# SELECTING A RISK-ADJUSTED SHAREHOLDER PERFORMANCE MEASURE

*Christian S. Pedersen and Ted Rudholm-Alfvin*[1]

***ABSTRACT***

The emergence of 'alternative' investment opportunities, the current bear market and the Wall Street analysts' conflict of interest debacle, have put pressure on current investment performance measurement methodologies. We present a survey of classic and modern performance measures and assess them against objective criteria. Depending upon the market, industry or group of assets studied and the preferences of investors, different measures gain favour and we propose key questions to address when selecting an appropriate performance measure. Our arguments are demonstrated empirically for the global financial services sector, for which we document strong evidence in support of using Sharpe Ratio-based measures. As a comparison, we also look at firms listed on the UK Alternative Investment Market (AIM) for which we illustrate a divergence of rankings based on alternative measures. General implications for risk management and asset allocations across different asset classes are discussed.

***Keywords:*** Risk-adjusted Performance, Downside Risk, Asset Management

## *1. INTRODUCTION*

The recent popularity of 'alternative' investment products has contributed to a search for performance measures capable of capturing different risk characteristics other than classic mean-variance based approaches. In particular, the traditional Sharpe Ratio, introduced by Sharpe (1966) has recently been attacked. A good recent survey of arguments and references can be found in the Institutional Investor article by Lux (2002). A scan of popular web-sites (e.g. www.cta-online.com, www.cradv.com or www.worthtrading.com) now reveals a wealth of different risk-adjusted performance measures attempting to meet increasing investor demand for rigour and transparency. This refocusing has been considerably fuelled by the present bear market and dismal performance of numerous fund managers, which accelerated fund withdrawals, and was most emphatically embodied in the Unilever vs. Merrill Lynch case, in which the core arguments centered on basic risk / return analysis.

We thus consider it timely to revisit some of the basic principles of performance measurement, and attempt to answer some key, but typically ignored, questions. Is it really the case that classic mean-variance assumptions underpinning the Sharpe Ratio and other measures are seriously violated over meaningful horizons? Will more complex performance measures actually give additional informational value for many asset classes? And is the focus on formulae distracting us from more practical issues, such as accounting for M&A activity and international or cross-industry differences? In our view, changing fundamentals or moving to more complex modelling techniques, at the potential cost of reduced clarity and ease of application, should come only when strong evidence of inconsistent measurement can be documented.

We survey existing performance measures from academia and the markets, and define objective criteria against which they could be evaluated. In the light of these we assess various alternatives to the classic measures and chart their usefulness given the properties of underlying equity data, investor preferences and market characteristics. To illustrate our points empirically, we use the financial services industry, which has undergone considerable consolidation and been hugely affected by the present bear market, and the Alternative Investment Market (AIM). Specifically, we compare and contrast a Sharpe Ratio-based rankings with more exotic alternatives, using Jarque-Bera tests and rank correlation analysis. Finally, we present a simple rule for locating other asset classes in relation to these examples and discuss the implications for best practice risk and asset management.

The paper is organised as follows: Section 2 presents properties we consider desirable for a performance measure and presents our survey of performance measures. We also characterise the properties of return data for several asset classes, which helps identify when the various measures are likely to be of most use. Section 3 contains our empirical analysis and Section 4 is reserved for our conclusions.

## *2. MEASURES OF RISK-ADJUSTED SHAREHOLDER PERFORMANCE*

There are several different concepts underlying quantitative numerous approaches to measurement of the performance of financial assets but incorporate notions of risk and are defined as a function of historical returns. Naturally, different tasks may call for different analytical approaches (e.g. shareholder value ranking, determining weights in investment portfolios, defining stop-losses etc.). However, to simplify our overview, we find it convenient to work with the following criteria, which we argue any 'best practice' performance measure should satisfy at a minimum:

- **Appropriateness:** The measure captures essential features of the asset return distribution, at a minimum risk and return

- **Foundation:** The measure should have a solid foundation either in finance theory or be a universally applied 'market standard'

- **Clarity:** The measure must be easy to explain to a non-technical individual

More rigorous restrictions can be imposed for special cases (e.g. consistency with a risk-return frontier when applied to asset allocation decisions). For international rankings, which we focus upon in Section 3, one more practical condition is added:

- **International Comparisons:** There must be consistent accounting for international differences without interfering with the fundamentals being assessed

With these properties in mind, we present our survey and discussion of performance measures from applied financial theory and the markets.

## 2.1 *Sharpe and Friends: The Classic Performance Measures*

All mainstream quantitative performance measures from the theory of finance are based on accurately capturing risk and returns. The origins of modern portfolio theory in Markowitz (1952) holds that investors have *mean-variance preferences*, i.e. that the assessment of asset returns only depends upon the average returns $\mu_p$ and volatility

$$\sigma_p = \left[ E(r_p - \mu_p)^2 \right]^{\frac{1}{2}} \quad \textbf{(1)}$$

Building on this, the Capital Asset Pricing Model (CAPM), attributed to Sharpe (1964), Lintner (1965) and Mossin (1969), uses mean-variance preferences combined with standard market conditions to deduce the following relationship between portfolio returns and the market risk premium

$$r_p - r = \alpha_p + \beta_p * (r_m - r) + \varepsilon_p \quad \textbf{(2)}$$

In (2), $r_p$ denotes the portfolio or asset returns, $r_m$ denotes the return on the 'market portfolio', r is the risk-free return, $\alpha_p$ is the excess returns and $\beta_p$ is the systemic risk measure, defined in terms of the covariance between portfolio and market returns. Finally, $\varepsilon_p$ denotes idiosyncratic risk, i.e. risk beyond that which can be attributed to systemic shifts in portfolio value via swings in the market.

Based on such fundamental finance theory, the 'classic' performance metrics are defined as

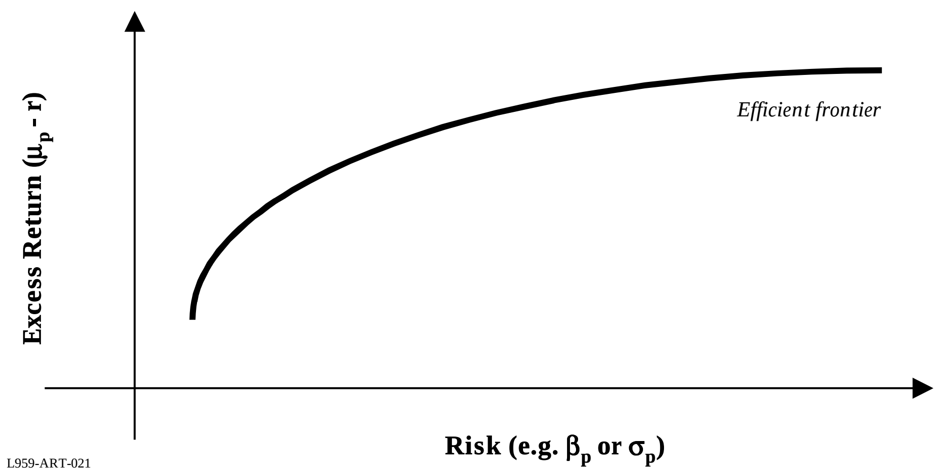
$$\text{Sharpe Ratio} = \frac{\mu_p - r}{\sigma_p} \quad \textbf{(3)}$$

$$\text{Treynor Index} = \frac{\mu_p - r}{\beta_p} \quad \textbf{(4)}$$

$$\text{Jensen's Alpha} = \alpha_p \quad \textbf{(5)}$$

These were introduced in Sharpe (1966), Treynor (1965) and Jensen (1972) respectively, and measure risk-adjusted performance relative to volatility, relative to systemic risk and in absolute terms. Note that the Sharpe Ratio is in fact independent of the CAPM results and market portfolio, $r_m$; this is a key point, which we return to in Section 2.2.1. In addition, the Sharpe Ratio and the Treynor Index have the convenient property of being completely captured by a risk-return frontier, as illustrated in Figure 1, from which one can interpret them as the slope of the efficient line, or *return per unit risk*.

