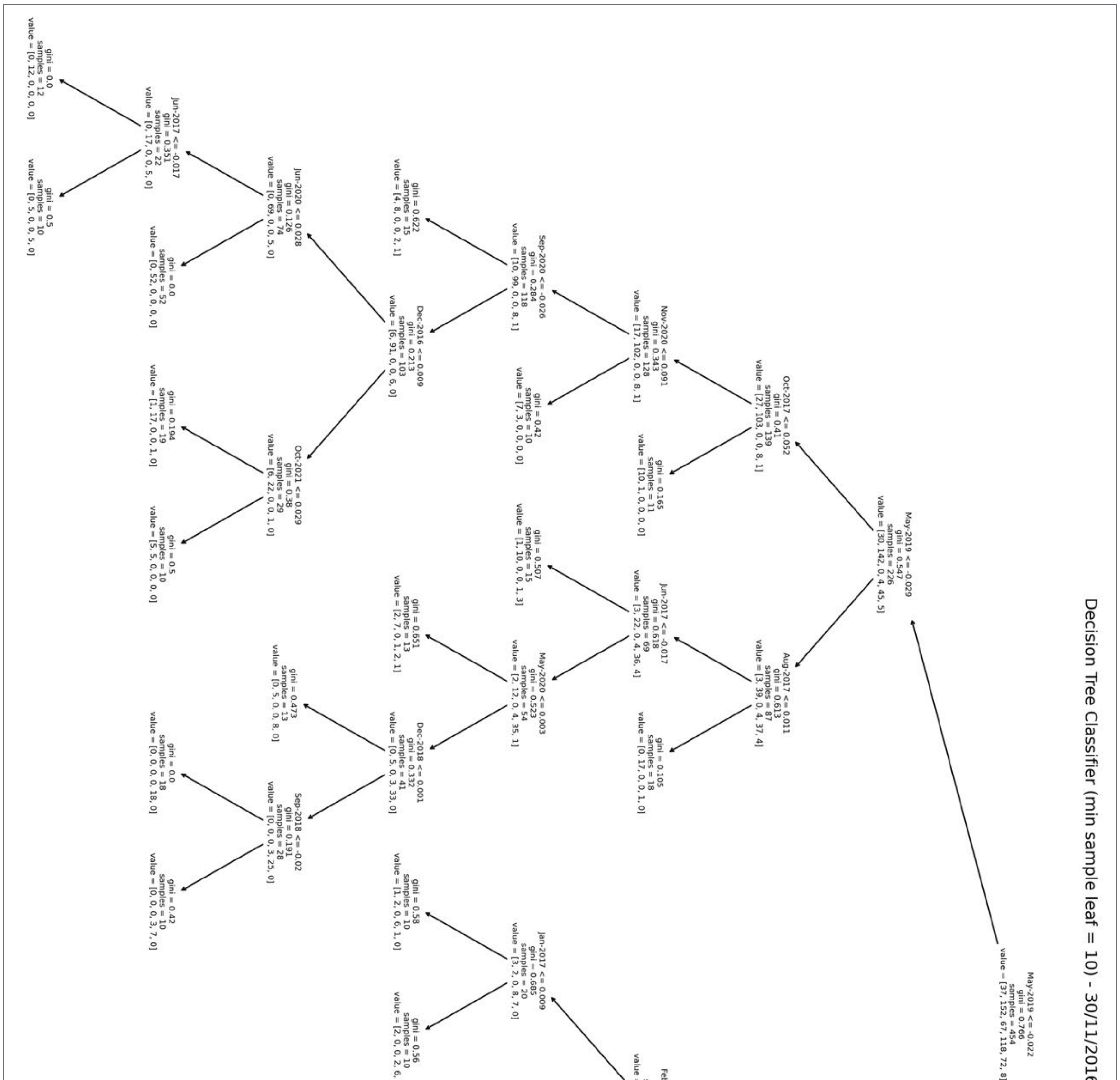


FIGURE 101 Decision tree classifier – decision tree – (min sample leaf = 10) – multi asset

three from the Industrial category (very surprising), and four from the Mid/Small Cap category (also not surprising).

