Michael Brown, VanEck

Michael Brown is a member of the Australian leadership team and a director of VanEck Investments Limited. He has over 30 years' experience in financial services and taxation. Prior to joining VanEck, Michael was executive director at boutique asset management consulting firm, Sunstone Partners. Previously he served as senior vice president at BT Funds Management and held roles as chief tax counsel at Perpetual Investments and MLC. Michael has always been at the cutting edge of developments in the asset management industry and government policy, using mathematics as a key to solving many problems.

WHY EQUAL WEIGHTING OUTPERFORMS

The Mathematical Explanation

Michael Brown

Equally weighted portfolios outperform their market capitalisation counterparts over the long term and over almost all short term periods. The evidence to support this is cited in the References and is demonstrated in Figure 1, which shows the performance of Australia's standard equal weighted index, the MVIS Australia Equal Weight Index, against Australia's standard market capitalisation weighted index, the S&P/ASX 200.

Various explanations have been offered since this phenomenon was first observed, such as the effect of selling high and buying low when rebalancing the portfolio. There hasn't however been a lot of mathematical analysis. This paper presents the data on individual stock returns to show why equal weighting has outperformed market capitalisation.

Equal weighting outperforming market capitalisation weighting can be explained by the statistical distribution of individual stock returns being skewed, which is contrary to the assumption that researchers generally make.

Skew

To get the maths right you have to start at the right point. The distribution of individual stock returns is not normal. That is, the distribution is not Gaussian.

It is now widely accepted that a normal distribution is a flawed way to explain financial markets because markets have 'fat tails' that normal distributions don't have. This is embodied in the delightful metaphor of a black swan. This is however only one way in which the actual distribution of stock returns differs from a normal distribution. It is time to discard the use of normal distributions.

The consensus description used for the distribution of individual stock returns that can be seen in the data is 'skewed'. Primarily this description indicates that the distribution is not symmetrical, as a normal distribution is. Rather, the actual distribution is pushed to one side, as can be seen in the histograms throughout this paper. There are however more differences than that.

The most comprehensive documenting of the skew of individual stock returns has been in the recent paper by Bessembinder. Bessembinder used the American monthly stock return database of the Center for Research in Securities Prices from July 1926 up

The quote

Equal weighting outperforming market capitalisation weighting can be explained by the statistical distribution of individual stock returns being skewed

Figure 1. Cumulative performance since inception of MVIS Australia Equal Weight Index

Source: VanEck, FactSet, as at 31 December 2017. Results are calculated to the last business day of the month and assume immediate reinvestment of all dividends and exclude costs associated with investing in the VanEck Vectors Australian Equal Weight ETF (MVW). You cannot invest directly in an index. Past performance of the Index is not a reliable indicator of future performance of MVW.

to December 2016, a period of 90½ years. His finding was that the distribution of returns from the individual stocks were highly skewed no matter what time period was chosen. He summarised what this means as follows:

"Simply put, very large positive returns to a few stocks offset the modest or negative returns to more typical stocks."

In Figure 2 the rectangle represents the whole of Bessembinder's data set of 25,332 companies with each company represented by the same area, about ½ square millimetre per company.

The standout result was how few top-performing companies it took to generate the same wealth as the total population. Half of the wealth over 90½ years was created by the companies represented by the blue rectangle in the top left hand corner and the other half was created by the companies represented by the green rectangle in the top right hand corner.

The grey space represents the lower-returning companies that in aggregate returned zero. Bessembinder tallied that the 25,332 companies created total wealth over the 90½ years of US$35 trillion dollars. The top 1,092 performers, ~4% of the population, on their own created the same US$35 trillion dollars of wealth. The other ~96% of companies totalled a zero return. The top 90 stocks, ~0.3%, created more than half of the $35 trillion dollars of wealth.

