NORGES BANK
INVESTMENT MANAGEMENT

06 | 2014

BENCHMARK DESIGN FOR AN ACTIVE INVESTMENT PROCESS

DISCUSSION NOTE

Most equity funds are benchmarked to market capitalisation weighted indices. A large investor may wish to design non-market capitalisation weighted benchmarks to maximise the benefit to the investor of the stock picking ability of active portfolio managers. We introduce a generic modelling framework which takes sector characteristics, manager skill, risk-taking appetite and market states into consideration. Our simulations may guide the design of customized benchmarks which offer investment capacity and increase the benefit of an active investment process.

Date 25/11/2014
ISSN 1893-966X

The Discussion Notes are written by staff members at Norges Bank Investment Management.

Norges Bank may use these notes as specialist references in letters on the Government Pension Fund Global. All views and conclusions expressed in the discussion notes are not necessarily held by Norges Bank.

dn@nbim.no
www.nbim.no

SUMMARY

BENCHMARK DESIGN FOR AN ACTIVE INVESTMENT PROCESS

Most equity funds are generally benchmarked to market capitalisation weighted indices. A large investor may wish to carve out some capital and allocate risk to active portfolio managers (PMs) with the aim of enhancing the fund return. Assuming the active investment process is organised along sector lines, the investor faces a benchmark design question with two key variables. The first is to decide on how many stocks should form part of the research list that can be covered by each sector PM, given resource constraints. The second is how should the sector benchmark be designed that is aligned with the investor's investment objectives and that exploits the PM's stock picking skills.

We advocate a diversity weighting scheme linked to market capitalisation in the design of the sector benchmarks. Diversity is defined by how evenly capital is distributed among the securities in the market. By lowering the concentration of capital in the largest stocks, the more even distribution of weights allows for a better representation of over and under-weight active decisions taken by the PM.

The optimal choice of the diversity parameter and the number of names in the benchmark ('research list') is a function of a number of competing variables. We present simulation results for a range of PM skill levels and styles (concentrated vs. diversified portfolios), conditioned on market states (volatility, dispersion, correlation and any prevailing size effects). We introduce a systematic framework which assesses a number of competing variables that include active returns, transfer coefficient and investment capacity, to determine the 'optimal' benchmark structure.

In our case study on the design of the European Banks research list for a skilled PM, we recommend a diversity parameter in the range 0.2-0.4 applicable for a range of 60%-80% of total number of names in the original cap-weighted sector benchmark. Our simulation framework can be applied for other sectors to arrive at the appropriate levels of diversification and is sufficiently versatile to reflect sector-specific characteristics and PM styles.

Introduction

Benchmarks continue to play a major role in the asset management industry. In particular, market capitalisation weighted benchmarks remain the bedrock of relative performance assessment of active decisions taken by portfolio managers. Researchers and practitioners have investigated apparent weaknesses in cap-weighted benchmarks. Challengers to market cap indices came in two broad forms. The first is a family of indices that assign weights to constituents based on attributes of securities other than prices. A well-known example is perhaps the 'fundamental' weighting scheme where weights are directly proportional to company valuation metrics such as dividends, sales and earnings (Arnott et al, 2005). They remain, however, not without their critics (see for example Perold (2007)). Second is a family of risk linked indices where weights are assigned according to volatilities and/or correlation statistics of the underlying securities. A well-known example under this category is the risk parity based scheme. The literature in this space has focused on performance metrics comparing such alternative weighting schemes (sometimes referred to as 'smart beta' indices) relative to the well-known cap weighted indices. We note that market cap indices enjoy a key feature that others do not; namely, that in absence of corporate actions and other scheduled rebalance screens (typically once or twice a year when index providers re-examine the representation of the markets), the owner of a cap-weighted replicating portfolio need not rebalance. That is, cap-weighted indices are 'passive' in a stricter sense than other alternatives. A survey on alternative weighting schemes can be found in Chow et al (2011). These alternatives have been gaining some popularity, partly driven by some of the issues raised in relation to cap weighting schemes (see for example Amenc et al (2006) and Hsu (2006)). A recent empirical evaluation of alternative equity indices using heuristic, optimized and fundamental weighting schemes can be also be found in Clare et al (2013 a,b).

The benchmark design problem we address in this paper does not focus on the choice of weighting scheme to arrive at a more 'efficient' beta representation of a market. To our knowledge, there is very little literature that articulates our specification. This may be due to a more unique investment process that demands such a formulation. Let us first assume an investor has already made a decision on the choice of the overall benchmark of their fund. For simplicity, we proceed with a market cap weighted benchmark at the total fund level. Let us assume further that the investor has capacity to take active decisions for a portion of the fund but not at the overall fund level (for example, due to liquidity, risk or other strategic constraints). Using the 'core-satellite' terminology sometimes used in the practitioner literature, the core is generally passively held (to the extent possible) and the 'satellite' is an active portfolio. In the typical problem setting the satellite portfolio is generally actively held against the same benchmark choice as the core (market cap weighted in our example). That implies that core (passive) and satellite (active) portfolios can simply be aggregated and compared to the same overall benchmark. The only decision to be made therefore is that of capital allocation and/or risk budgeting between the two portions.

However, one could argue that a cap weighted benchmark may not be well suited for an active portfolio manager. One drawback in a typical cap weighted index which we will address formally in the note is the lack of 'diversification' embedded in the benchmark. Market cap weighted indices typically have a high concentration of weights in a small number of the largest securities typically resulting in low weights in a large number of mid and small cap stocks. In a long only portfolio context, this will generally limit the diversification benefits that enable active PMs to express broader active views across names and size spectrum. In order to overcome the drawbacks of actively managing long-only funds against capitalization weighted indices, practitioners have attempted to adopt index 'extension' schemes such as '130/30', where the manager can go long and short up to 30% of the fund value as long as the net exposure to the index remains 100%. While such a structure has theoretical merits (see for example Leibowitz et al (2009)), the limited experience some long-only managers have had with 'shorting' securities is perhaps the reason for the mild take-up rate of these strategies. The active benchmark design problem is to construct a suitable custom benchmark[1] on the portion to be carved out from the original market cap index to become the yardstick for the active manager to beat. The core portfolio manager would then 'neutralise' that custom benchmark in the overall fund while allowing the active manager the freedom to take positions against it. It is worth noting that the absolute benchmark performance itself per se is not relevant to our problem design since its risk relative to the original benchmark is hedged in the context of the overall fund. Moreover, turnover in the context of our benchmark design problem is not relevant since the combination of the resulting custom benchmark for the active PMs and the residual benchmark managed by the core PMs is equivalent to the original cap-weighted benchmark, which by construction remains the lowest cost passive exposure. Turnover may become pertinent if management decides to re-design the custom benchmarks on a regular basis, which is unlikely in practice.

The practical implication is to build tailored research lists for the active managers according to their specialisations which will form the universe of stocks for the design of the custom benchmark. Typically managers are organised by regional or global industries and the research lists would be designed accordingly. Given time and resource constraints in terms of coverage, a typical research list constitutes a subset of names in a given sector specialisation (regional or global) and generally excludes the smallest names in the original benchmark. The key decisions in the sector benchmark design problem therefore become a choice of the number of names in the research list (universe), and the choice of the weighting scheme that is suitable for the investor's active investment process. In broad terms, an optimally diversified sector benchmark should incentivise each portfolio manager to utilise their stock-picking skills, and at the same time to be able to enhance the fund's overall performance in a scalable way.

More specifically, our sector benchmark design problem has two defined objectives in mind. First is to maximise potential for outperformance by limiting the number of benchmark names to allow portfolio managers to express high conviction positions while maintaining sufficient coverage of the sector.

1 *We use custom benchmark and research list interchangeably throughout the note*

Second is to embed diversification in the choice of weighting scheme. This can be achieved by moving away from market cap and towards equal-weighting to allow managers to take on meaningful active positions across their research lists, both 'long' (over-weight) and 'short' (under-weight) positions.

It is perhaps useful to set some additional questions at this stage which may inform the modelling process. Below and in no order of priority are some of the issues in the portfolio design process:

- What is the optimal number of names selected in the research list[2] that balances 'breadth' of sector coverage (cross-sectional dispersion across sufficient number of securities) and resources (in terms of manageable number of names) required by the PM to conduct detailed company research?

- What is the ideal weighting scheme for sufficient 'diversification' that enables the PM to better express a wider set of relative views given the limitations imposed by the small cap tail inherent in a cap-weighted benchmark?

- What is the impact on investment capacity of moving towards a more diverse research list (in the limiting case, equally weighted list)?

We address the above issues and make recommendations on the appropriate sector benchmark design that is suited to the active portfolio manager. The recommendations are backed by empirical evidence based on comprehensive simulations. The note is organised as follows. We first present the theoretic components of our model where we introduce diversity weighting and contrast it with other weighting schemes. Following which we define parameters necessary for our simulation framework. These include key variables that would impact our performance metrics. Specifically we expand on how we simulate a portfolio manager's skill. We then address the portfolio construction process of a 'typical' sector PM used in our simulations. We next introduce the last component of our simulation framework where we define market states characterised by volatility and relevant factor exposures. Finally, we utilise our systematic framework to determine the optimal benchmark structure, and illustrate the approach using European Banks as our case study.