**Figure 1: Illustrative Risk/Return Diagram**



Despite Jensen's Alpha not being consistent with such a boundary, it retains popularity as it gives a '\$- amount' of risk-adjusted out-performance[2]. We thus claim that when mean-variance preferences are supported, our three first principles (appropriateness, clarity and foundation) are supported - i.e. if quantitative techniques are to be used the classic performance measure should be fully applicable in such conditions.

However, mean-variance preferences do not always give a fair description of the financial markets. To see this note that mean-variance preferences hinge on either or both of the following conditions being satisfied:

*Condition 1:* Portfolio returns can be characterised completely by the first two moments[3] (i.e. mean and variance) of the return distribution

*Condition 2:* Investors care only about the two first moments of the return distribution

These conditions, and the CAPM (1), have faced extensive challenges from a wide range of critics, especially on questions regarding appropriateness and the requirement for international comparisons[4]. We now discuss the practical situations where this assumption can be justifiably made and the instances where alternative approaches are called for.

### 2.1.2 Condition 1: Behaviour of Asset Returns in Practice

A consequence of Condition 1 is that the classic measures automatically satisfy appropriateness when returns are normally distributed. In practice there are many asset classes where returns are very seldom normal, due to frequent presence of skewness (third moment effects) and / or fat tails (fourth moment effects, also called kurtosis). Typical drivers of such asset return properties include illiquidity, which create fat tails, and bankruptcy risk, which drive negative skewness, but in some instances specific industry characteristics may be the main factor (e.g. profit-caps for regulated utilities). To illustrate using our data sample of equity returns, UK micro-firms have median skew of 1.15 vs. 0.25 for large OECD equities. Note that boom-and-bust periods leading to high volatility are not in themselves enough to reject the classic approach – most frequently, it is the asymmetric nature of returns, which imply that some risks may not be captured by classic performance measures.

Significant deviations from normality has been explicitly demonstrated for emerging markets (e.g. Estrada (2000), Harvey (1995), Eftekhari and Satchell (1996) and Hwang and Pedersen (2002b)), portfolios with derivatives (e.g. Leland (1999) and Pedersen (2001)) and small companies (e.g. Dimson and Marsh (1998) and Hwang and Pedersen (2002a)). Moreover, return asymmetry is structural in the case of low-grade credit products and start-ups/venture capital, which encompass significant bankruptcy risk. In Section 3, we compute deviance from normality for several asset classes and discuss implications.

It is noteworthy that analysis performed on high frequency return distributions will also fall foul of Condition 1, since they are less affected by the Central Limit Theorem than lower frequency returns[5]. This was demonstrated in the recent article by Dimson and Jackson (2001), who assessed the effect of high frequency data on performance monitoring. Also, where data histories are sparse (such as for micro-firms or in certain emerging markets) or markets thin, one can observe large asymmetries in even low frequency return data. Consequently, whilst the question of whether the classic measures satisfy our principles is always a matter of degree, it is clear that in some cases, underlying data assumptions are undoubtedly violated.

#### 2.1.3. Condition 2: Complexity of Investment Decision-making

In the instances where returns are not well behaved, Condition 2 – the assumption that investors' preferences can be described by quadratic utility - may still justify use of the classic measures. Unfortunately, both theory and empirical analysis challenge this claim. The most frequently cited theoretical criticism holds that the underlying framework, which ensures Condition 2 is satisfied (specifically, you require utility to be a quadratic function of wealth), implies increasing risk aversion, i.e. as investors' get wealthier they get more risk-averse; this is both counter-intuitive and almost universally contradicted empirically.

Perhaps even more damaging critique, however, emanates from a large but less known literature in experimental psychology and management science (see e.g. Fishburn and Kochenberger (1979), Luce (1980), Sarin (1987), Fishburn (1980) or Fishburn (1981) for early works). Based on experimental evidence and surveys, these argued that individuals care more about avoiding 'downside' volatility only (rather than volatility per se) and presented neat and practical risk models accounting for this. Recently, such views have received renewed attention, as they reflect part of popular theories in behavioural finance (see Thaler (1993) for an excellent overview of these), including applications of the infamous Prospect Theory of Kahnemann and Tversky (1979). Moreover, private investors are not the only class of investors exhibiting such behaviour (e.g. Sortino and Price (1994), Balzer (1994) and Damant and Satchell (1996) amongst others). For instance, pension and asset fund managers are typically judged relative to a benchmark and punished more severely when failing to meet target returns (e.g. no bonus, loss of reputation and funds, or even dismissal) than they are rewarded when they beat targets (e.g. proportional bonus). This is bound to lead to loss aversion rather than applied mean-variance optimisation and thus influence trading decisions governing the funding of large volumes of assets on today's financial markets.

Hence, the mean-variance assumption is also tested via Condition 2, in that – although admittedly hard to measure - preferences of investors who are not always only determined by mean and variance. Finally, we argue that financial investors most likely to be trading off returns with downside or extreme risks, rather than volatility, are typically attracted to assets where returns are less well-behaved and such features more explicit. Consequently, Conditions 1 and 2 may go hand in hand for these types of assets, compounding the need for alternative approaches to measuring performance.

### *2.2. Alternative Measures: The Theory*

As alluded to above, critics of volatility as a risk measure, mean-variance, CAPM and the classic performance measures have advanced alternative approaches, which may work well when the underlying assumptions are violated. The most quoted class of risk measures alternative to volatility, which can capture asymmetries in both investor preferences and asset returns, is the Lower Partial Moments, $LPM_p(t, k)$, defined by

$$LPM_p(t,k) = \left[ E(t - r_p)^k \middle| t > r_p \right]^{\frac{1}{k}} \qquad \textbf{(6)}$$

Here, only returns, which are below a general target, t, actually contribute to risk so that, unlike for volatility, large positive deviations are not viewed as 'risky'. Typically t is the risk-free rate (i.e. t = r), mean returns (i.e. $t = \mu_p$) or a specific external benchmark against which performance is measured. This type of measure may fit with the way in which some private equity investments are assessed, and also reflect the reluctance of money managers to take on risk which may mean their overall portfolio value drops below pre-set targets. The variable, k, explains the sensitivity to extreme losses; the higher the value of k, the more extreme losses contribute to risk in relative terms.