Many of the portfolios in the Resources category, also get predicted to fall into the Large Cap category, which is perhaps not too surprising, although we did previously note just how different some of these portfolios are from all other equity portfolios. This is surely just a function of the hyperparameter chosen, and choosing a different value for this parameter, or choosing another hyperparameter, would likely change the results substantially.

```
Decision Tree Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[  4   0   0   0   0   0   0]
 [  4 112   0   1   0   0   0]
 [  3   0   0   0   0   0   0]
 [  0   4   0   7   0   0   0]
 [  4   2   0   0   0   0   0]
 [  0   1   0   5   0   0   0]
 [  0   1   0   0   0   0   0]]
                                precision    recall  f1-score   support

    South African EQ Financial       0.27      1.00      0.42         4
      South African EQ General       0.93      0.96      0.95       117
   South African EQ Industrial       0.00      0.00      0.00         3
    South African EQ Large Cap       0.54      0.64      0.58        11
South African EQ Mid/Small Cap       0.00      0.00      0.00         6
    South African EQ Resources       0.00      0.00      0.00         6
 South African EQ Unclassified       0.00      0.00      0.00         1

                      accuracy                           0.83       148
                     macro avg       0.25      0.37      0.28       148
                  weighted avg       0.79      0.83      0.80       148
```

FIGURE 102 Decision tree classifier (min sample leaf = 10) – equity

The decision tree for the Equity tier can be seen in Figure 103. November 2021 appears to lead the charge by splitting out most of the Large Cap and Resources portfolios from the rest.

4.9.3 SOUTH AFRICAN – INTEREST BEARING

The results for the Interest Bearing tier can be seen in Figure 104. Although at first glance the results may not appear to be very different from the KNN classifier, some of the differences are significant. In particular, the fact that the very low duration portfolio that is in the Variable Term category has not been predicted to fall into the Short Term category is somewhat baffling.

Unfortunately, this classification method chooses the feature that best splits the data in order to best fit the data within the hyperparameters chosen, and this results in all the portfolios in the Variable Term category being split from all other portfolios. In other words, the model assumes that the labels are correct, and learns the best way to create splits in the data (months and values) so as to obtain the results being supplied.

Luckily, we can see exactly how the decisions were made by the classifier, by looking at the classification tree provided in Figure 105. You will note that the first node is based