Theoretical and Empirical Framework

Introducing Diversity Weighting

Portfolio managers are generally incentivized to outperform their benchmarks. The design of the benchmark is therefore critical to the PM's investment style and the associated risk appetite. An active stock picker who aims to outperform a benchmark makes two decisions – selection of names for inclusion in the portfolio, and active weighting of the selected names in the portfolio. The first decision is a function of the size of the research list under

2 *PMs may have some freedom in choosing non-benchmark names. Further research is required to determine the optimal level of freedom allowed.*

his coverage – more names implies more diversity in choice but more resources required for company research. This results in choosing outperformers ('ins') vs. underperformers ('outs') and is implicitly a long-short portfolio decision unrelated to the weighting scheme in the benchmark. The second decision is then to translate the relative views on the outperformers – stock X is preferred over stock Y and as such will receive a higher active over-weight against its benchmark weight relative to Y, all other things being equal. The active over- and under-weights are therefore a function of the distribution of weights in the benchmark. As noted earlier, market cap weighted benchmarks tend to have high concentration (in weight terms) in the larger cap stocks, thus limiting the ability of an active stock-picker to express bottom-up views across the names. Specifically, a tail of small cap stocks with low weights will not warrant significant research attention given that a 'short' position would only imply a low active under-weight (weights cannot be smaller than 0 in the portfolio).

This is where the diversity weighting scheme could be useful. Diversity measures how evenly capital can be distributed among stocks in the market or in a portfolio. When diversity is low, capital is more concentrated (e.g. the case of a typical stock which is market cap weighted) and when diversity is high, capital can be more evenly distributed (in the limiting case, equal weighting across all stocks in the market). Fernholz et al (1998) define market or portfolio diversity, at a given point in time, using the function:

$$D(p) = \left(\sum_{1}^{n} w_i^p\right)^{1/p} \qquad (1)$$

where n is the number of stocks, w_i is the capitalisation weight for stock i and $0 < p < 1$, the diversity parameter. It can be shown that $1 \le D(p) \le n^{(1-p)/p}$ and the measure of diversity $D(p)$ generates a diversity weighted portfolio with the resulting stock weights as a function of market cap weights and p:

$$w_i = w_i^p \Big/ \sum_{i=1}^{n} w_i^p \qquad (2)$$

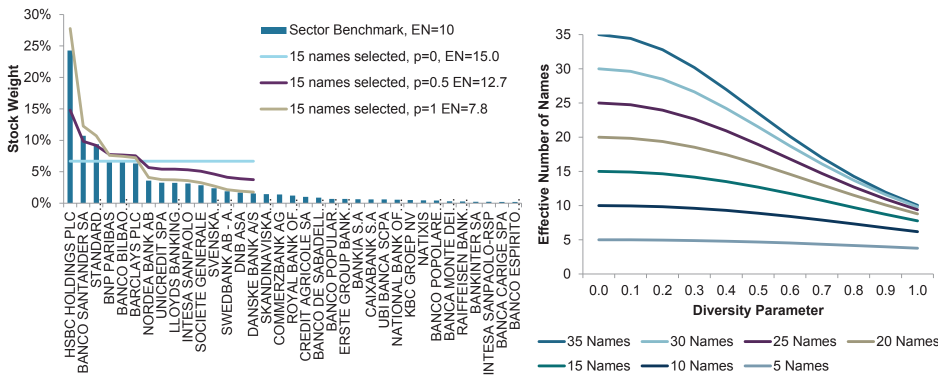
Compared to the market cap weighted portfolio, a diversity weighted portfolio under-weights the larger stocks and over-weights the smaller stocks[3]. We note that when $p = 1$ the portfolio is market cap weighted, and when $p = 0$ the portfolio becomes an equally weighted one (i.e. maximum diversity is achieved under this measure). One further desirable feature of diversity weighting is that a link is established between the new weights and the original market cap weights. This has implications on investment capacity in the weighting scheme which we will re-visit later. Alternatively, diversity can also be measured by the 'effective' number of names against which the PM can be measured (Strongin et al, 2000):

$$n_{eff} = 1 \Big/ \sum_{i=1}^{n} w_i^2 \qquad (3)$$

3 Fernholz et al (1998) go on and show that if diversity in a stock market is mean reverting and there is a positive equity risk premium, then a diversity-weighted portfolio is likely to outperform the market.

We illustrate the concept through an example of the European Banking sector which will form the base of our case study throughout the paper.

Chart 1: Weighting scheme comparison for European Banks (left panel) and trade-off contour for choice of names and diversity parameter (right panel) as of 31 January 2012



Source: MSCI, NBIM calculations

The original market benchmark has 35 names shown in order of their weights in chart 1. Let us assume that the research list for the active PM has been set to the largest 15 names. The lines shown in the left panel of chart 1 represent the weighting schemes for two p values contrasted with cap weights ($p = 1$) and equal weights ($p = 0$). On the right panel of chart 1 we show, for a given n names in the research list, the trade-off contours between effective number and the diversity parameter $(0 < p < 1)$. Some observations can be made that may guide the choice of names and diversity parameter. In the case of a 15-name research list, for example, one can consider p values in the range up to 0.3-0.4 that will still not adversely affect the effective number of names relative to the maximum possible (15). One can also see the lack of diversity present in the cap weighted banking sector. In fact, for $n = 15, 25, 35$, the effective number of names is almost the same (approximately 8 to 10).

Simulation framework and performance metric

In this section, we present our simulation framework which examines the behaviour of key competing variables that can influence the characteristics of an optimal benchmark structure. In determining the diversity parameter and number of names in the benchmark, the variables we analyse include manager skill, risk aversion level, transfer coefficient, diversification level, investment capacity and their impact on active return. In line with a fundamental PM's long term investment horizon, we simulate 1-year return forecasts that represent buy-and-hold strategies. In particular, we rebalance our portfolio annually at the end of January each year, and hold the same portfolio until the next rebalance. To investigate the effect of benchmark structure on our performance metrics, we first vary the proportion of benchmark names chosen to form the research list. For a given number of names selected in the research list, we then replicate a range of representative manager skills by running multiple simulations[4] to generate stock forecasts, such that their correlation with next-period returns is a function of a pre-specified skill level

4 *100 simulation runs for a given manager skill level*

ranging from 'high' to 'low'. We also include results for underperforming managers. To further examine the effect of systematically moving from cap to equal weighting, we decrease the diversity parameter, p from 1 (cap weighting) to 0 (equal weighting) in increments. Lastly, we examine how risk tolerance levels impact our performance metrics by constructing portfolios with different concentration levels (or active risk relative to benchmarks). Figure 1 provides an overview of our simulation framework described above.

Figure 1: Overview of simulation framework

For each year
- **For some percentage of benchmark names selected**
  - **For some manager skill level**
    - Simulate stock forecasts for a given skill level
    - **For some benchmark diversity level**
      - Calculate investment capacity
      - **For some portfolio style (concentrated and diversified)**
        - Construct typical portfolio and measure relevant performance metrics such as excess return, information ratio, transfer coefficient
      - **Next**
    - **Next**
  - **Next**
- **Next**

Next

Source: NBIM calculations

Post portfolio construction, we pinpoint and measure some performance metrics such as active returns, information ratio (IR) and transfer coefficient (TC). The transfer coefficient is defined as the cross-sectional correlation between active weights and forecast returns. In other words, it measures how well a PM can express his over/ underweights. The relationship between IR and IC was established in the seminal work by Grinold (1989) and Grinold and Kahn (2000), which they referred to as the Fundamental Law of Active Management. Clarke et al (2002) go on to show that the information ratio and hence expected active return is intimately linked to both the transfer coefficient and information coefficient through the generalized fundamental law of active management[5]:

$$IR \approx TC \times IC \times \sqrt{N} \quad (4)$$

where N = number of independent bets. In terms of the expected active return, equation (4) becomes:

$$E(R_A) \approx TC \times IC \times \sqrt{N} \times \sigma_A \quad (5)$$

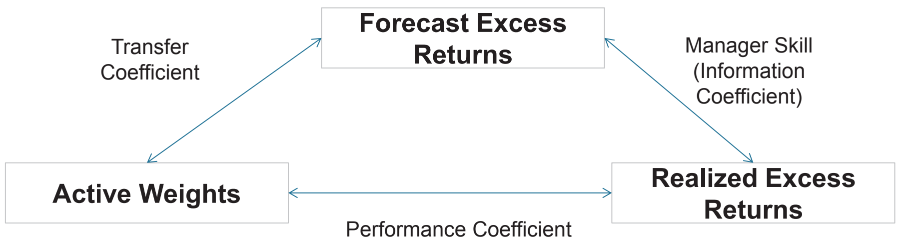
where σ_A = active risk of the portfolio. In correlation form, the authors show that the generalised fundamental law can be expressed as:

$$PC \approx TC \times IC \quad (6)$$

5 *The most critical simplifying assumption in the mathematical derivation of the generalised law is the assumption of a diagonal residual covariance matrix, i.e. residual stock returns are perfectly uncorrelated with each other after stripping out the 'market' effect and other systematic risk factors. Stubbs (2013) shows that the information coefficient in the equation summarising the generalised law may be incorrectly estimated if both realised and forecast returns are not neutralised for systematic risk factors. The author further shows that the return of the IC will be realised only if the TC is equal to one and if IC is measured appropriately as above. However, throughout the paper, we simply refer to the transfer coefficient as the correlation between active weights and raw forecast returns to better reflect how fundamental PMs stock pick in practice. We also prefer to 'purify' the alpha signals in the portfolio construction phase as detailed in the note as it better reflects the risk management process fundamental PMs employ in practice. Note that while we do not use the equations related to the generalised law directly per se, we may refer to them to motivate our results.*

where PC is the performance coefficient, defined as the expected correlation between active weights and subsequent excess returns. Equations (4) to (6) imply that, for a given N number of independent bets and some target active risk, a higher expected performance coefficient translates into higher returns. Figure 2 shows the inter-relationship between performance coefficient, transfer coefficient and manager skill. It illustrates that performance (as measured by the performance coefficient) is a function of both manager skill (the right leg of the triangle) and the constraints imposed in the portfolio construction process (the left leg of the triangle). In the absence of portfolio constraints, the transfer coefficient is equal to 1. However, PMs rarely enjoy the luxury of a completely unconstrained portfolio. Portfolio constraints such as no short sales in a pure long-only setting limit the full transfer of information into active weights.