Several sub-cases of LPM(t, k) will be familiar, including absolute shortfall (AS),

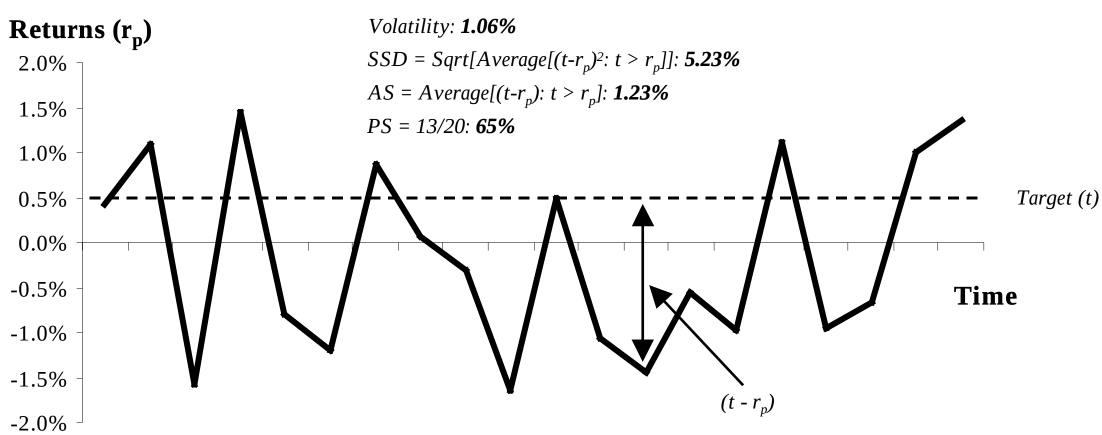
$$AS = LPM_p(t,1) = \left[ E(t - r_p) \middle| t > r_p \right] \qquad \textbf{(7)}$$

and the probability of shortfall, (PS),

$$PS = LPM_p(t,0) = \Pr ob(r_p < t) \qquad \textbf{(8)}$$

These define average returns below a present target, and the chance of such a shortfall occurring, respectively. Figure 2 illustrates the calculation of downside risk measures. Note that AS (7) is the *coherent* measure of Artzner, et al. (1997); also, when t = Value-at-Risk, then (8) denotes the associated confidence level (e.g. 95%). Asset selection models, building on mean-risk decisions using AS and PS as risk measures, have been developed by Kijima and Ohnishi (1993) and Roy (1952), respectively. However, the analytical restrictions prevent CAPM-equivalents to be derived, thus not admitting comparable versions of Treynor's Ratio (4) and Jensen's Alpha (5) to be defined with sufficient foundation.

**Figure 2. Sample Computation of Downside Risk Measures**




L959-ART-021

The natural risk-adjusted alternatives to the Sharpe Ratio based on AS and PS are, unsurprisingly, the Return on Absolute Shortfall (ROAS)

$$ROAS = \frac{\mu_p - r}{\left[E(r - r_p) \middle| t > r_p\right]} \qquad \textbf{(9)}$$

and the Return on Probability of Shortfall (ROPS)

$$\text{ROPS} = \frac{\mu_p - r}{\text{Prob}(r_p < t)} \qquad \textbf{(10)}$$

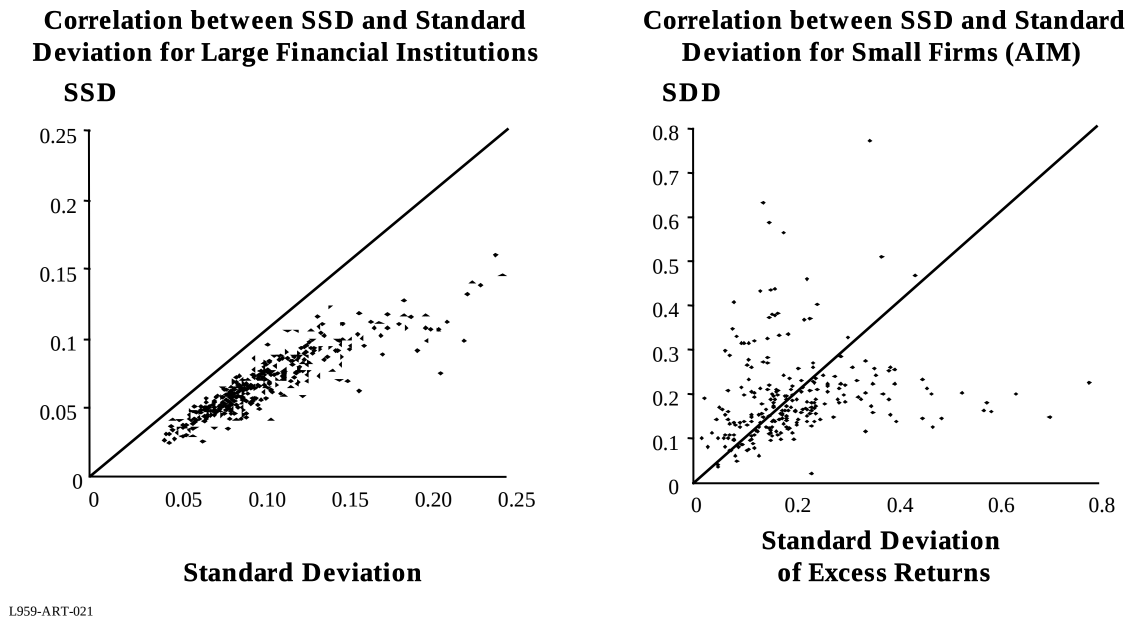
The convenience of the risk-return frontier can be maintained by plotting Figure 1 with the respective risk measures on the x-axis in place of $\sigma_p$ or $\beta_p$, thus allowing intuitive trade-off analysis to be performed. One can thus argue that in cases where the Sharpe Ratio is not suitable, appropriateness and foundation can be preserved.

One risk measure, which sits within the class (7) and has received particular support is the Semi-Standard Deviation (SSD) defined as

$$\text{SSD} = \text{LPM}_p(r,2) = \left[ E(r - r_p)^2 \middle| r > r_p \right]^{\frac{1}{2}} \qquad \textbf{(11)}$$

In practice, this offers a 'happy medium' between the extremes of PS and volatility, as the size of drops affect risk to a similar degree than volatility, but only those returns below the risk-free rate contribute (i.e. t = r and k=2 in (6)). As illustrated in Figure 3 below, SSD and volatility differ when returns are 'non-normal': In the case for large financial institutions, SSD and volatility are highly consistent whilst, for small company (AIM) stocks, no relationship is apparent. This has obvious consequences for any use of risk measures (e.g. asset allocation or hedging), and also illustrates why individuals with different preferences may view some assets similarly but others completely independently. In particular, risk-adjusted performance will be highly dependent upon choice of risk measure for AIM firm but not for financial services; we return to this point in Section 3.

**Figure 3: Asymmetric Returns and Deviation of Risk Measures**




Moreover, the theoretical foundations of an "SSD – CAPM", have been firmly established; in this, mean-SSD preferences replace mean-variance preferences (see for instance Hogan and Warren (1972), Bawa and Lindenberg (1977), Harlow and Rao (1989) or Satchell (1996). This leads to an equation exactly like that for the CAPM (1), except that the systemic risk measure, $\beta_p^{SSD}$, is defined in terms of the 'semi-covariance' with the market, i.e. only co-movements between market and portfolio returns in times when market under-performs the risk-free rate, contribute to systemic risk. Consequently, three measures emerge, which are in direct parallel with the classic measures in (3), (4) and (5):

$$\text{Sortino Ratio} = \frac{\mu_p - r}{SSD} \quad \textbf{(12)}$$

$$\text{Treynor Index (SSD)} = \frac{\mu_p - r}{\beta_p^{SSD}} \quad \textbf{(13)}$$

$$\text{Alpha(SSD)} = \alpha_p^{SSD} \quad \textbf{(14)}$$

Here the superscripts SSD denotes a co-efficient in the version of equation (1) corresponding to the SSD-CAPM rather than CAPM.