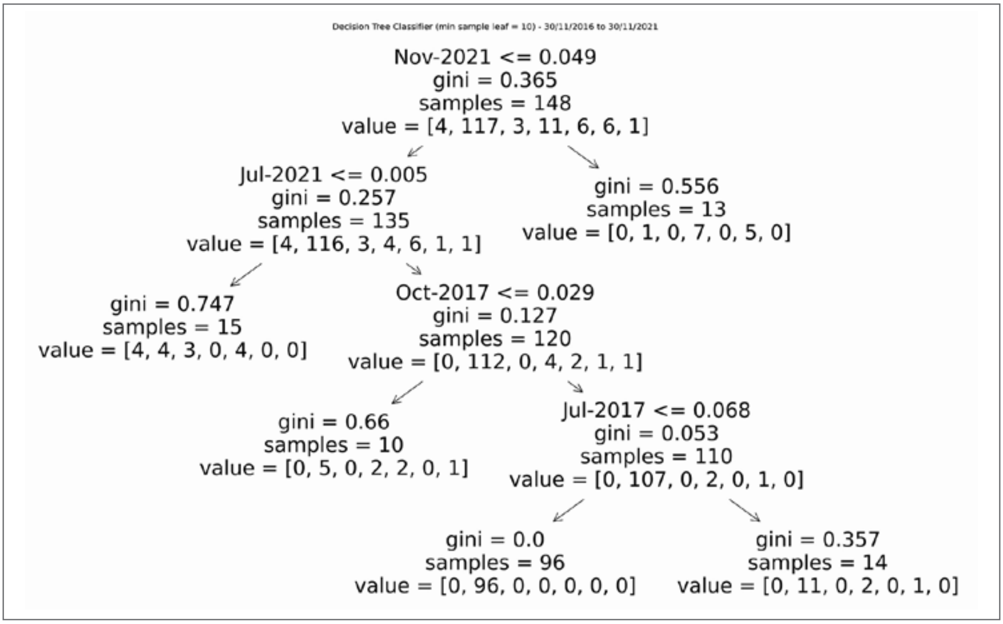


FIGURE 103 Decision tree classifier – decision tree – (min sample leaf = 10) – equity

on the returns for May 2018 being less than or equal to 0.003 (or 0.3% or 30 basis points). Clearly bonds did badly back then (ALBI delivered –1.95% for the month), so this forms a clean way to separate Variable Term portfolios from the rest (Money Market and Short Term).

Let us briefly describe the rest of the numbers in the tree so that it is easier to understand what is going on with the predictions. The first row tells you the feature (month) and value (return) that creates the split in the tree. The third row gives you the sample size before the split. The final row gives you a vector of the number of values in each of the categories, sorted by the same order as the classification report and confusion matrix. For our purposes, ignore the second row (gini) which is used by the algorithm.

```
Decision Tree Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[28  0  0]
 [ 2 25  0]
 [ 0  0 24]]
                                precision    recall  f1-score   support

 South African IB Money Market       0.93      1.00      0.97        28
   South African IB Short Term       1.00      0.93      0.96        27
South African IB Variable Term       1.00      1.00      1.00        24

                      accuracy                           0.97        79
                     macro avg       0.98      0.98      0.98        79
                  weighted avg       0.98      0.97      0.97        79
```

FIGURE 104 Decision tree classifier (min sample leaf = 10) – interest bearing

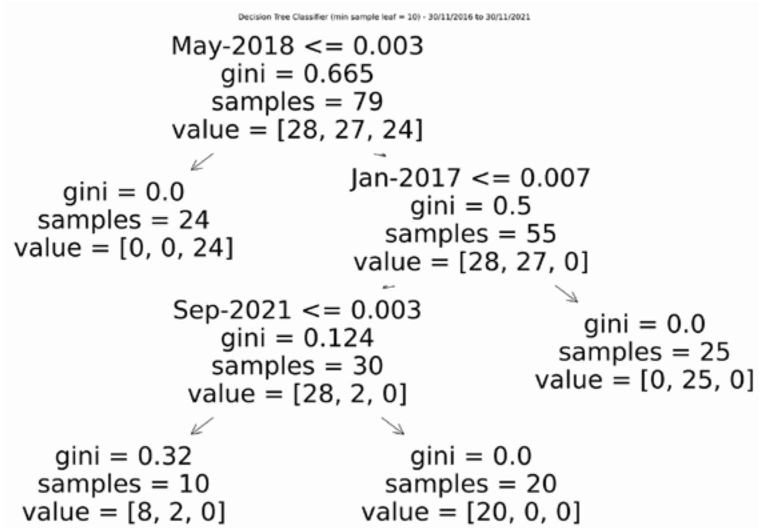


FIGURE 105 Decision tree classifier – decision tree – (min sample leaf = 10) – interest bearing

5. THOUGHTS ON ALTERNATIVE CLASSIFICATION

5.1 Multi asset

5.1.1 TOTAL EQUITY ALLOCATION

One possible remedy is to constrain the total equity exposure, not just with an upper limit, but also with a lower limit. You already have 0% and 100% as the absolute minimum and maximum. You also have the Equity categories from 80% to 100% total equity exposure. The High Equity category should be changed to fall between 60% and the 80%, being upper limit for the Medium Equity category, and the lower limit for the equity category respectively. There are two primary reasons for this. The first is that if it is a High Equity category, the total equity allocation should be high, and not medium or low, or some unknown value upfront within a very wide range. The second is that a Flexible category, if it is decided to keep it, can cater for variable equity allocation from 0%. There is also the obvious point that any portfolio could simply be Unclassified, if it does not fall neatly into an otherwise homogeneous category.