At the other end, 3,071 individual stocks, ~12%, lost all or nearly all of the money invested. For the curious, Bessembinder reported that the biggest return was from Exxon, contributing 2.88% of the total wealth creation. The second was Apple contributing 2.14%. Bessembinder provided similar results in respect of ten-year returns, annual returns and monthly returns.

There is extreme skew in each time horizon. He presented a lot of statistical calculations that demonstrate the skew but the numbers are unintuitive and there is not enough data in the paper to present the findings in a simpler way. Similar results from a second recent paper are easier to understand. This one was by Edwards, Lazzara, Preston and Pestalozzi of S&P Dow Jones Indices. Figure 3 is their chart of the returns of S&P 500 constituents from March 2003 to December 2017. There are three observations to make about Figure 3, reading from left to right:

Observation 1: The lower returning groups are bunched together, not spread out in a tail

A stock cannot do worse than -100%.

The authors' rendition of the distribution groups together all stocks that return between -50% and -100%. If the data had been presented continuously rather than as a bar chart, it would have shown that there are a large number that return exactly -100%. That is, a large number where there is a total loss of investment. This density at or near -100% conflicts with any attempt to use a normal distribution to describe the data.

Observation 2: The peak is way to the left

Many stocks have relatively low returns, as Bessembinder identified.

In a normal distribution the peak would be in the middle and the average and the median would coincide with the peak. It can be seen in Figure 3 that the average is higher than the peak and well above the median, which itself is well above the peak.

Observation 3: The tail to the right is very long
A small number of stocks have a very high return, again as Bessembinder identified.

This particular rendition of the distribution groups together all stocks that return more than 1,000%. This was necessary because otherwise the x axis would have been far too long to fit neatly on the page. It would however have looked even more tail-like.

The long tail could be said to be consistent with a normal distribution. What is important though is that the right side of the chart is totally different to the left side. On the positive side there are many companies well above 100%. On the negative side the limit is -100%. This is the important difference to the normal distribution.

Edwards et al are consistent with Bessembinder in finding that this shape is not a feature of the particular time period chosen but that it can also be seen over shorter time periods and over periods ranging back to the 1920s.

Figure 2. Individual stock returns over 90½ years

Source: Bessembinder, VanEck.

Skew is not just a US phenomenon. Edwards et al also show a very similar result for the constituents of the S&P Europe 350 Index. Another paper by Ganti and Lazzara of S&P Dow Jones Indices has similar findings for Japan's S&P/TOPIX 150, the S&P Pan Asia Ex-Japan & Taiwan BMI and our own S&P/ ASX 200.

To delve deeper into this phenomenon needs more data than is made available in either of these papers. The following analysis uses a data set of the 200 largest companies on ASX at 12 May 2015 and the return they each generated over the following three years.

For comparison with the chart on the previous page, Figure 4 presents this data as a histogram, with the returns rounded to the nearest 10%. The shape can be seen to match the shape in Figure 3.

Compared to Figure 3, which represents 25,332 stocks over 90½ years, Figure 4 only represents 200 stocks over 3 years. There are some observable differences that are to be expected:

- Figure 4 is not as filled out as Figure 3 because there are less stocks
- In Figure 4 there has been less time for stocks to go completely bust so the left hand side is barer than in Figure 3
- In Figure 4 there has been less time for the most successful stocks to build up a return so the right hand tail is not as long as it is in Figure 3.