Figure 2: The correlation triangle



Source: Clarke et al (2002)

Clarke et al (2002) further decompose the realized performance coefficient, $\varphi_{\Delta w,r}$, into components arising from realised IC (manager skill), $\varphi_{\alpha,r}$, and noise associated with portfolio constraints, $\varphi_{c,r}$:

$$\varphi_{\Delta w,r} \approx TC\varphi_{\alpha,r} + \sqrt{1 - TC^2}\varphi_{c,r} \quad (7)$$

where r is the next-period residual return and c can be thought as the 'optimal weight not taken' on each stock because of portfolio constraints. Equation (7) is significant as it analyses the drivers of performance coefficient. The first term in (7) is due to manager skill, whereas the second term is an exogenous noise component associated with portfolio constraints. When $\varphi_{c,r}$ is positive, performance is increased because the portfolio construction process forces higher weights than expected on outperforming securities and lower weights than expected on underperforming securities in order to satisfy the portfolio constraints. The relative weighting between both terms depends on the transfer coefficient. This exogenous noise component, which is outside the PM's control, is generally a function of the benchmark design (in terms of the number of names in research list and weighting scheme) and market state.

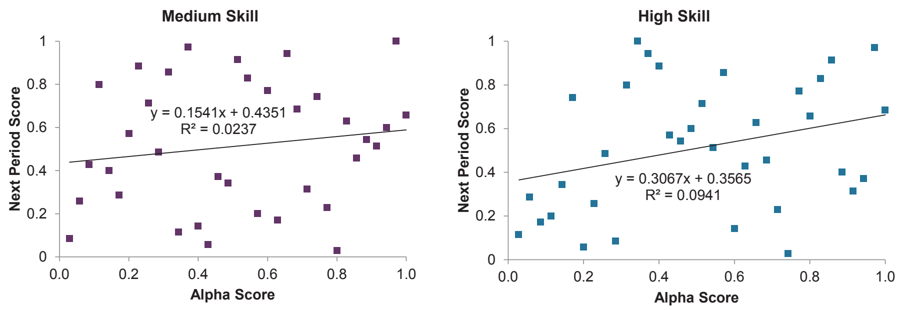
Defining and simulating manager skill

In our simulations, we assume manager skill to be the correlation between the ranks of alpha signals and subsequent realised excess returns, that is, we are primarily interested in the ordering of the signals relative to the realised excess returns. To simulate a specific manager skill level, we first normalise the next-period return into rank scores, $\boldsymbol{y}$, such that they range between 0 and 1:

$$y_i = \frac{rank(r_i)}{n} \qquad (8)$$

where r_i = next-period return for stock i and n = number of names in the research list. Next, we simulate our alpha scores by generating a uniform random variable, $\boldsymbol{x}$, which ranges between 0 and 1, such that its correlation with $\boldsymbol{y}$ represents the desired manager skill. We simulate manager skills for ICs ranging from -0.3 to 0.3 using 0.15 intervals. One can consider IC=0.3 to imply a highly skilled PM, and IC=0.15 a medium skilled PM. A negative IC implies an underperforming PM and analyses are included for completeness. Chart 2 below shows a typical scatterplot of scores for simulated forecasts and next-period returns for both medium and high-skilled managers from a single simulation run.

Chart 2: Manager skill simulation from a single run – An illustration



Source: MSCI, NBIM calculations

Portfolio construction

In this section, we present our portfolio construction framework to build a stylised replication of active portfolios typically managed by a fundamental sector PM. We set up an optimisation problem which maximises our portfolio-weighted simulated alpha scores, subject to some active risk budget, market 'beta' and size neutrality. The latter constraints are introduced such that the portfolios are maximally exposed to the stock specific alphas, and do not reflect any unintended systematic risk factors. We have assumed two types of portfolio management styles – concentrated and diversified with an annualised active risk of 8% and 4% respectively[6]. Throughout this exercise, we assume perfect foresight on active risk and stock betas relative to the cap-weighted sector benchmark. To achieve size neutrality, we first rank stocks belonging to the research list in descending order, such that stocks with the largest weight in the original sector benchmark will get the highest rank scores. We then normalise the rank scores by their cross-sectional average. Specifically, the rank score on security i at time t is:

$$Size\ score_{it} = c_t\left(rank(w_{it}) - \frac{\sum_{i=1}^{n} rank(w_{it})}{n}\right) \qquad (9)^7$$

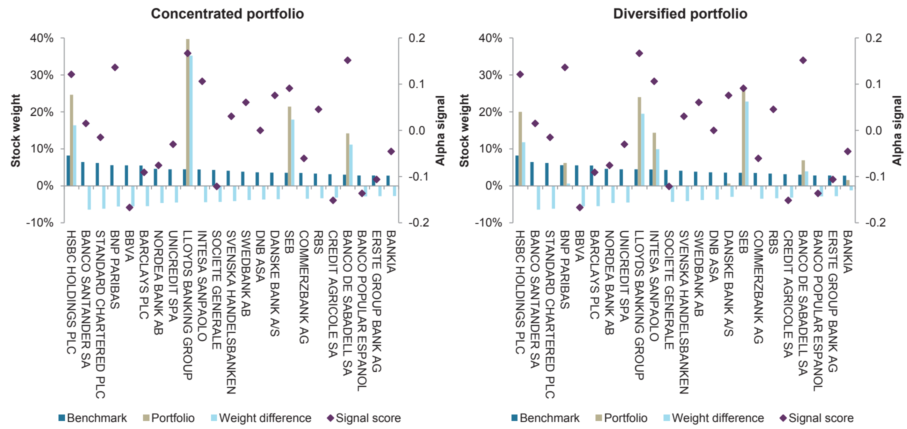
where w_{it} is the weight of stock i in the original sector benchmark, and c_t is a scaling factor such that the rank scores are scaled to one unit positive

6 *In our framework, active risk budget is set a priori (for example, by management). However, one can re-design the simulations such that active risk or level of 'aggressiveness' is an additional decision variable.*

7 *Note that* $\sum_{i=1}^{n} rank(w_{it})$ *can be simplified into* $n(n+1)/2$

and one unit negative. Size neutrality is then achieved by setting the portfolio-weighted size score to zero. Beta neutrality with respect to the custom benchmark is also realized in a similar manner.

Chart 3: Illustration of typical portfolios constructed on 31 Jan 2012 (IC=0.15, p=0.3 and 23 out of 35 sector benchmark names selected)



Source: MSCI, NBIM calculations

Chart 3 shows the typical profile of portfolios constructed relative to a fairly diverse research list (23 sector benchmark names selected with p = 0.3) on 31 January 2012 for a medium-skilled PM (IC=0.15). For the concentrated portfolio (active risk of 8%), we see that there are 4 high conviction long positions, namely HSBC, Lloyds, SEB and Banco De Sabadell. For the diversified portfolio (active risk of 4%), the long positions are spread out more evenly across 6 names (maximum overweight of 22.8% in diversified portfolio vs. 35.2% in concentrated portfolio). Note also that both portfolios are size and beta neutral by construction. As noted earlier, because of portfolio constraints (in our case, our portfolio constraints include no short sales, beta and size neutrality), the cross-sectional correlation between the signal and active weights is not perfect; for example, in our concentrated portfolio, BNP has an underweight exposure even though its signal is amongst the highest[8].

Market states – Volatility and size regimes

We first define market regimes in terms of two variables: sector volatility and any prevailing size effect. We define the size effect as the simple linear correlation between current benchmark weights and next one-year returns across the stock universe. To construct our market regimes, we first calculate the average sector volatilities, and size effects across all benchmark structures[9] for each period, and then, bucket the averaged metrics across all periods into quartiles. Finally, we classify a period based on a factor as being high (top quartile), low (bottom quartile) or medium (all other quartiles). For each regime, we also calculate the corresponding dispersion and average pairwise correlation between stocks[10]. In the case of the size effect, note that

8 Note that Lloyds and BNP have approximately similar historic market 'beta' and size attributes. Given that the PM prefers Lloyds to BNP (on a relative basis), has a concentrated portfolio management style, and is keen to neutralise 'beta' and size effects, Lloyds is chosen over BNP.