The Medium Equity category can then fall between 40% and 60% for the same reasons as above, and the Low Equity category can fall between 20% and 40%, to keep the total equity allocation widths the same. That means changing the Income category to fall between 0% and 20% (above the current 10%). There are again a couple of reasons for this.

Firstly, the Income category can currently allocate 25% to local listed property, which has historically been more volatile than local listed equity. More volatile generally means more risky, so this limit to property should probably also change (perhaps a maximum of 5% for the Income category is more appropriate).

```
Decision Tree Classifier - Confusion Matrix & Classification Report - 30/11/2016 to 30/11/2021

[[  0   3   0   0   0   0   0   0   0   0   0   1   0   0   0   0   0]
 [  0 106   0   4   0   0   0   0   0   0   0   7   0   0   0   0   0]
 [  0   0   0   3   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   3   0   8   0   0   0   0   0   0   0   0   0   0   0   0   0]
 [  0   4   0   0   0   0   0   0   1   0   0   0   1   0   0   0   0]
 [  0   5   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   1   0   0   0]
 [  0   0   0   0   0   0   0  27   0   0   0   0   1   0   0   0   0]
 [  0   0   0   0   0   0   0   2  21   0   0   0   4   0   0   0   0]
 [  0   0   0   0   0   0   0   0   1  22   0   0   1   0   0   0   0]
 [  0  11   0   1   0   0   0   1   1   0   5  12   2   2   2   0   0]
 [  0   6   0   0   0   0   0   0   0   0   5 125   0   6  10   0   0]
 [  0   0   0   0   0   0   0   3   1   0   0   0  62   1   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   1   3 111   3   0   0]
 [  0   0   0   0   0   0   0   0   0   0   4  14   0   8  46   0   0]
 [  0   0   0   0   0   0   0   0   0   0   1   3   0   0   4   0   0]
 [  0   4   0   0   0   0   0   0   0   0   0   3   1   0   0   0  20]]
                                    precision    recall  f1-score   support

    South African EQ Financial         0.00      0.00      0.00         4
      South African EQ General         0.75      0.91      0.82       117
   South African EQ Industrial         0.00      0.00      0.00         3
    South African EQ Large Cap         0.50      0.73      0.59        11
South African EQ Mid/Small Cap         0.00      0.00      0.00         6
    South African EQ Resources         0.00      0.00      0.00         6
 South African EQ Unclassified         0.00      0.00      0.00         1
 South African IB Money Market         0.79      0.96      0.87        28
   South African IB Short Term         0.84      0.78      0.81        27
South African IB Variable Term         1.00      0.92      0.96        24
     South African MA Flexible         0.33      0.14      0.19        37
  South African MA High Equity         0.75      0.82      0.79       152
       South African MA Income         0.83      0.93      0.87        67
   South African MA Low Equity         0.86      0.94      0.90       118
South African MA Medium Equity         0.71      0.64      0.67        72
  South African MA Target Date         0.00      0.00      0.00         8
      South African RE General         1.00      0.71      0.83        28

                      accuracy                             0.78       709
                     macro avg         0.49      0.50      0.49       709
                  weighted avg         0.74      0.78      0.75       709
```

FIGURE 106 Decision tree classifier (min sample leaf = 10) – all tiers & categories

Secondly, many of the portfolios in this category are too conservatively managed. This may be because the name of the category is thought of too literally, but many portfolios are indistinguishable from the South African Interest Bearing Short Term category. Again, when you are thinking about classifying homogeneous portfolios, this group should be more aggressive to distinguish it from other categories, although you will still be keeping the 0% lower limit.