The Sortino Ratio was introduced by Sortino and Price (1994), and Pedersen and Satchell (2002) proved that the risk/return frontier, when risk is defined by SSD (11), exhibits the same desirable convexity properties of the traditional mean-variance frontier, thus rendering it amenable for portfolio analytics. Variants of the Treynor Index (SSD) and Alpha (SSD) were applied by Henriksson (1984) and Henriksson and Merton (1981) to measure the performance of market timers in Bull and Bear markets. Moreover, their statistical properties were compared to their CAPM counterparts (Treynor Index (4) and Jensen's Alpha (5)) in Pedersen and Satchell (2000).

When bearing in mind our three first criteria, these are strong alternatives to the classic performance measures in the case of mean-variance assumptions breaking down. As we now discuss, however, the additional principle of international comparison adds a new angle to the problem and puts a question mark against measures, which are CAPM or SSD-CAPM dependent.

### 2.2.1 Reservation over CAPM-Dependent Measures in International Comparisons

The Sharpe Ratio (3), Sortino Ratio (12), ROAS (9) and ROPS (10) are defined only on historical time series of portfolio returns. However, for two versions of the Treynor Index ((4) and (13)) and the two versions of Jensen's Alpha ((5) and (14)), the choice of portfolio driving market returns, $r_m$, is fundamental (see Roll (1977) or Jagannathan and McGratten (1995)). For the purposes of performance rankings internationally, the problem of choosing the 'right' market portfolio has an additional twist.

Suppose one wants to compare the performance of a large German firm (Germany Inc) with a US firm (US Inc). There are three relevant proxies for the market (S&P 500, Dax, World Market Index). Choosing the former implies that the low correlation between the S&P 500 and Germany Inc leads to a low systemic risk, $\beta_p$, for Germany Inc and an upwardly biased measure of its performance relative to US Inc. The same happens with US Inc's performance relative to Germany Inc if the Dax is chosen. Finally, if a World Market Index is used, the performance numbers are comparable, but both are arguably hugely biased upwards, since neither firm will have a strong correlation with the world market and thus unrealistically small systemic risk.

Although true for all international comparisons, this phenomenon is particularly well documented for emerging-market stocks. Although generally considered high-risk, Harvey (1995 and 2000) has demonstrated that when applying the CAPM with market portfolio set at the MSCI World Index, they appear virtually risk-less, and are high performers on a risk-adjusted basis when using Treynor (4), Jensen (5) and their SSD counterparts (13) and (14). In such conditions, the Sortino Ratio (12), ROAS (9) or ROPS (10) are more appropriate as they are independent of the market and are sensitive to asymmetry in returns.

### *2.3 Alternative Measures: Market Favourites*

The Sharpe Ratio undoubtedly remains the most popular performance measure in practice. Nonetheless, other groups of measures also feature in everyday analysis. Total Shareholder Return is very simply the overall return over the period studied (including share performance as well as dividends), i.e. in our notation

$$TSR = r_p \quad \textbf{(15)}$$

This measure is unique in that no risk-adjustments are made. To place it within the previous discussion, it would clearly have clarity but is only convenient for an investor who is neutral (does not care) about risk and captures the first moment only in the return distribution. Consequently, a large question mark has to be put against both appropriateness and foundation.

Much in the spirit of the Lower Partial Moment (6) and related downside risk and performance measures, a series of industry measures pay attention to the maximum draw-down (MDD). Maximum draw-down is defined as the maximum sustained percentage decline (peak to trough), which has occurred in the stock within the period studied.
Two popular versions of these are the Calmar Ratio (CR), which is defined as

$$CR = \frac{r_p}{MDD} \qquad \textbf{(16)}$$

and the Sterling Ratio (SR),

$$SR = \frac{r_p}{\text{Average MDD} + 10\%} \qquad \textbf{(16)}$$

In the latter, Average MDD is taken over typically 2 to 5 years. These two measures, and other variants, are discussed and assessed in Kestner (1996) and Young (1991). It is obvious that, whilst no explicit support from theory is present, CR and SR are closely related to ROAS (8) and ROPS (9), since both are driven exclusively by below-target or negative returns. Finally, the Information Ratio (IR) is defined by

$$IR = \frac{r_p}{\sigma_p} \qquad \textbf{(17)}$$

This measure is almost identical to the Sharpe Ratio (3), except the numerator is total, rather than excess, returns. For domestic comparisons, it thus has the same benefits and shortcomings as the Sharpe Ratio, with additional lack of theoretical foundation. For international comparisons, however, these 'market favourites' lack a particular quality, which constitutes our second main observation.

### 2.3.1 Reservation over Interest Rate Adjustments in International Comparison

We make the case that by not adjusting for the risk-free interest rate, and thus using absolute rather than excess returns, the above measures (15) to (18) face potential problems when applied to international comparisons. Using similar nomenclature as our previous example, suppose US Inc has an absolute return, $r_p$, of 12% and excess returns of 9% (i.e. risk-free rate r = 3%) whilst Turkey Inc has absolute return of 60% and excess returns of 5% (i.e. an inflation-fuelled risk-free rate r of 55%). Under the above measures, Turkey Inc is likely to strongly outperform (indeed it is 5 times as 'good' as US Inc under IR (16)), as they fail to adjust for differences in local risk free rates. This is not an issue with the measures outlined in Section 2.2, including the Sharpe Ratio (3) and the Sortino Ratio (12).

It may be argued that by subtracting the risk-free rate, r, in the numerator of the relevant measures, one solves the problem. In fact, this would lead to IR becoming the Sharpe Ratio and CR and SR closely approximating ROPS and ROAS respectively. Hence, there is strong consistency between theory and practitioner approaches, albeit 'extreme' losses carrying more explicit weight in the measures advanced by the latter.

## *2.4 Summary*

To summarise, we note that in theory all measures can be applied when mean-variance preferences are observed. Thus, it makes little sense to go beyond Sharpe, Treynor and

Jensen since no extra information will be provided. We also argue that the clarity principle is satisfied by all the measures we have discussed so far. However, foundation in theory and/or markets, appropriateness where mean-variance preferences are not justified, and suitability for international comparisons, vary greatly. Table 1 below summarises these.

**Table 1. Main Differences in Performance Measures**

| | APPROPRIATENESS IN NON-MV SPACE | FOUNDATION THEORY | FOUNDATION MARKET | INTERNATIONAL COMPARISONS |
|---|---|---|---|---|
| SHARPE RATIO | ✗ | ✓ | ✓ | ✓ |
| TREYNOR INDEX | ✗ | ✓ | ✓ | ✗ |
| JENSEN'S ALPHA | ✗ | ✓ | ✓ | ✗ |
| ROAS | ✓ | ✓ | ✗ | ✓ |
| ROPS | ✓ | ✓ | ✗ | ✓ |
| SORTINO RATIO | ✓ | ✓ | ✗ | ✓ |
| TREYNOR (SSD) | ✓ | ✓ | ✗ | ✗ |
| JENSEN (SSD) | ✓ | ✓ | ✗ | ✗ |
| TSR | ✗ | ✗ | ✓ | ✗ |
| INFORMATION RATIO | ✗ | ✗ | ✓ | ✗ |
| CALMAR RATIO | ✓ | ✗ | ✓ | ✗ |
| STERLING RATIO | ✓ | ✗ | ✓ | ✗ |

✗ Denotes relative weakness ✓ Denotes relative strength

Clearly individual applications will help determine the final choice of particular measure to use and these criteria are of course open to some degree of interpretation. Although by no means set in stone as a selection rule, Table 1 does however inform us of key questions to ask when selecting a quantitative performance measurement philosophy:

- Are mean-variance assumptions satisfied?
  - What data frequency is used for the analysis?
  - Are the markets covered liquid?
  - Is there a sufficient data history to carry out statistical investigations?
  - Is there evidence of the presence of other risk characteristics (e.g. downside risk) among investors?