9 For each year, we average the corresponding sector volatilities and size effect across all benchmark names and diversity parameters.

10 Note that dispersion and pairwise correlation is only dependent on benchmark names.

a positive (negative) number implies the presence of a strong large cap effect (strong mid/ small cap effect).

Chart 4 shows that large caps typically outperform (underperform) mid/ small caps during periods of high (low) volatility in our sample period. This is expected as there is typically a 'flight to quality' and away from 'riskier' small caps when volatilities spike during distressed periods. On the other hand, mid/ small caps tend to do relatively better in rising markets when volatility is benign and during 'risk-on' environments. De Silva et al (2001) linked US mutual fund performance to cross-sectional security return dispersion, a measure of the opportunity set available to an active PM. Yu and Sharaiha (2006) show that return dispersion is closely linked to (time-series) volatility and derive the following approximation for market dispersion, χ_t^2:

$$\chi_t^2 \approx \overline{\sigma^2}(1-\bar{\rho}) \qquad (10)$$

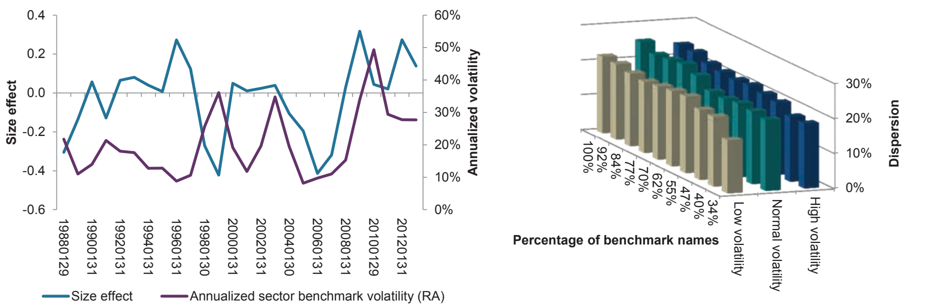
where $\overline{\sigma^2}$ is the average stock volatility and $\bar{\rho}$ is the average pairwise correlation between stocks. Dispersion is therefore expected to fall when average stock volatility falls and/ or when average correlation rises. This is precisely what we observe. Chart 4 shows that the much lower pairwise correlation offsets the low stock volatility to produce an 'alpha' opportunity set (i.e. cross-sectional dispersion) approximately equal to that available during high volatility periods.

Chart 4: Measuring dispersion, pairwise correlation and size effects across volatility regimes

Panel A: Median market-state variables conditional on volatility regimes

	Average size effect	Average sector benchmark volatility	Dispersion	Average pairwise correlation
High volatility	0,04	34,3%	21,3%	0,62
Normal volatility	0,03	18,6%	22,3%	0,44
Low volatility	-0,17	10,2%	21,1%	0,26

Panel B: MSCI Europe Banks - Historical benchmark volatility and size effect (left chart); Dispersion across volatility regimes (right chart)



Source: NBIM calculations

Performance drivers and interaction effects

Before presenting summary results from our simulations in the next section, it is useful to understand how different components (as described earlier) including benchmark structures, active risk levels, manager skill and market states interact with each other to impact the transfer coefficient, which in turn affects the excess returns. We present summary findings in this section. For a more detailed discussion, please refer to the appendix. Table 1 summarises the impact of benchmark structure, manager style and sector volatility on both transfer coefficient and performance independently.

Table 1: Drivers of transfer coefficient and performance[11] for a skilled manager

Dependent variable	Increase in explanatory variable			
	Benchmark diversity	Benchmark names	Active risk	Volatility
Transfer coefficient	Increase	Decrease	Decrease	Increase
Excess returns	Increase	Increase	Increase	Decrease
Risk-adjusted excess returns	Increase	Increase	Decrease	Decrease

Source: NBIM calculations

We start with the leftmost explanatory variable – benchmark diversity. As the benchmark shifts from cap-weighting (diversity parameter of 1) to equal-weighting (diversity parameter of 0), the transfer coefficient improves as the PM's relative views are better expressed[12], leading to higher absolute and risk-adjusted excess returns. Despite the fall in transfer coefficient with benchmark names (recall that the names in the benchmark are included in order of their size, which means more of a small cap tail as more names are incrementally added), we find that the diversification effects from a greater number of 'independent' bets more than offset the corresponding fall in transfer coefficient, which in turn translates into better performance[13]. Whilst the increase in active risk generally leads to higher excess returns, the fall in transfer coefficient means a less than proportional increase in performance per unit of risk. This is expected as a high-conviction PM who is less risk-averse will require larger active weights which will make portfolio constraints more binding, and hence adversely affecting his/ her ability to express relative views. The explanatory variables mentioned so far can be generally controlled by management. Volatility regimes, on the other hand, are beyond such control. We find that high volatility regimes tend to be associated with high pairwise correlation or co-movement between stocks which reduces diversification benefits. The smaller number of available 'independent' bets more than offsets the improvement in transfer coefficient in such a scenario, therefore adversely impacting performance.

Putting it all together

In this section we provide a generic framework that will assist in the design of sector benchmarks aiming to enhance the excess return potential of an active stock picker. Management may need to take into consideration an

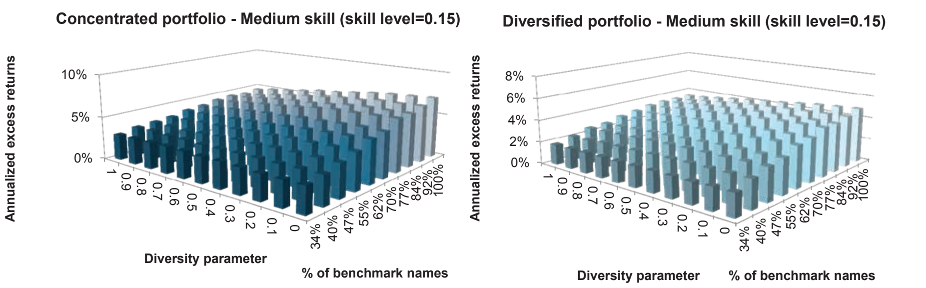
11 *Increase in benchmark diversity implies a decrease in* p*, the diversity parameter.*

12 *Although the short sale constraint for large cap stocks becomes more binding as one shifts from cap to equal weighting, the relaxation in the short sale constraint for mid/ small cap stocks more than offsets this effect, contributing to the improvement in transfer coefficient.*

13 *Note that this result assumes that manager skill is constant even with increasing name coverage which may not be true in practice.*

additional number of factors or constraints that will be specific to the PM or equity sector universe. Examples of these include resource constraints that may dictate the maximum number of names to be selected from the original benchmark, and the number of specialisations within a sector which the PM may need to cover. In the case of the latter one may further decompose the sector into its constituent sub-groups. The framework can handle additional sector-specific constraints. The results we present here can help guide the parameter choices in the final design of a sector benchmark.

Chart 5: Annualised excess returns across different benchmark names and diversity parameters for different portfolio concentrations (medium skill, IC=0.15)

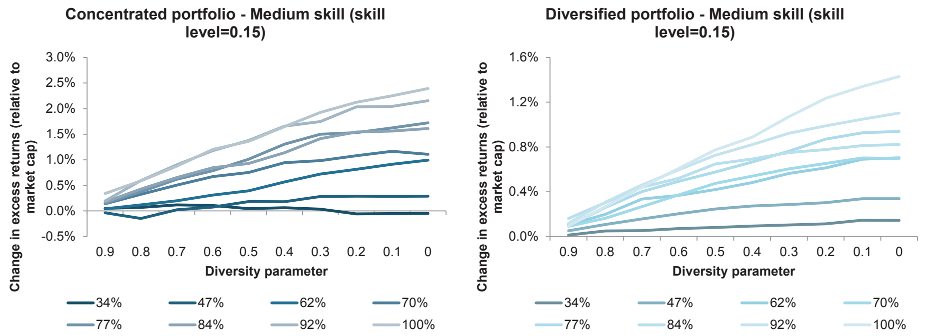


Source: NBIM calculations

Chart 5 shows the annualised excess returns of our simulations when the number of benchmark names and the diversity parameter are varied. For brevity, we focus on a portfolio manager with medium skill (IC=0.15) with two different active styles. The left panel presents the results for a concentrated portfolio choice, a stylised version of an active PM who believes in taking on a limited number of stocks with a higher active risk ('tracking error' versus his sector benchmark). The right panel presents the results for a more diversified portfolio choice with lower active risk.

Starting with the left hand panel, it is clear that the excess returns (bar heights) increase, on average, with an increasing number of names chosen and level of diversity (decreasing p). It is worth noting, however, that the pattern is non-linear in the choice of diversity parameter p. For a given number of names selected we observe some concavity in the curves describing the positive relationship between excess returns and decreasing values of p from 1 (market cap) to 0 (equal weights). In particular the excess return tends to fall more sharply as p is above 0.3-0.4 indicating that the effective number of names becomes a critical part of the performance.