The above proposed solution does not magically fix all of the problems with the existing category definitions, but it does go some way to addressing some of them. The bands for total equity exposure are still quite high at 20%, with portfolios with 60% total equity exposure being different to portfolios with 80% total equity exposure, but the bands are certainly much better than the current 75% range for the High Equity category, or the current 60% for the Medium Equity category.

Two portfolios in different categories could be more similar to each other than two portfolios in their own categories, e.g. one in the Medium Equity category with total equity exposure of 59% is more similar to one in the High Equity category with total equity exposure of 61%, than either is similar to portfolios within their own categories with total equity exposure at the opposite extreme allowed by the category.

5.1.2 TOTAL GROWTH ASSETS (EQUITY AND PROPERTY COMBINED)

We finally get back to a point that has been alluded to throughout the document, and we thank you for your patience if you have made it this far. Why were some portfolios in the High Equity category so different, despite having similar volatility, or why did a portfolio in the Medium Equity category have as much volatility as the highest volatility portfolio in the High Equity category, despite having an upper limit on total equity exposure of 15% less than in the High Equity category?

The reason could be in part because the portfolio does in fact have a higher allocation to equity than allowed by the category limits. Portfolios are allowed to breach the limits as long as the breach is remedied within a specified period of time (too long in our opinion). This is however not the reason.

The reason is that although equity is very distinct from bonds and money market instruments, in terms of risk, it is not that different from property. In fact, property has a fairly high correlation with equity, and even forms part of equity indices in many cases. Because property shares represent an ownership share in the underlying property companies, they are not unlike equities more generally. Technically, they are the same in many respects, although they can also differ in some important ones.

Ultimately though, property is considered a growth asset class, along with equities, albeit a different asset class for most. We should therefore be potentially thinking about these two asset classes in combination, when setting limits. The reason this has possibly not been done historically is because property shares really represented only a very small proportion of the overall equity market (less than 5%). Why set limits on something that has such a small weight in the overall scheme of things.

This could also be an overhang from Regulation 28 of the Pension Funds Act, that allows 25% exposure to property, but this envisaged direct property exposure, which many pension funds had exposure to. Although as risky as listed property, essentially being the same, volatility on direct exposure would be a lot lower because direct property is marked to model, not to market.

In the High Equity category, you can theoretically have 75% allocation to equity (upper limit), and 25% to property, taking the total exposure to growth assets to 100%. This mix is probably not ‘right’ as property is a much smaller part of the market than regular equities, at less than 5%, so we may want to adjust this ratio a little, if not a lot.

One potential solution would be to change the limits for the High Equity category to 80% equity, and to introduce a limit of 20% to property. At 20%, which is a quarter of 80%, this allocation is still very high given the size of the property market, but it is better than the current ratio of one third (75%/25%), and many portfolios would probably not use the full allocation to property, just like they do not currently.

We could then adjust the upper limits for the rest of the categories proportionally down as follows: 60% equity and 15% property for the Medium Equity category; 40% equity and 10% property for the Low Equity category; and 20% equity and 5% property for the Income category. Note that we are intentionally reducing exposure to property to keep the ratios constant, and because these categories are becoming (or should be becoming) less risky, and having 25% allocation to property for all of them is simply not appropriate.

This change to the Income category is probably the most significant, as some portfolios have 25% allocation to property in this category, and property has had higher volatility than equities for most periods of time. Although the upper limit for total equity exposure would increase from 10% currently, to a proposed 20%, the total exposure to growth (and very volatile and risky) assets, would actually decrease from 35% to 25%, which is probably more appropriate for this category.