- Does the particular application of measures permit an approach not necessarily consistent with fundamentals?

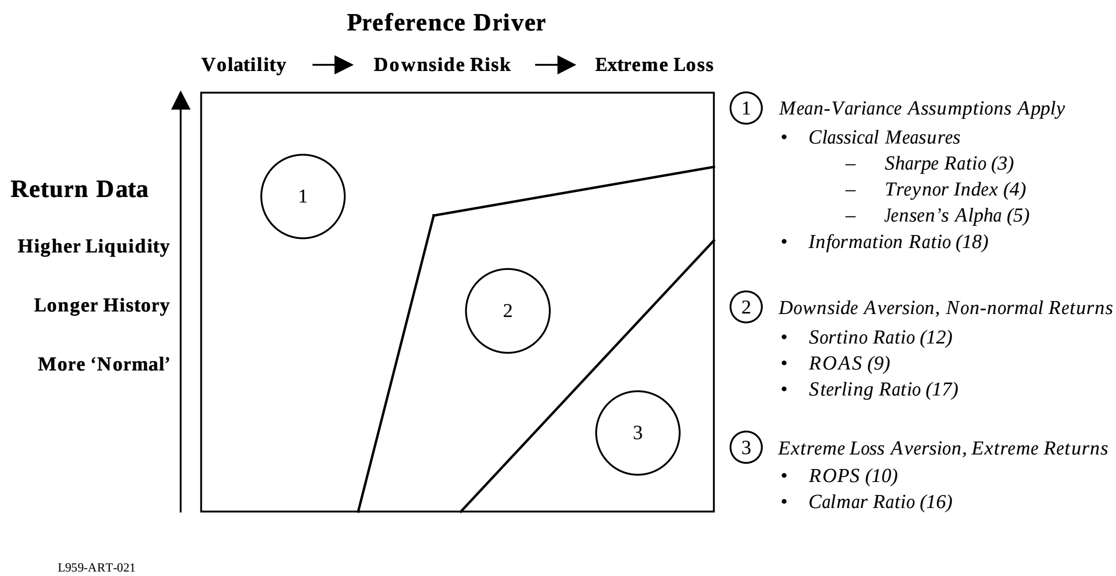
- Can international performance comparisons be made?

- Are returns measured in absolute terms or relative to a benchmark?

- Is there evidence of differences in rankings implied by alternative measures?

The first bullet-point, which addresses the degree to which the traditional measures can be meaningfully applied, requires deepest analysis and has the biggest consequence for the applications of performance measures. In Figure 4, we have expanded the second column of Table 1 to summarise this complex problem and give our suggestions for a broad selection rule that can sensibly be applied.

**Figure 4. Performance Measures and the Mean-Variance Assumptions**



When returns are approximately normally distributed, the classic measures can be applied and satisfy all types of investor preferences. Also, when investors truly are only concerned with volatility and are not especially sensitive to losses, the classic approach applies regardless of the return distribution. This is captured by Area 1. However, applying the Sharpe Ratio in Area 2, in which for instance the typical class of small cap stocks being assessed by a downside risk-averse investor is located, would fail to capture some of the risk to the investor; an asymmetric risk measure is thus recommended. This is even more pronounced in Area 3, which would cover private equity or derivatives investment for investors trading off extreme loss avoidance with expected return. Calculating, for instance, a risk budgeting mechanism across a portfolio spanning all areas and accounting for marginal risk contributions by asset class, would thus be somewhat complex. The key question would change from: *How much risk?* to: *How much of what type of risk?*

## *3. EMPIRICAL ANALYSIS*

The arguments outlined above provide a set of criteria, which can be used to guide decisions on performance measurement methodology. We now illustrate how these can work in practice. As referred to in Section 2.1, significant evidence exists across numerous markets to challenge the assumption that return data is sufficiently 'well-behaved' to justify use of the classic measures. Table 2 below summarises the results of a 'normality test' using the Jarque-Bera statistic (defined in Appendix 5.1) on different asset classes.

The results give a clear signal of the degree to which the different assets conform to mean-variance assumptions. Highly liquid investment grade bonds and OECD equity indicies (despite the recent turbulence) reject normality less than 20% of the time and the high-volume on FX markets also appear 'normal'. At the other end of the spectrum, we find more illiquid assets such as emerging market stocks, hedge funds, low-grade bonds and the ill-fated telecoms sector, all of which have rejection rates above 67%. These are consistent with result of Hwang & Pedersen (2002a and 2002b), who found ten-year rejection rates of 75-80% for both emerging markets stocks and small UK firms (in the FTSE SmallCap, larger than the AIM firms used in this article).

**Table 2. Jarque-Bera Test for Asset Returns**[7]

| ASSET CLASS | JB REJECTION* | MEDIAN* | MEDIAN SKEW |
|---|---|---|---|
| **Investment Grade Bonds** | 14.3% | 2.1 | -0.27 |
| **OECD Market Indices** | 15.8% | 2.2 | -0.33 |
| **FX Rates** | 26.1% | 1.1 | 0.33 |
| **Large OECD Equities** | 36.9% | 3.3 | 0.25 |
| **Large Financial Institutions** | 37.9% | 3.7 | -0.04 |
| **Emerging Market Country Indices** | 43.8% | 5.3 | 0.47 |
| **OECD Mid Cap Equities** | 47.5% | 5.4 | 0.35 |
| **Financial Institutions** | 49.8% | 5.9 | 0.24 |
| **Gas** | 52.4% | 6.5 | 0.47 |
| **Mid Cap Financial Institutions** | 53.9% | 6.9 | 0.35 |
| **Electricity** | 62.0% | 11.6 | 0.35 |
| **Low Grade Bonds** | 66.7% | 18.6 | -0.26 |
| **Emerging Market Stocks** | 71.0% | 20.8 | 0.93 |
| **Telecoms** | 79.3% | 41.1 | 1.20 |
| **Hedge Funds** | 84.6% | 63.5 | -1.05 |
| **Tiny Firms (AIM)** | 89.0% | 67.4 | 1.15 |

* % assets rejecting Jarque-Bera test at 5% level; median value of Jarque-Bera statistics

Hence, traditional measures can be expected to deal comfortably with most liquid and unregulated assets, whilst we expect the size of their bias on several key emerging asset classes and those which exhibit high bankruptcy rates, to render their application inappropriate. For these, alternative investigation is needed for identifying candidate performance measures. Finally, it is worth mentioning that analysis such as that supporting Table 3 may have to be re-calculated regularly to update methodology and capture market developments (e.g. convergence between new and established markets).

We now illustrate in more detail the type of calculations, which can help distinguish between alternative performance measures. Our focus is on the financial services industry and the Alternative Investment Market (AIM), but the approach is applicable to any underlying return data.

### *3.1 Financial Institutions Data*

The financial services sector has experienced severe turbulence during the present Bear market and also seen a high level of consolidation. Moreover, events such as September 11, product mis-selling scandals in the UK and the present plight of many European insurers have contributed to (and largely been driven by) huge downside returns. However, the largest firms‘ stocks have remained highly liquid and so this sector provides an interesting case study in the light of our previous discussion.