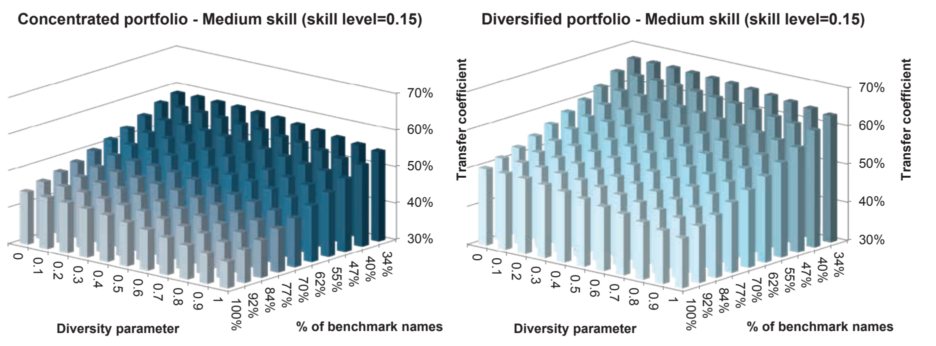
Chart 6: Change in excess return (relative to cap-weighted benchmark) for choice of names and diversity parameter across entire sample period (medium skill, IC=0.15)



Source: NBIM calculations

The concavity is better illustrated in chart 6. In deciding on the value of p to represent sufficient sector diversity, and acceptable number of names selected from the benchmark, we can measure the incremental improvement in excess returns relative to the market cap benchmarks. Each contour represents a fixed choice of average number of names over the sample period (47% represents about 21, 62% about 28 and 84% about 38). In the case of the concentrated portfolio, an excess return increase of approximately 100 bps relative to the market cap benchmark can be achieved for a p value of 0.5 for a wide range of name selection (60-85% name inclusion). However incremental excess return through additional diversity by lowering value of p from 0.5 to 0.3 now depends on the number of names selected in the benchmark. In the case of larger number of names (35 names representing 77% of the benchmark) the excess return improvement increases from 1% to 1.5%, whereas in the case of a lower number of names (15 and 21 names representing 34% and 47% of the benchmark respectively), there is no marked improvement in performance. Since the names in the benchmark are included in order of their size, a limited number of names selected imply less of a small cap tail and hence the benefits of a diverse weighting scheme become less pronounced compared to the market cap weighted benchmark. The picture is not dissimilar in the case of the diversified portfolio management style.

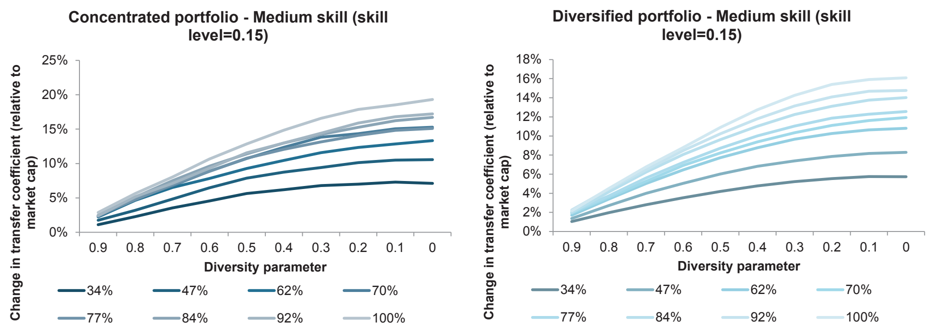
Chart 7: Sensitivity of transfer coefficient to benchmark names and diversity parameter across entire sample period (medium skill, IC=0.15)



Source: NBIM calculations

As discussed in the previous section, the performance coefficient is driven by both the information coefficient (manager skill) and noise related to portfolio constraints, where the latter component is deemed to be exogenous. To minimise the contribution from the noise component, a high transfer coefficient is desired. Chart 7 shows the sensitivity of transfer coefficient to both benchmark names and curvature for a medium-skilled manager. For the concentrated portfolio, we note that the transfer coefficient begins to fall materially for p greater than 0.4 and for larger choice of benchmark names. This relationship is better illustrated in chart 8 where we measure the improvement in transfer coefficient relative to market cap benchmarks for different choice of names and p values.

Chart 8: Percentage change in transfer coefficient (relative to cap-weighted benchmark) for choice of names and diversity parameter across entire sample period (medium skill, IC=0.15)



Source: NBIM calculations

Finally, in deciding the appropriate benchmark structure, management may need to take into consideration capacity constraints, where capacity is defined as the maximum proportion of the cap-weighted sector benchmark at the original fund level which can be carved out without impacting the no-short constraint in the fund.

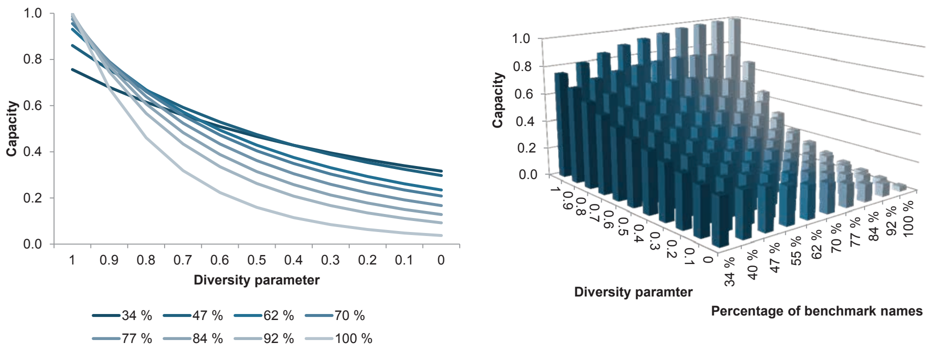
In table 2, we illustrate the concept of capacity through a simple example of carving out a custom benchmark from a hypothetical cap-weighted index consisting of four names. For simplicity we choose an equally weighted custom benchmark. In this example, the smallest security in terms of market capitalization (stock D) is the key determinant in setting the capacity limit. To adhere to equal weighting, the dollar value of the maximum carve out is 60 MM out of original 100 MM available (i.e., a capacity of 60%). This is equivalent to the ratio of the weight of the binding stock in the cap weighted benchmark to its weight in the custom benchmark.

Table 2: Example of carving out an equal-weighted custom benchmark

	Stock A	Stock B	Stock C	Stock D	Total
Cap-weighted benchmark weight	35 %	30 %	20 %	15 %	100 %
– *Corresponding notional value (MM)*	35	30	20	15	100
Equal-weighted custom benchmark weight	25 %	25 %	25 %	25 %	100 %
– *Corresponding notional value (MM)*	15	15	15	15	60

Chart 9 shows the capacity trade-off contours for choice of benchmark names and p values. Recall that market capitalisation still plays a key role in the diversity weighting scheme by reducing/ increasing stock weights proportional to their caps, therefore preserving as much capacity as possible for a given diversity level. Whilst capacity decreases as p falls, the rate of decrease depends on the number of names selected in the benchmark. We note that the rate of decrease is much more linear for a small number of names, whereas the rate of decrease falls faster as one moves away from cap weighting for a large number of names. For p >0.8, we find that capacity is less sensitive to the number of names selected as the benchmark is less diversified and closer to original market cap version. For p <0.5, the number of names becomes more critical as adding names by order of size implies a larger small cap tail.

Chart 9: Sector benchmark capacity for choice of benchmark names and concavity[14]



Source: NBIM calculations

One can arrive at the 'optimal' choice of p by considering the trade-off between transfer coefficient, active returns and investment capacity for a given skill level[15]. For a medium-skilled manager benchmarked to European Banks, we find that the range of p = 0.3 to 0.4 generally achieves a level in transfer coefficient and active returns close to the maximum improvement over cap weighting, without severely impacting investment capacity. Having determined the optimal p range, we now focus on the number of names one should include in the research list for a concentrated portfolio choice that may inform management's choice.

14 Averaged over our entire sample period from Jan 1987 to Jan 2012.

15 Alternatively, one could approach this optimization problem of a multiple objective function in a more formal framework (see for example Davies et al (2009) on the design of a portfolio of hedge funds).