We could then also introduce lower limits for equity for each of these categories as per above, being the upper limits of the category directly adjacent to it, so for example, the High Equity category would have a lower limit to total equity exposure of 60%, which is the upper limit for the Medium Equity category. The lower limit to property would need to be 0% as this sector is fairly small. This can therefore still create a level of overlap between categories, but equity exposure would at least be somewhat controlled and the overlap minimised to drift only, which could also be controlled.

For example, a Medium Equity portfolio could have 60% total equity exposure, along with 15% total property exposure, for a total exposure to growth assets of 75%, versus a High Equity portfolio with only 60% total equity exposure, for a total exposure to growth assets of 60%. This is not ideal, but much better than the current classification. The alternative would be to change the limits (upper and lower) to growth assets in total (equity and property combined) and address the problem thus by having no overlap.

This could take the following form: the Equity category already allows 80% to 100% equity exposure so the High Equity category limits would be 60% to 80% total exposure to growth assets, in any combination chosen, but perhaps an overall limit of 20% to property; the Medium Equity category would then have total growth asset exposure of 40% to 60%, and again perhaps a lower limit of 15% on property as before; the Low Equity category limits would be 20% to 40%, with property limits as before; and the Income category limits would be 0% to 20%, again with property limits as before.

The allowance to drift past the limits should also be heavily curtailed, perhaps requiring remedy by the month end in which the breach occurs. This would not only help portfolios stay within their exposure bands, but also help investors to know to some extent what they were buying into, which would be better than the current situation, where – as per an actual example – they could be buying a high equity portfolio with 0% exposure to equity.

5.1.3 FLEXIBLE CATEGORY

Where does this leave the Flexible category? Although many portfolios would legitimately want some flexibility to potentially move over two categories at most, most would comfortably sit within one of the categories, and should be moved into those. These should then just use the flexibility provided by the category, which is still substantial. The few remaining portfolios that want to remain truly flexible could remain in a Flexible category, but this would require this category to continue to exist, and that would create an incentive for the same portfolios to continue to use it, even those that do not belong in the Flexible category.

A better alternative would be to get rid of the category altogether and require truly flexible portfolios to be moved into an Unclassified category. Think back to the objective of having the categories, as having a homogeneous group of portfolios. A Flexible category does nothing to achieve this, and does not help investors or advisers understand the portfolios within this category any better, so rather make these few portfolios stand out as being very different from any well-defined category. It is also important to note that some of these portfolios hold more than 80% total equity exposure and should therefore be moved into the Equity category.

5.2 Equity

For the Equity tier, the debate can take a slightly different direction from the Multi Asset tier discussed above, where the justification was about creating more homogeneous categories of portfolios. For the Equity tier, we can ask the question of why multiple categories should exist at all. For the Multi Asset tier, the justification was predominantly on the basis of risk, but this is not a legitimate reason for multiple categories within the Equity tier.

Additionally, should investors be invested in sector portfolios at all? If so, what would be the reasoning for this? The objective when investing in a single asset class should be to maximise diversification, while perhaps trying to extract alpha if choosing to invest actively. This is especially true in South Africa, where our local equity market is fairly small, with only a limited number of shares being extremely liquid (say 60).

The Cap categories (Large, and Mid/Small) are also unnecessary for the same reasons as above, that investors should not be invested on this basis. Even where investors decide to combine large cap and mid/small cap themselves using different portfolios, they will require a decision on how much to allocate to each strategy, and how to change this exposure over time.

Portfolios wanting to distinguish on the basis of sectors or market cap could still do so, but would need to sit in the Unclassified category, highlighting that they were in fact very different from the General category. The General Equity category already has 170 portfolios, whereas the other six categories have a combined 40 portfolios. The difference in AuM is also very large. This provides evidence of the very limited appeal of these categories to the general public and even institutional investors, and it merely creates another source of confusion and opportunity for bad advice and decision-making.