The Shareholder Performance Index (SPI)[8] is a ranking of the world’s leading financial institutions by their risk-adjusted performance for shareholders, designed to provide executive managers with a tool for judging their firm’s relative shareholder performance. The index universe includes the 400 largest financial services firms by market capitalisation and draws from all general areas of financial services (e.g. banking, insurance, brokerages, asset managers) as well as specialists, such as Capital One (consumer credit), Charles Schwab (brokerage) and First Data (processor)). Hence, international comparisons are included, and considerable M&A activity is accounted for (487 mergers in the last 5 years covering more than $1.5 TN in transaction value). It is also important not to mix currency fluctuations with share-price performance and total returns were thus measured in local currency and effectively adjusted for inflation differences by using a local three-month (or nearest maturity) treasury rate in each home country.

The underlying measure of the SPI is a 5-year Sharpe Ratio (3) measured on monthly data. The median Sharpe Ratio for the universe of 400 institutions is then used to define a firm’s performance index (SPI)

$$\text{SPI(firm)} = \frac{\text{Sharpe Ratio (firm)}}{\text{Median Sharpe Ratio (universe)}} \text{ x } 100 \qquad \textbf{(19)}$$

Given the requirement of making international comparisons, the Treynor Index (4) and Jensen's Alpha (5) lost favour in comparison with the Sharpe Ratio (3). Also, we note that other measures based on the distribution of the Sharpe Ratio in the universe covered could be applied (i.e. quartiles or mean-volatility standardisation). However, the present re-scaling seemed most appropriate for our purposes.

### 3.1.1. Testing the Mean-Variance Assumptions for Financial Institutions

To test if the underlying return data justifies the mean-variance assumptions we first refer back to the Jarque-Bera results in Table 3. Of the 400 companies, 50% reject the assumption of normal returns at the 5% level but the median Jarque-Bera value was just 5.9 and financial institutions had the lowest median skew of all samples tested. Hence, this shows relatively weak deviation from mean-variance. Whilst this may seem surprising given recent market activity in this sector, it is a sobering reminder that one should not necessarily discard standard approaches even in hugely volatile periods.

Naturally, even within the large liquid financial sector individual examples of large deviations from normality can be observed. The largest JB value in our sample corresponded to Mitsubishi Tokyo Financial Group (JB = 3,332), which underwent major corporate reorganisation by merging two entities to form a third during our sample period. White Mountains Insurance Group (JB= 1,317) had the second largest value, largely driven by single > 60% return on one day when Berkshire Hathaway supplied cash for them to acquire the US operations of CGNU. Similar 'one-off' stories characterise most other large deviations from normality, and on the whole the data appears to behave reasonably enough to fully support the traditional approach.

The second direction from which mean-variance assumptions can be challenged is to look at investor preferences, which of course are more difficult to infer. We argue that the typical investors in large, liquid equities such as those covered by the SPI are likely to be more traditional buy-and-hold strategists or investors seeking to diversify across industries and/or regions, whilst the more exotic risk-takers would move towards less well-behaved asset classes.

Combining these observations, we advance the view that mean-variance assumptions are reasonable approximations to use and the large financial services firms are amongst the classes of asset likely to be located towards the upper left corner of Figure 3. This would thus indicate that classic and in particular Sharpe Ratio-based measures are appropriate for this segment.

### 3.1.2. Divergence Between Alternative Performance Measures

To add more evidence to support our view, we note that since the classic framework is automatically implied when mean-variance assumptions hold, differences in using alternative approaches diminish in such conditions. In practice, this is typically manifested in extensions of the CAPM not explaining underlying data better than the traditional CAPM (1) - for instance, see Harlow and Rao (1989) or Eftekhari and Satchell (1996) for emerging markets. Alternatively, it can be shown by demonstrating there is little statistical difference between asymmetric risk measures and those implied by mean-variance (e.g. see Price et al. (1982) or Chow and Denning (1994)).

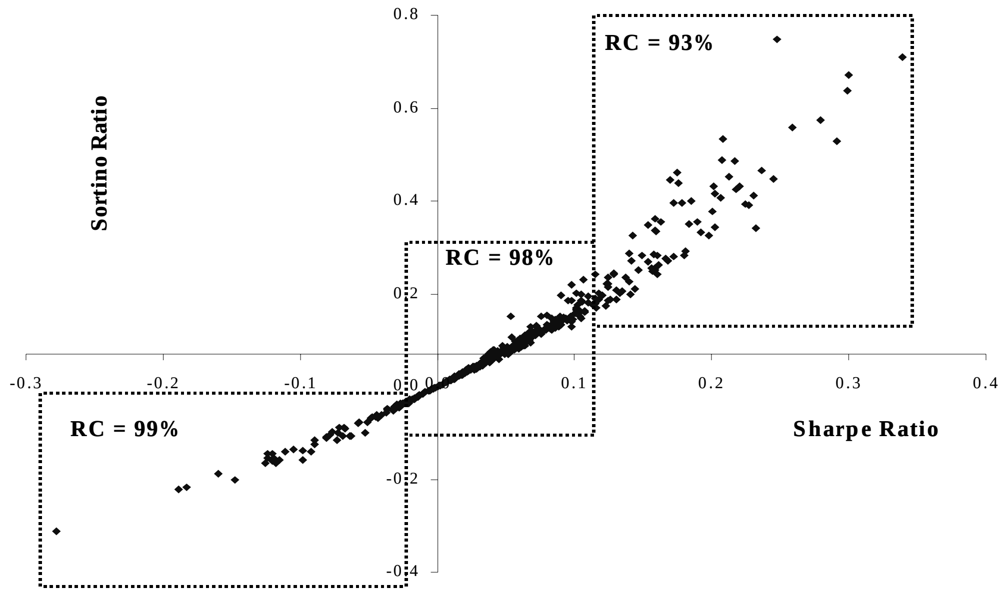
Employing rank correlation analysis, we have examined the potential additional value from alternative measures for financial institutions. Technically defined in Appendix 5.2, rank correlation calculates the similarity between two rank orderings, as opposed to standard correlation analysis, which measures linear correlation. We have thus calculated the SPI, Sortino Ratio, ROAS, ROPS, Information Ratio, Sterling Ratio and Calmar Ratio on our 400 financial services firms and estimated the pair-wise rank correlation for all measures. Note that measures like Treynor Index (4) and Jensen's Alpha (5) are not included as they are derived using the CAPM-based approach requiring market portfolios which, as discussed in Section 2.1.4, is not appropriate for international comparisons

The results, given in Tables A1 in Appendix 5.2, are striking. In each of our three quartile-based subdivisions (top, middle two, bottom), the Sharpe Ratio (or SPI) ranking is on average over 70% correlated with rankings based on other performance measures. A smaller correlation exists between SPI and the extreme downside practitioner ratios (Sterling and Calmar), particularly where the absolute values of the indices are small, and also with ROPS (10) for the top quartile of performers. The correlation between SPI and Information Ratio (16) suggests there is some loss in accuracy when the interest differentials across countries are ignored, especially for low performers (which are more likely to be located in high-inflation environments). Nonetheless, all performance measures analysed are remarkably consistent in producing a shareholder ranking and across the whole sample, each pair of measures have rank correlation over 80%.

Moreover, as is demonstrated by Figure 5, the Sharpe Ratio-based ranking is almost identical with the Sortino Ratio, the most-used downside performance measure. When reading this (and Figure 6), 'RC' denotes rank correlation.