We first define the capital allocation ratio as the proportion of the original fund's cap-weighted sector that is allocated for active investing by sector PMs. The top panel of chart 10 compares the realised and target capital allocation ratios[16] for p = 0.3 to 0.4, where each contour represents a specific target capital allocation ratio. We define the realised capital allocation ratio, $car_{realized}(p, n)$, to be the minimum of capacity and the target capital allocation for some diversity parameter, p, and choice of names, n:

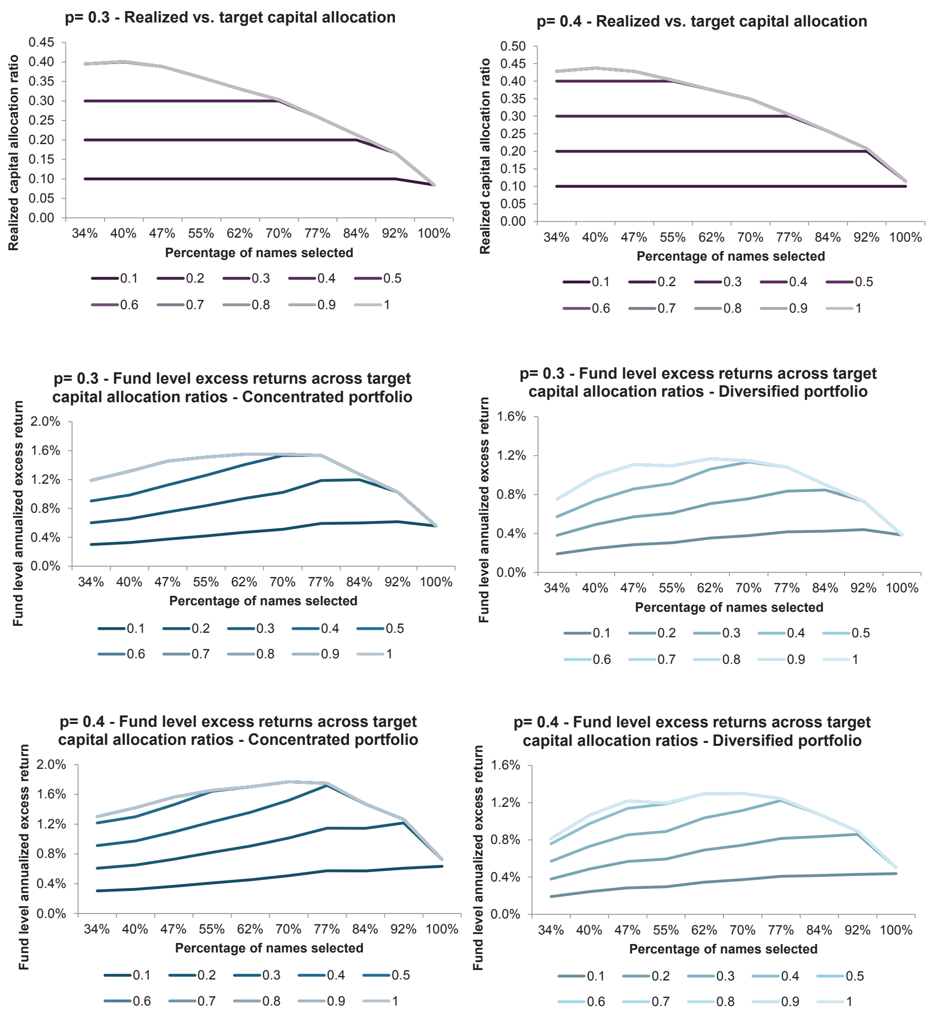
$$car_{realized}(p, n) = min(car_{target}, c(p, n)) \quad (11)$$

where car_{target} is the target capital allocation and $c(p, n)$ is the capacity of the custom benchmark characterised by p and n. The fund-level excess return of the sector, $R_{A,fund}(p, n)$, is then defined as:

$$R_{A,fund}(p, n) = R_{A,sector\ PM}(p, n) \times car_{realized}(p, n) \quad (12)$$

where $R_{A,sector\ PM}(p, n)$ is the active return of the PM who is benchmarked to a research list characterised by p and n.

Chart 10: Realised capital allocation ratios (top panel) and fund-level excess returns for a medium-skilled manager across benchmark names for different portfolio concentrations and p *values (p=0.3, middle panel; p=0.4, bottom panel)*

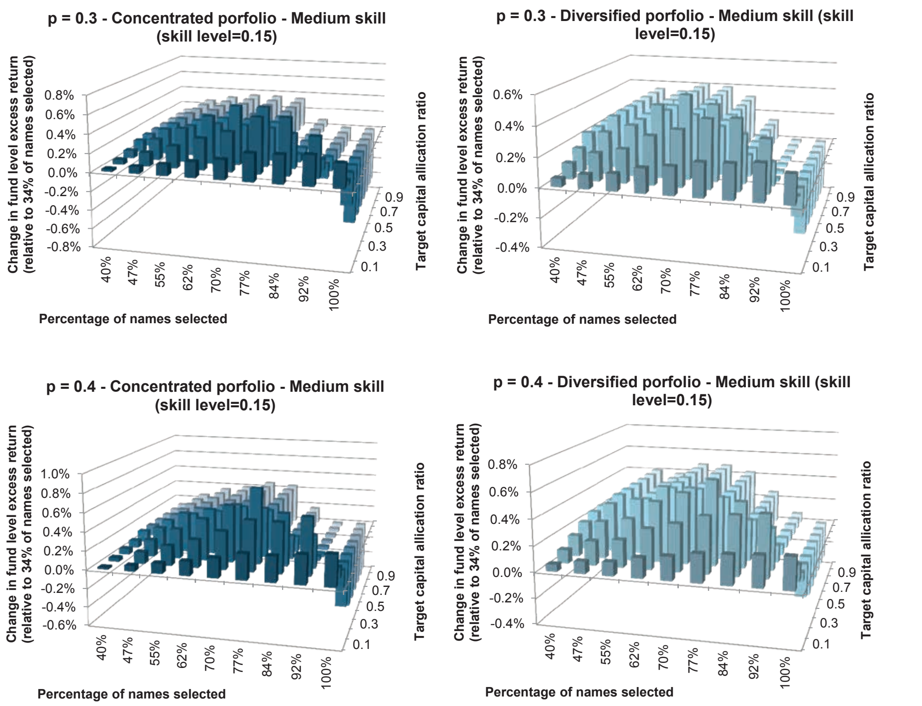


Source: NBIM calculations

16 *Management can set target capital allocation ratios for sector PMs according to various considerations including complexity of sector, manager skill and risk-taking appetite.*

For example, even if the intention is to put 100% of capital to use into a fairly diverse (p =0.3) research list, capacity is approximately 40% for small choice of benchmark names, which decreases with the number of benchmark names (see top left panel in chart 10). On the other hand, a lower target proportion of original cap-weighted sector to be carved out is naturally less binding as the benchmark structure is able to accommodate the lower capital. The middle and bottom panels in chart 10 show the fund-level excess return trade-off contours for choice of benchmark names and target capital allocation ratios. These contours illustrate the trade-off between investment capacity and improvement in excess returns relative to the tailored research list that has been carved out from the original sector benchmark.

Chart 11: Change in fund-level excess returns for a medium-skilled manager across benchmark names and target capital allocation ratios for different p *values (p=0.3, top panel; p=0.4, bottom panel)*

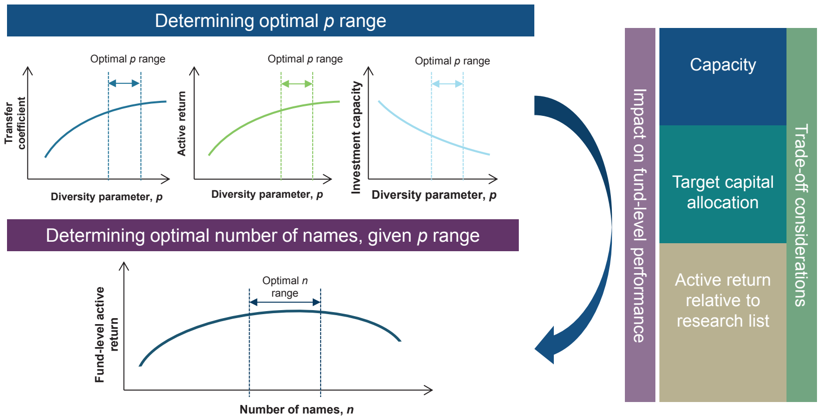


Source: NBIM calculations

Chart 11 shows the absolute change in fund-level excess returns (relative to 34% name inclusion) with the number of benchmark names selected for inclusion. It appears that approximately 60-80% name inclusion is 'optimal' for a wide range of target capital allocation. For smaller target capital allocation ratios (0.1-0.2), around 80% name inclusion is optimal assuming that manager skill does not deteriorate with wider coverage of names.

Putting it all together and considering the trade-offs (active returns, transfer coefficient, diversification benefits and capacity), the optimal range of parameters defining our custom European Banking sector benchmark for a medium skilled manager is 0.3< p <0.4 applicable for 60-80% name inclusion from the original sector benchmark.

Figure 3: Schematic of generic framework for determining optimal benchmark diversity and names for a skilled manager



Source: NBIM calculations

Figure 3 presents the schematic summarising our systematic framework for determining the optimal benchmark structure. We can first determine the optimal benchmark diversity by observing the concavity in both the transfer coefficient and active return with p, taking into account the falling capacity with p. Recall that a high transfer coefficient ensures the performance coefficient is sufficiently driven by manager skill. From the derived optimal p range, we then decide on the optimal sector coverage by looking at the behaviour of its corresponding fund-level active returns with key competing variables in the form of the sector PM's active return, research list investment capacity and target capital allocation ratio. We can also take into consideration impact of market states on these metrics to form robust views.

Using the framework presented above, we extend our analysis to include results for other skill levels. Table 3 summarises our results. Across all manager skills, we note that the transfer coefficient profiles are similar in both pattern and magnitude. For a highly-skilled manager (IC=0.3), our results favour a slightly more diverse benchmark (p ranging from 0.2 to 0.3). In this case, a marginal increase in transfer coefficient translates into greater value-add in terms of performance relative to a medium-skilled manager. On the other hand, PMs may find stock selection particularly challenging during distressed periods typically characterised by high volatility regimes. In such a scenario, we advocate a less diverse benchmark, both in terms of p values and choice of names[17].