If we wanted to rather take the approach of looking for homogeneous categories, you could reasonably conclude that you would only really need two categories for this tier, namely a General category, and a category for Resources, as these portfolios are substantially different to the General category as resources companies are generally very risky by comparison. The portfolios in the Resources category are amongst the very few that should probably be moved into an Unclassified category, perhaps along with some Mid/Small Cap category portfolios.

This could make the Equity General category rather large, and there is potential justification for dividing this up on the basis of local-only portfolios, versus portfolios that have an allocation to global equities, but this will be discussed below as a general point for distinguishing portfolios within categories.

5.3 Interest bearing

This tier is both easy and difficult to classify. The Variable Term category should exist with portfolios that are only managing strictly to the All Bond Index as a benchmark, even with strict duration limits around the duration of the All Bond Index. This would require moving out inflation-linked bond portfolios and portfolios managing to flexible duration, and these should probably just be moved into an Unclassified category.

Many people do not understand the need to have a long-duration bond portfolio, and would baulk at the idea of having to have long duration within a portfolio, just because the benchmark does. Long-duration bond portfolios are however very often good hedges for certain liability profiles and hence represent a legitimate way of investing and hedging, and an important asset class for many investors.

The Money Market category is also really well and narrowly defined creating no source for big differences between portfolios, and hence represents a very homogeneous group of portfolios.

The complication comes with the Short Term category, because it again has no lower limits on duration, and portfolios could end up looking very similar to portfolios in the Money Market category. This however does not create a problem in that the portfolios in this category are very heterogeneous, and hence it could be left untouched. The alternative would be to create a lower duration and/or maturity limit of say one year (or possibly less), to distinguish it from the Money Market category.

5.4 Local-only categories

Another source of variability of returns within categories, not covered previously, is global exposure in South African portfolios which some portfolios have, and others do not. While this may seem like an investment decision that the manager has full discretion over, and should therefore be left as a source of similarity or difference between portfolios, this is not always the case. Many portfolios have trust deeds specifically prohibiting investments offshore, and managers have often used these portfolios for investors wishing to separate their local and global investment decisions, building solutions on a 'specialist' basis.

Where a source of returns can lead to a high degree of variability, it is best to consider separating this out into categories that include this source of returns, and those that do not. This may be the case for global exposure within many of the categories covered above. If only a few portfolios within a category are different from the rest, then it may be better to just move those into an Unclassified category, instead of creating a new category for a few portfolios, as is currently the practice.

5.5 Unclassified

One more option to consider is to get rid of the Unclassified category at each second tier level, and rather have a single Unclassified category for each tier one. This could be justified on the basis that these portfolios should never be compared directly without further investigation into each of the portfolios and what may make them similar or different to all other in the same category.

6. RECOMMENDATIONS ON CLASSIFICATION

6.1 Multi asset

We recommend four categories, including High Equity, Medium Equity, Low Equity, and Income. The limits should be changed as recommended above, or some version of that. This means that portfolios currently in the Flexible category must either be reclassified into one of the proposed categories above, or into the Equity category, or an Unclassified category if they are truly flexible in terms of their asset allocation. This is appropriate as it will send a strong signal to investors and advisers that they should not expect a specific asset allocation from these portfolios, as they can invest very differently over time.

6.2 Equity

One category, namely General. Large Cap portfolios could be reclassified into the General category, or the Unclassified category. Sector portfolios, and Mid/Small Cap portfolios should be reclassified into the Unclassified category. It is probably worth thinking about the need for these portfolios, as we think they should completely disappear in time.

6.3 Interest bearing

We recommend three categories, including Money Market, Short Term, and Variable Term. Inflation-linked bond portfolios should be moved into the Unclassified category, as should flexible duration portfolios. The Variable Term category should be reserved for portfolios strictly managed to the All Bond Index, with duration limits around the duration of the index. Again, this will help to make the category homogeneous, and will let investors and advisers know what to expect from these portfolios.

6.4 Real estate

A single category as is currently the case.

6.5 Unclassified

A single category for all portfolios that do not belong anywhere else.