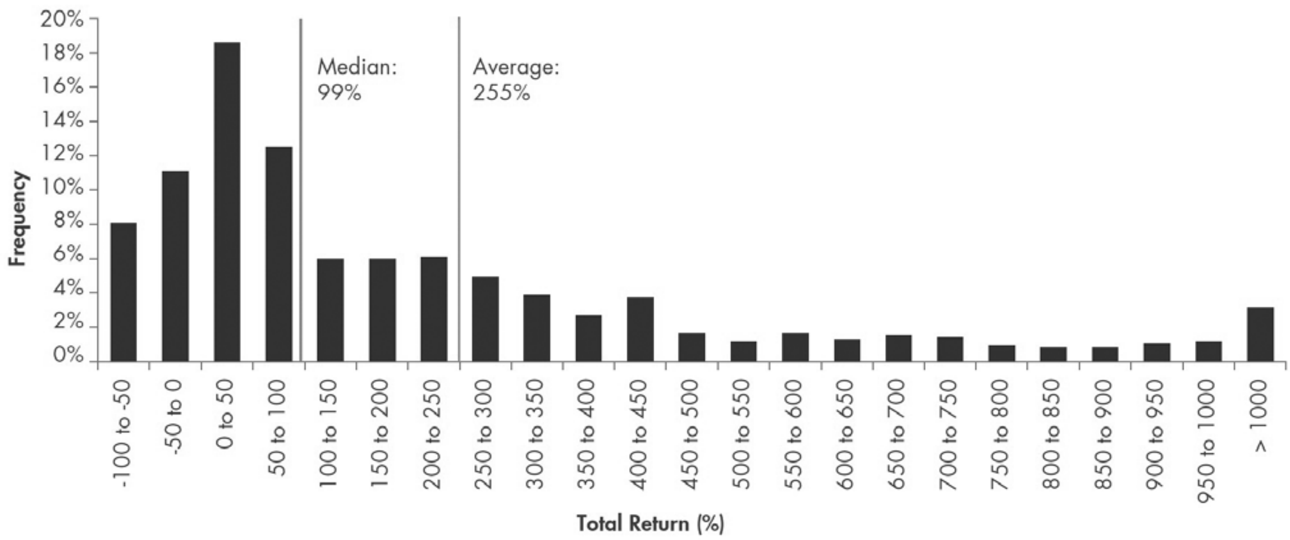
Consistent with the data in Figure 3, the average in this Australian data set is 34%, well above the median of 21% which itself is well above the peak of the histogram at 10%.

For the curious, the two bottom performers in the Australian data were Arrium and Slater & Gordon. The two top performers were BlueScope Steel and Regis Resources. A steelmaker at each end.

Ergodicity

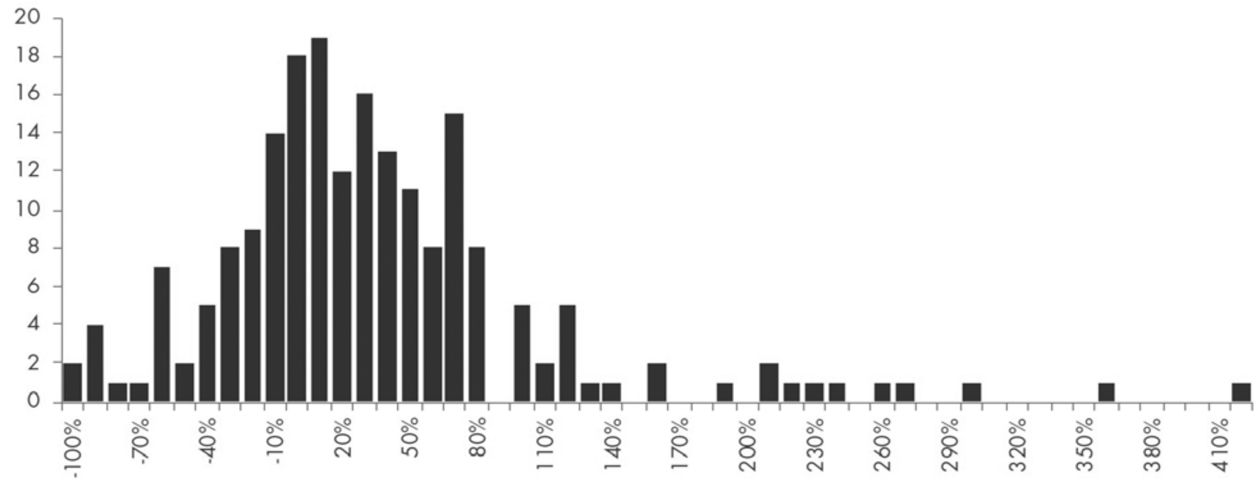
Another significant difference between this actual distribution of stock returns and the way a normal distribution would typically be applied is that this distribution is not ergodic. Ergodicity is the property of each constituent having the same chance to be at any point of the distribution as any other constituent.