17 For PMs with negative skill, the optimal benchmark parameters become a mirror image of the positively-skilled manager. In this case, the PM's enhanced ability to express relative views and greater diversification benefits from a more diverse benchmark becomes detrimental to performance.

Table 3: Summary of optimal parameters for different manager skills and active risk

	Optimal parameters			
Manager skill	Diversity parameter, *p*	Percentage of benchmark names	Number of names	Comment
High skill (skill level=0.3)	0.2–0.3	60-80%	21–28 out of 35-name benchmark	Optimal choice of names nearer to upper bound when target capital allocation ratio is small
Medium skill (skill level=0.15)	0.3–0.4	60-80%	21–28 out of 35-name benchmark	Optimal choice of names nearer to upper bound when target capital allocation ratio is small
No skill (skill level=0)	Higher *p* values	Lower percentage of names	Lower number of names	Less diverse benchmark implies more 'free-riding' on the small/ mid cap effect
Negative skill	Higher *p* values	Lower percentage of names	Lower number of names	Optimal parameters mirror image of positive skill

Source: NBIM calculations

Conclusion

We advocate a diversity weighting scheme linked to the market caps in the design of tailored research lists for the active sector PMs. By lowering the concentration of capital in the largest stocks, the more even distribution of weights allows for a better representation of relative views taken by the PM. The optimal choice of diversity parameter, p, and choice of number of names in the benchmark (research list) is a function of a number of competing variables in the investment process. For an assumed skill level and target active risk, optimal diversity parameter and benchmark names depend on following trade-off considerations: active return, transfer coefficient, diversification benefits, or number of 'independent' bets available and investment capacity. In this paper, we have introduced a systematic framework for determining the optimal benchmark structure, and presented simulation results for a range of PM skill levels and styles (concentrated vs. diversified), conditioned on different market states (sector volatility and size effects). Our simulation framework is generalisable to other benchmark design questions related to active investment.

References

Amenc, N., F. Goltz, and V. Le Sourd, (2006): "Assessing the Quality of Stock Market Indices", EDHEC Risk Institute Publication.

Arnott, R., J. Hsu and P. Moore (2005): "Fundamental Indexation", Financial Analysts Journal, 61(2).

Chow, T., J. Hsu, V. Kalesnik, and B. Little (2011): "A Survey of Alternative Equity Index Strategies", Financial Analyst Journal, 67(5).

Clare, A., N. Mosten and S. Thomas (2013a): "An Evaluation of Alternative Equity Indices Part 1: Heuristics and Optimised Weighting Schemes", SSRN working paper.

Clare, A., N. Mosten and S. Thomas (2013a): "An Evaluation of Alternative Equity Indices Part 2: Fundamental Weighting Schemes", SSRN working paper.

Clarke, R., d. Harindra and S. Thorley (2002): "Portfolio Constraints and the Fundamental Law of Active Management", Financial Analysts Journal, Sep/Oct 2002, 58, 5.

Davies, R.J., H.M. Kat and S. Lu (2009): "Fund of Hedge Funds Portfolio Selection: A Multiple-Objective Approach", Journal of Derivatives and Hedge Funds 15(2), 91-115.

Fernholz, R., R. Garby and J. Hannon (1998): "Diversity-Weighted Indexing", Journal of Portfolio Management, Winter 1998, 74-82.

Grinold, R.C. (1989): "The Fundamental Law of Active Management", The Journal of Portfolio Management, Vol. 15, No. 3: pp. 30-37.

Grinold, R.C. and R.N. Kahn (2000): "Active Portfolio Management", McGraw-Hill, New York.

Harindra, d.S., S. Sapra and S. Thorley (2001): "Return Dispersion and Active Management", Financial Analysts Journal, Vol. 57, No. 5, pp. 29-42.

Hsu, J. (2006), "Cap-Weighted Portfolios are Sub-optimal Portfolios", Journal of Investment Management, 4(3), pp. 1-10.

Kroll, B., D. Trichilo and J. Braun (2005): "Extending the Fundamental Law of Investment Management", JP Morgan Asset Management US Equity Quantitative Research.

Leibowitz, L.M., Emrich, S. and Bova, A. (2009): "Modern Portfolio Management: Active Long/Short 130/30 Equity Strategies", Wiley Finance.

Perold, A.F. (2007): "Fundamentally Flawed Indexing", Financial Analysts Journal, Vol. 63, No. 6, pp. 31-37.

Strongin, S., M. Petsch and G. Sharenow (2000): "Beating Benchmarks", Journal of Portfolio Management, Summer 2000, Vol. 26, No. 4: pp. 11-27.

Stubbs, R.A. (2013): "Consistent Portfolio Management: Alpha Construction", Axioma research paper No. 044.

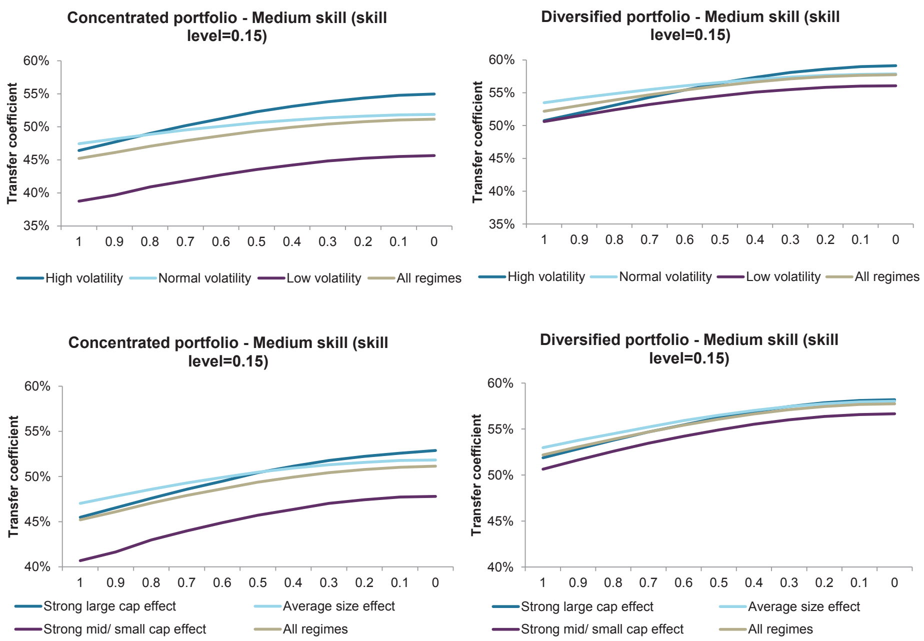
Yu, W. and Y.M. Sharaiha (2007): "Alpha Budgeting – Cross-Sectional Dispersion Decomposed", Journal of Asset Management (2007) 8, 58-72. doi:10.1057.

Appendix

A closer look at the transfer coefficient

Chart A1 examines the impact of volatility and size regimes on the transfer coefficient across different benchmark diversity curvatures[18] for medium-skilled PMs with different conviction levels. We have showed empirically the general linkage between volatility and size regimes. Indeed, the trade-off contours for choice of benchmark curvatures across both regime types are similar in pattern and magnitude.

Chart A1: Impact of volatility (top panel) and size (bottom panel) on transfer coefficient across different benchmark curvatures and portfolio concentrations (medium skill level, IC=0.15)

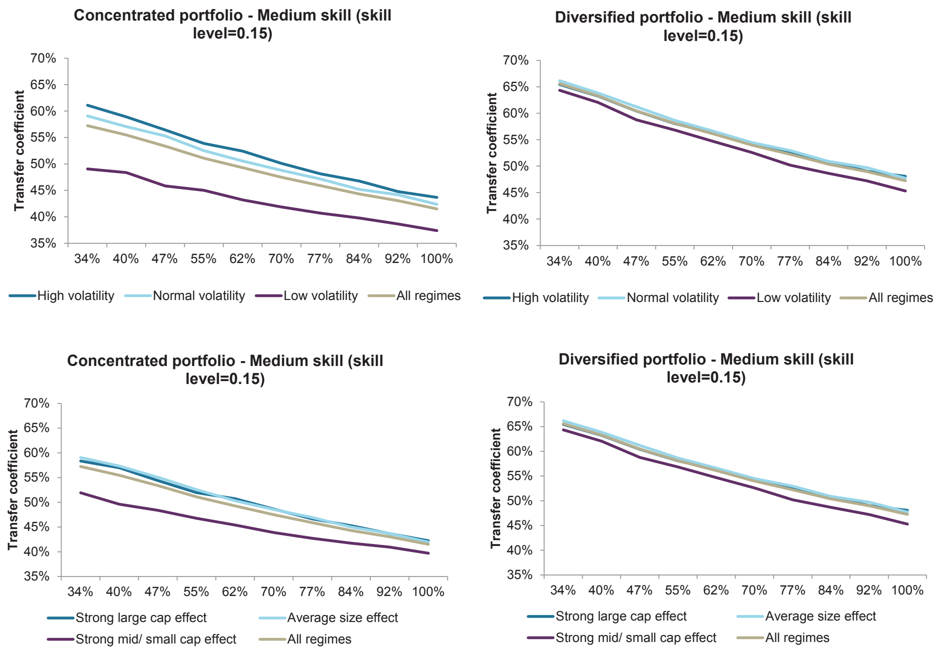


Source: NBIM calculations

For both levels of active risk, we notice that the transfer coefficient improves as the benchmark shifts from cap-weighting (diversity parameter of 1) to equal-weighting (diversity parameter of 0), indicative of better expression of over/ underweight views. We also note that the transfer coefficient is generally lower for a high-conviction PM. In fact, the fall in transfer coefficient is most drastic as the PM increases his/ her active risk during periods of low volatility or strong small/ mid cap effect. This is expected as larger active weights are required for the same level of active risk in such regimes. That is, the PM would need to take on a small set of long stock positions which makes the benchmark less relevant. We also notice that moving away from cap weighting to diversity weighting is most beneficial during periods of high volatility. In our example of European Banks, our overall results across PM skill/style and market states show that the increase in transfer coefficient begins to plateau when the diversity parameter is around 0.3 to 0.4.