**Figure 5. Comparison of SPI and Sortino Ratio:**

**Financial Institutions**



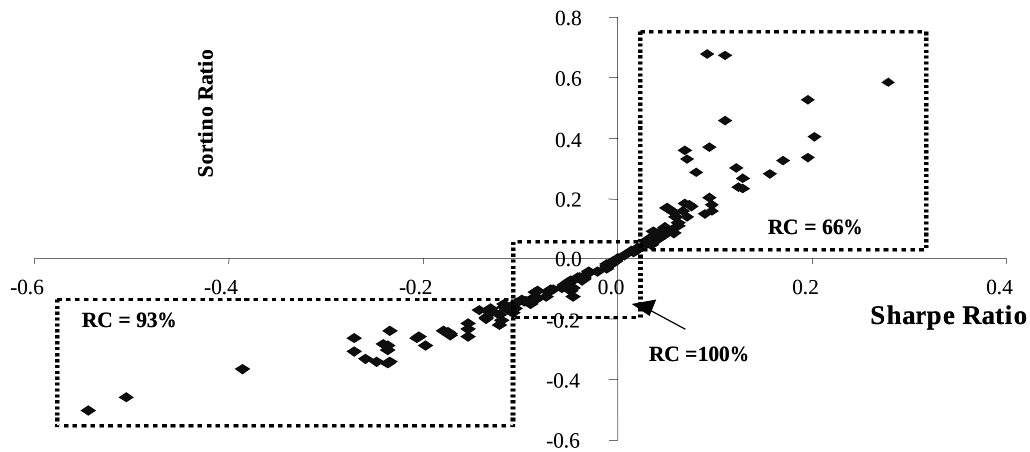

This essentially means that applications using either measure will give very consistent outcomes (e.g. hedging, asset allocation or risk-adjusted performance measurement), although at the extreme ends of the spectrum one should still take care.

Our earlier assertions appear strongly supported, in that there would be little gain in abandoning classic measures. The Sharpe Ratio-based approaches will give an almost identical ranking, capture the necessary features of the data, and thus – by retaining traditional support and analytical convenience – remain appropriate for measuring the performance of the top 400 financial services firms.

### *3.2 Comparison with the AIM data*

To compare with the global financial services we focus on the bottom candidate in the Jarque-Bera tests in Table 3, the micro-firms from the UK's Alternative Investment Market (AIM). The AIM contains firms in the UK, which are too small to enter the FTSE All-Share, and typically covers 2-3% of total UK market capitalisation. They are typically successful start-ups or mono-line or specialist firms with highly concentrated business models and are thus heavily industry-sensitive. Combining this with the obvious bankruptcy risks and thinly traded stocks, we have the key drivers of the high skew and median Jarque-Bera statistic of 67.4 reported in Table 3.

**Figure 6. Comparison of SPI and Sortino Ratio: Tiny Firms (AIM)**




Making the assertion that the mean-variance assumptions are fundamentally violated in this case, we have repeated the above analysis. We have plotted the Sharpe-Sortino comparison diagram in Figure 6 and computed rank correlation matrices for all measures (see Appendix 5.2, Table A2). From Figure 6, it is clear that for outperforming stock the measures diverge considerably. However, divergence is far more acute as we consider the set of measures as a whole in Table A2. Firstly, the Sterling and Calmar Ratio rankings, which are most sensitive to extreme loss, are highly correlated. However, their correlation with other rankings dips below 5% in some instances, in particular in the lowest quartile. Thus, in contrast with the large financial institutions, the return asymmetry of the AIM firms breaks the consistency in measurement of performance.

Interestingly, the Total Shareholder Return, which ignores the risk-free rate, roughly follow Sterling and Calmar (also not adjusting for changes in interest rates), whilst ROAS relates to SPI and the Sortino Ratio, which do make such adjustments. Hence, the role of the interest rate, as described in Section 2.3.1 for international comparison, is significant even when judging performance domestically - the mere question of whether absolute or relative returns are used may distort results. To give a few particular examples to highlight the differences due to skews, we include a sample of results across five individual companies in Table 3 below.

**Table 3. Examples of Divergence in Alternative Risk Measures**

**Comparison of Performance Metrics and Ranks by Firm**

| NAME | AVG MONTHLY EXCESS RETURN | ST DEV OF EXCESS RETURN | SSD OF EXCESS RETURN | SHARPE (3) | | SORTINO (12) | | ROPS (10) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | RATIO | RANK/ 254 | RATIO | RANK/ 254 | RATIO | RANK/ 254 |
| Heath (Samuel) | 1.5% | 5.0% | 2.7% | 29% | 1 | 53.4% | 10 | 153% | 43 |
| Yeoman GP | 8.8% | 95.5% | 13.0% | 9% | 38 | 67.8% | 5 | 111% | 6 |
| Parallel Media | 7.7% | 108.9% | 23.4% | 7% | 54 | 32.9% | 25 | 58% | 8 |
| West Bromwich Albion | -1.7% | 6.4% | 5.6% | -27% | 248 | -30.6% | 243 | -61% | 171 |
| Gaming Insight | -3.4% | 31.4% | 16.0% | -11% | 187 | -21.2% | 214 | -30% | 234 |

Note that Yeoman Group would not have been an out-performer under the Sharpe Ratio but when a very low downside deviation is incorporated the performance looks much better under both Sortino and ROPS. Conversely, the top Sharpe Ratio-ranked firm (Heath Samuel) drops to 43rd in RPOS! Conclusions based on applications of these measures would thus differ greatly with major consequences for risk and asset management.

## *4. CONCLUSION*

We have shed light on key aspects of quantitative performance measurement and suggested basic criteria for clarifying appropriate use of a range of alternative approaches, illustrating our proposition using equity return data from global financial services institutions and UK micro-firms. It is strikingly apparent that whilst some asset types require careful treatment in terms of the performance measures used in different applications, others have returns aligned with mean-variance preferences and react similarly to different approaches.

Our analysis sheds light on several fundamental issues for risk and asset management. Firstly, risk management across a range of investment products typically requires quantification of the impact of different portfolios on overall returns. Tracking performance of portfolio segments will be tricky as marginal distributions may differ greatly from the portfolio distribution. Both institutional investors and banks need to invest in methodology and technology as 'alternative' investments gather momentum and risk budgeting rules such as defined in Rahl (2000) and Sharpe (2002) for mean-variance assets will have to be extended to cover more exotic investments.

Secondly, when building asset allocations, the classic mean-variance optimisation implicitly maximise the Sharpe Ratio (3) and failure to account properly for violation of the underlying assumptions can have a huge impact on final allocations (see for instance Eftekhari et al. (1998) and Damant and Satchell (2002)). The result in Figure 5 indicates that high performance assets, which typically get assigned the highest investment weights, imply the largest divergence between measures. This only exaggerates the potential damage from erroneous application.

Familiarity with these issues is of increasing importance to the investment management community as margins and volumes are currently squeezed and we expect significant cost savings can be realised by realigning risk management efforts across the asset classes. Furthermore, tightening reporting requirements (e.g. the revised Basel Accord for banks) means that investors will be more informed about risk-adjusted performance matching their own levels of "risk-aversion". This will pressure sophisticated money managers to structure increasingly customised portfolios and consequently increase return on customer relationship and marketing investment.

## 5. APPENDIX

### 5.1. Jarque-Bera Tests

The Jarque-Bera statistic given by the following expression

$$JB = n\left[\frac{skewness^2}{6} + \frac{(kurtosis - 3)^2}{24}\right]$$

where *n* is the number of observations in the sample period. This statistic has a chi-squared distribution (with 2 degrees of freedom) under the null hypothesis of normality. Note that a normal distribution (indeed any symmetric distribution) has skewness = 0 and the kurtosis of a normal distribution is 3.