Figure 3. Individual US stock returns March 2003 to December 2017



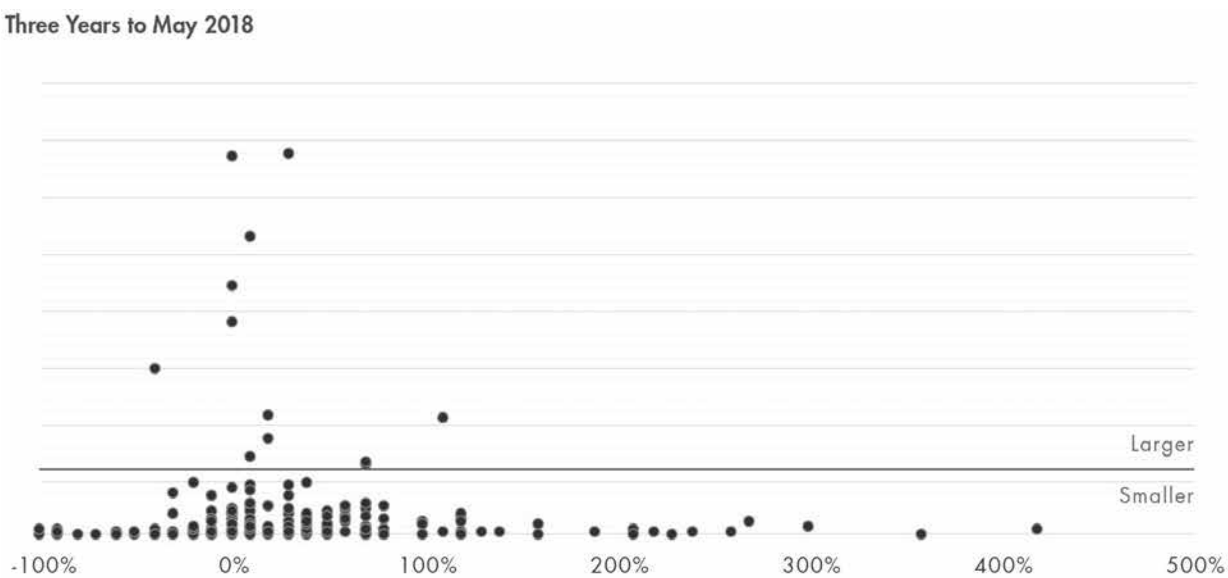
Source: Edwards, Lazzara, Preston and Pestalozzi

Figure 4. Returns for the 200 largest Australian stocks three years to May 2018



Source: Bloomberg, VanEck

Figure 5. Returns for the 200 largest Australian stocks versus their market capitalisation