18 *We average the transfer coefficient across all proportions of benchmark names for a given diversity parameter.*

Chart A2: Impact of volatility (top panel) and size (bottom panel) on transfer coefficient across benchmark names and portfolio concentrations (medium skill level, IC=0.15)



Source: NBIM calculations

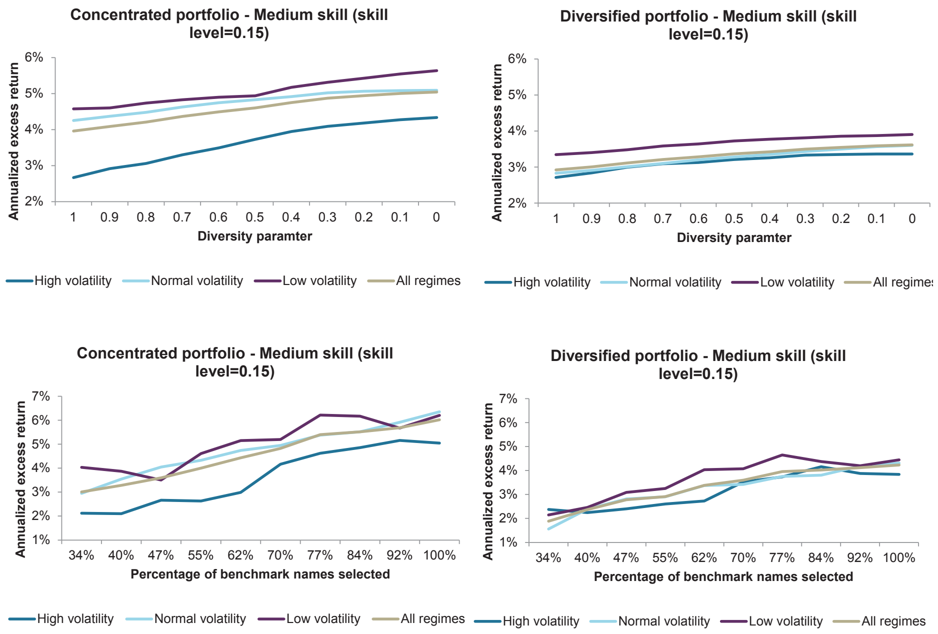
Chart A2 explores the relationship between the transfer coefficient and benchmark names across different volatility and size regimes. We notice that the transfer coefficient falls monotonically with the number of benchmark names due to the increasing difficulty in expressing over/ underweight views. Recall that the names in the benchmark are included in order of their size, which means more of a small cap tail as more names are incrementally added.

From transfer coefficient to excess returns

Continuing with our medium-skilled PM example, we examine the profile of excess returns for choice of benchmark curvature and names under different volatility regimes, and link this to the corresponding transfer coefficient profile. In the same spirit of Clarke et al (2002), we also calculate the excess return of the 'optimal' portfolio[19] in the absence of constraints; that is the excess return when the PM can fully express his/ her relative views. This could shed light on the drivers of excess returns – manager skill, noise that may be generated as a by-product of the long-only constraint and the size of alpha opportunity set.

19 *The 'optimal' portfolio weights are precisely proportional to the alpha scores. The portfolio is then scaled accordingly to achieve the desired tracking error target (8% and 4% for concentrated and diversified portfolios respectively).*

Chart A3: Sensitivity of excess returns to benchmark curvature[20] (top panel) and proportion of benchmark names selected[21] (bottom panel) by portfolio concentrations (medium skill, IC=0.15)



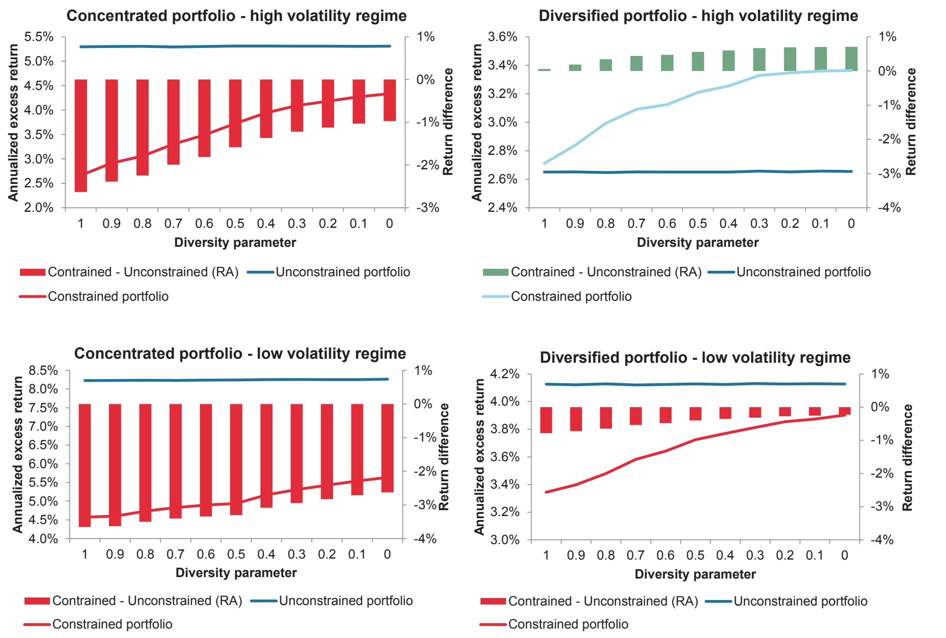
Source: NBIM calculations

We see from the top panels of chart A3 that excess returns, on average, improve across both portfolio concentrations across all volatility regimes as the benchmark shifts from cap weighting to diversity weighting, consistent with the improvement in transfer coefficient and increased potential diversification. We note that the increase in excess return across both portfolio concentrations is most profound in high volatility periods due to the greater improvement in the transfer coefficient (see top panels of chart A1). The bottom panels in chart A3 shows that excess returns generally improve with the number of benchmark names. This is because the diversification effects from a greater number of 'independent' bets more than offset the corresponding fall in transfer coefficient due to the increasing difficulty in expressing over/underweight views.

20 We average the excess returns across all proportions of benchmark names for a given diversity parameter.

21 We average the excess returns across all diversity parameters for a given proportion of benchmark names selected.

Chart A4: Comparing excess returns of constrained and unconstrained portfolios

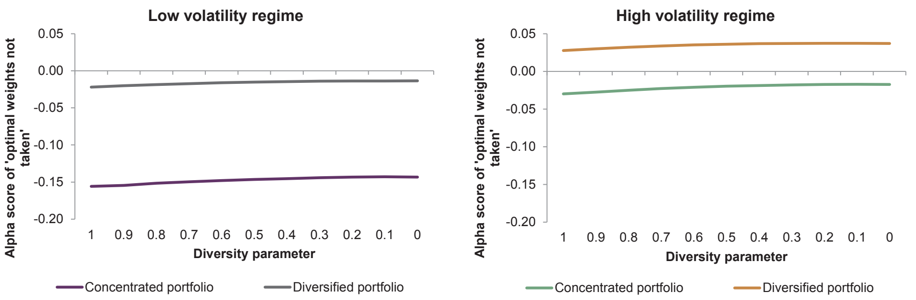


Source: NBIM calculations

Despite having the same active risk, we notice that the excess returns implied from our simulations tend to be higher (lower) during low (high) volatility periods (see chart A3). Chart A4 compares the excess returns of constrained and unconstrained portfolios in different volatility regimes. In low volatility regimes, we observe higher unconstrained portfolio excess returns mainly due to the lower average pairwise correlation between stocks. This implies that the PM can take on a greater number of 'independent' bets and improve his or her risk-adjusted return.

Chart A4 also measures the how close the constrained portfolios are relative to their unconstrained equivalents in excess return terms. Consistent with our findings related to the transfer coefficient, the deterioration in performance is most visible for concentrated portfolios, especially during low volatility periods. We also observe that the constrained returns are much closer to their unconstrained equivalents for diversified portfolios. In fact, the unconstrained diversified portfolio appears to be outperforming its constrained equivalent when volatility is high. This is driven by the higher alpha scores of the constrained portfolios (see right panel in chart A5). In such a scenario, portfolio constraints (size and 'beta' neutrality, and no short sales) are more easily met, hence allowing the alpha maximising PM to outperform the unconstrained gold standard.

Chart A5: Alpha score of 'optimal weights not taken'



Source: NBIM calculations