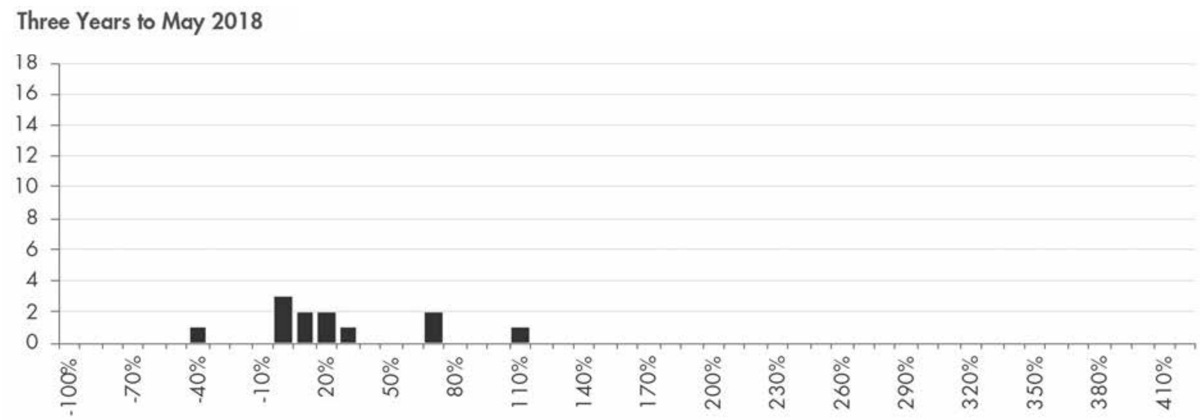
Source: Bloomberg, VanEck

In this context, for the distribution of stock returns to be ergodic, at the beginning of the period each stock would have had to have had the same likelihood as any other stock of being in the extreme right or extreme left. If the distribution was ergodic there would be no characteristic of the individuals stocks that could predict that certain stocks had a lesser chance than other stocks of being at either extreme.

Figure 5 plots the returns from the Australian data set against the stock's market capitalisation at the beginning of the period. It can be seen in Figure 5 that the companies that were the largest at the beginning of the period produced returns that are far more narrowly distributed than the rest of the stocks. The extremes are populated by the smaller stocks. The word 'smaller' is used with its precise meaning, as a relative term. There is no suggestion that the stocks referred to as 'smaller' are actually small.

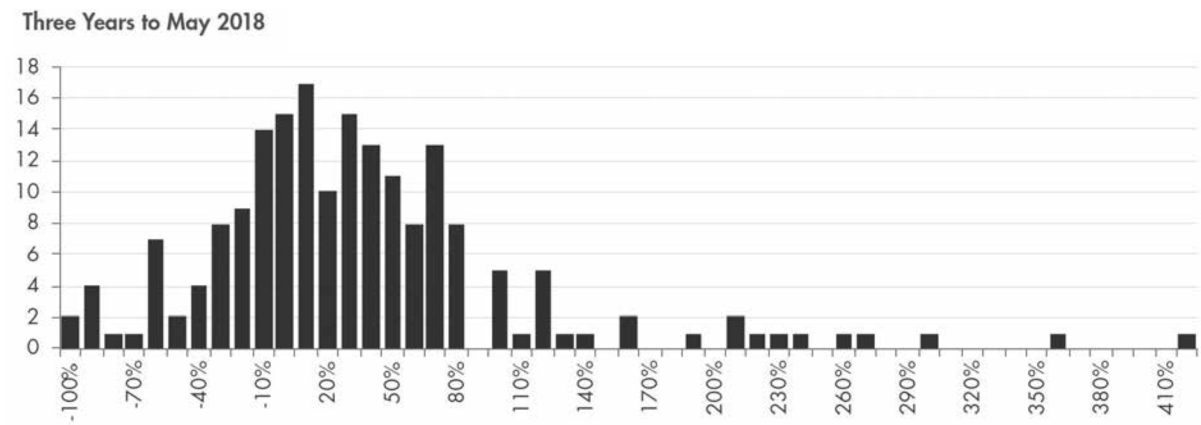
It is somewhat arbitrary to draw a line in Figure 5 and say one side of the line are larger capitalisation stocks and the other side are smaller capitalisation stocks. Nevertheless, this has been done in

Figure 6. Returns for the 12 largest Australian stocks versus their market capitalisation



Source: Bloomberg, VanEck

Figure 7. Returns for the next largest 188 Australian stocks versus their market capitalisation



Source: Bloomberg, VanEck

what can be seen to be a meaningful way and at a natural break. The twelve blue dots above the red line feel like a separate group to the 188 below.

The twelve larger cap stocks have a performance range of -36% (Telstra) to 107% (CSL) compared to the complete range of -100% to 416%.

The distribution of the twelve larger cap stocks however is still skewed in the sense used in this paper because the average is well above the median which is above the peak. The average for these 12 stocks is 25% and the median is 15%. The peak, the average of the two highest dots, is 14%.

That the returns from the larger caps are more narrowly distributed feels intuitively correct. The larger cap stocks are predominately big mature businesses so are less likely to go completely bust. On the other hand they probably already have a big market share for their main products so their growth is more limited than is typical for smaller, less mature businesses.

So the skewed distribution of individual stock returns can be separated into the sum of two separate skewed distributions with different parameters.

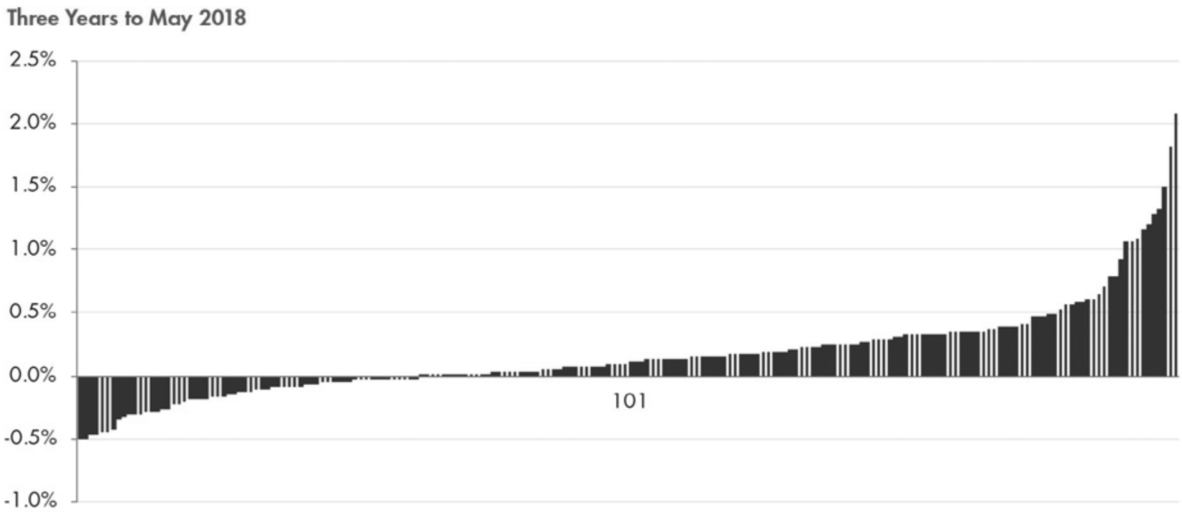
One for the larger stocks and one for the smaller stocks. For comparison with Figure 4, Figures 6 and 7 show the corresponding histogram for these two separate distributions.

Figures 6 and 7 demonstrate that the smaller stocks are more likely than the larger stocks to be in the extreme left and more likely to be in the extreme right, that is, the extremely low returns and the extremely high returns.

The smaller stocks outperform the larger stocks

It was shown above that the smaller stocks are more likely to be at the extremes of the skewed distribution. The next part of the explanation is to show that this means that the smaller stocks outperform the larger stocks.

Figure 8. Returns for the 200 largest Australian stocks in numerical order



Source: Bloomberg, VanEck

In other words, the smaller stocks on the right hand side contribute more than the smaller stocks on the left hand side take away. Because the right hand side goes as far as 416% but the left hand side can go no lower than -100%, this hypothesis would be the intuitive conclusion.

Looking a bit deeper, there is a similar intuition if you look at Figure 3, where the median is indicated in the chart, you can see that the bottom half covers the range -100% to 99% and the top half covers the range 99% to well over 1,000%. The corresponding data for Figure 4 is that the bottom half covers the range -100% to 21% and the top half covers the range from 22% to 416%. In both cases the top half seems weightier than the bottom half.

Figure 8 shows this imbalance visually. The return of each of the 200 stocks is charted in numerical order. Stock 101 is marked to indicate the start of the top half.

It can be seen from this chart that there is more blue from stock 101 up than there is below that mark. The right hand side can be seen to outweigh the left hand side. As your eye travels out from the centre to the extremes you can see that more is being added than is being taken away.

Turning to the mathematical explanation, this is a consequence of the average being higher than the median and the median being higher than the peak. This is why a skewed distribution gets more from its exposure to the right hand side than it suffers from its exposure to the left hand side.

The statements above are useful because in order to understand how stocks perform we all need to learn about skewed distributions. The better performance of the smaller stocks though can also simply be seen from the fact that the average for the smaller stocks is higher than the average for the larger stocks.

While the idea that smaller stocks outperform has been folklore at least since Fama and French's three-factor model, the mathematics above do more to show the nature of this phenomenon than the three-factor model attempted to do.

The explanation of equal weighting's outperformance

This finding that the smaller stocks outperform the larger stocks immediately explains the consistent outperformance of equal weighting over market capitalisation weighting. Equal weighting has consistently given greater exposure to the smaller stocks than market capitalisation weighting does. It is as simple as that. Greater exposure to smaller stocks which outperform larger stocks.

The way forward

The finding of a distribution of individual stock returns that is shaped as described above and that is non-ergodic when size is considered is a strong effect that cannot be ignored. Any analysis of relative performance between two different portfolios should isolate the effect of these two characteristics before trying to argue that any other effect is present.

Past suggestions that the explanation of equal weighting's outperformance lies in what happens when an equal weight portfolio rebalances are thrown into doubt. The data sets above show outperformance even though there is no rebalancing. A possible line of future research is whether, after adjusting for the exposure to smaller stocks, rebalancing adds or detracts from the performance.

The other common suggestion, following Fama and French's three factor model, is that equal weighting outperforms due to a greater exposure to value stocks. Having found a strong effect from the exposure to smaller stocks, the question of whether a measure of value would also be useful is difficult. It would only 'also' be useful if it was an additional explainer and not just a second way of looking at the same phenomenon.

Mathematically, something can only be an additional explanation if it is statistically independent of the first explanation. This is a hurdle. The factor analysis that every academic now has programmed into Excel assumes that all factors are independent, but there is no work done to validate this assumption. Factor analysis should not proceed until the independence of the second factor is established.

The finding that smaller stocks outperform larger stocks is not necessarily an argument to exclude larger caps completely. Performance is not the only objective. The narrower range of the distribution of larger cap returns means less drawdown and less variance.

While diversification can address these matters to a large extent, there can still be a role for anchoring a portfolio with stocks from this narrower range of returns. Particularly if an equal weighted portfolio is seeking to outperform a market capitalisation index and tracking error is to be controlled.

Appendix

Rejecting the lognormal distribution and other mathematical functions

One attempt to get away from the error of assuming a normal distribution has been to assume a lognormal distribution instead. The fatal flaw in this attempt is Observation 1 (see 'Skew' section). A lognormal distribution starts at x=0, y=0, which severely underrepresents the number of very low returning stocks.

Part of the attraction to the lognormal distribution is that researchers always want to use a distribution that can be expressed as a continuous mathematical function. This is because once you have a continuous mathematical function you can do a lot of manipulation and dissection very easily.

Unfortunately, if you start with an invalid assumption the conclusions you draw from your manipulation and dissection will also be invalid. The easy conclusions may fill out a research paper but they have no other value.

We all have to be realistic and recognise that there is no continuous mathematical function that can accurately represent these distributions of stock returns. Admitting this kills off a lot of conclusions we could have otherwise drawn but since those conclusions would have been invalid, we are better off without them. There is investors' money at stake so we should stick to what is valid. **FS**