### 5.2. Spearman's Rank Correlation

Suppose you have a set of firms (1,2,3,4,….,n), which have been ranked by two regimes (x and y). Let x(i) and y(i) be the rank (i.e. 1 the highest and n the lowest) of firm i in the two regimes respectively and define $d_i$ = x(i) - y(i) as the distance between these rankings. Then, Spearman's rank correlation is given by

$$\text{RankCorrelation} = 1 - 6 \times \left[\frac{\sum_{i=1}^{n} d^2_i}{n\left(n^2 - 1\right)}\right]$$

The following two pages contain the results of rank correlation estimates for the SPI and AIM data. (Results for the other asset classes are available upon request from the authors.) In addition to results for the entire samples, these are presented by quartiles (top, middle two and bottom). In each case the quartiles are worked out based on a ranking of the measures along the horizontal axis – as these differ by measure, note that the matrix is not symmetric.

## TABLE A1. RANK CORRELATION OF PERFORMANCE MEASURES (SPI)

### Rank Correlation Coefficient for SPI400 (Financial Firms)

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=400** | | | | | | | | |
| SPI | 0.996 | 0.997 | 0.974 | 0.904 | 0.864 | 0.860 | 0.853 | 1.000 |
| TSR | 0.848 | 0.849 | 0.821 | 0.934 | 0.992 | 0.986 | 1.000 | |
| Calmar Ratio | 0.854 | 0.856 | 0.813 | 0.941 | 0.994 | 1.000 | | |
| Sterling Ratio | 0.859 | 0.862 | 0.820 | 0.945 | 1.000 | | | |
| Information Ratio | 0.901 | 0.902 | 0.876 | 1.000 | | | | |
| ROPS | 0.976 | 0.975 | 1.000 | | | | | |
| ROAS | 0.998 | 1.000 | | | | | | |
| Sortino Ratio | 1.000 | | | | | | | |

### Rank Correlation Coefficient for SPI400 (Financial Firms) Top Quartile

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=100** | | | | | | | | |
| SPI | 0.933 | 0.944 | 0.635 | 0.898 | 0.792 | 0.815 | 0.726 | 1.000 |
| TSR | 0.545 | 0.525 | 0.550 | 0.669 | 0.826 | 0.720 | 1.000 | 0.541 |
| Calmar Ratio | 0.708 | 0.680 | 0.493 | 0.849 | 0.918 | 1.000 | 0.755 | 0.736 |
| Sterling Ratio | 0.697 | 0.684 | 0.562 | 0.836 | 1.000 | 0.923 | 0.851 | 0.717 |
| Information Ratio | 0.774 | 0.768 | 0.516 | 1.000 | 0.810 | 0.834 | 0.690 | 0.829 |
| ROPS | 0.671 | 0.658 | 1.000 | 0.487 | 0.387 | 0.331 | 0.538 | 0.579 |
| ROAS | 0.958 | 1.000 | 0.673 | 0.818 | 0.689 | 0.721 | 0.645 | 0.937 |
| Sortino Ratio | 1.000 | 0.962 | 0.679 | 0.808 | 0.677 | 0.712 | 0.646 | 0.926 |

### Rank Correlation Coefficient for SPI (Financial Firms) Middle Quartiles

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=200** | | | | | | | | |
| SPI | 0.984 | 0.991 | 0.909 | 0.693 | 0.563 | 0.558 | 0.553 | 1.000 |
| TSR | 0.583 | 0.598 | 0.497 | 0.837 | 0.989 | 0.979 | 1.000 | 0.605 |
| Calmar Ratio | 0.539 | 0.556 | 0.470 | 0.789 | 0.990 | 1.000 | 0.980 | 0.565 |
| Sterling Ratio | 0.520 | 0.529 | 0.451 | 0.787 | 1.000 | 0.989 | 0.986 | 0.536 |
| Information Ratio | 0.698 | 0.702 | 0.632 | 1.000 | 0.785 | 0.778 | 0.780 | 0.704 |
| ROPS | 0.909 | 0.909 | 1.000 | 0.654 | 0.532 | 0.527 | 0.544 | 0.908 |
| ROAS | 0.991 | 1.000 | 0.916 | 0.676 | 0.570 | 0.563 | 0.565 | 0.992 |
| Sortino Ratio | 1.000 | 0.989 | 0.909 | 0.678 | 0.578 | 0.571 | 0.578 | 0.984 |

### Rank Correlation Coefficient for SPI (Financial Firms) Last Quartile

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=100** | | | | | | | | |
| SPI | 0.994 | 0.996 | 0.949 | 0.683 | 0.617 | 0.603 | 0.537 | 1.000 |
| TSR | 0.554 | 0.554 | 0.617 | 0.546 | 0.889 | 0.862 | 1.000 | 0.559 |
| Calmar Ratio | 0.624 | 0.619 | 0.670 | 0.637 | 0.906 | 1.000 | 0.859 | 0.626 |
| Sterling Ratio | 0.648 | 0.653 | 0.685 | 0.731 | 1.000 | 0.920 | 0.897 | 0.659 |
| Information Ratio | 0.635 | 0.640 | 0.610 | 1.000 | 0.726 | 0.680 | 0.600 | 0.646 |
| ROPS | 0.952 | 0.949 | 1.000 | 0.655 | 0.664 | 0.665 | 0.637 | 0.949 |
| ROAS | 0.992 | 1.000 | 0.949 | 0.673 | 0.602 | 0.587 | 0.526 | 0.996 |
| Sortino Ratio | 1.000 | 0.992 | 0.952 | 0.668 | 0.612 | 0.605 | 0.538 | 0.994 |

# TABLE A2. RANK CORRELATION OF PERFORMANCE MEASURES (AIM)

### Rank Correlation Small (AIM) Firms

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=254** | | | | | | | | |
| SPI | 0.991 | 0.991 | 0.938 | 0.908 | 0.732 | 0.688 | 0.651 | 1.000 |
| TSR | 0.638 | 0.631 | 0.676 | 0.792 | 0.792 | 0.993 | 1.000 | |
| Calmar Ratio | 0.676 | 0.670 | 0.699 | 0.831 | 0.983 | 1.000 | | |
| Sterling Ratio | 0.716 | 0.714 | 0.725 | 0.873 | 1.000 | | | |
| Information Ratio | 0.892 | 0.889 | 0.856 | 1.000 | | | | |
| ROPS | 0.956 | 0.941 | 1.000 | | | | | |
| ROAS | 0.995 | 1.000 | | | | | | |
| Sortino Ratio | 1.000 | | | | | | | |

### Rank Correlation Small (AIM) Firms Top Quartile

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=64** | | | | | | | | |
| SPI | 0.670 | 0.727 | 0.199 | 0.895 | 0.812 | 0.812 | 0.826 | 1.000 |
| TSR | 0.822 | 0.850 | 0.713 | 0.896 | 0.939 | 0.912 | 1.000 | 0.859 |
| Calmar Ratio | 0.691 | 0.733 | 0.474 | 0.892 | 0.975 | 1.000 | 0.912 | 0.761 |
| Sterling Ratio | 0.704 | 0.751 | 0.501 | 0.903 | 1.000 | 0.975 | 0.939 | 0.786 |
| Information Ratio | 0.528 | 0.591 | 0.236 | 1.000 | 0.865 | 0.858 | 0.852 | 0.837 |
| ROPS | 0.722 | 0.668 | 1.000 | 0.145 | -0.025 | -0.012 | 0.077 | 0.359 |
| ROAS | 0.966 | 1.000 | 0.579 | 0.513 | 0.477 | 0.490 | 0.528 | 0.717 |
| Sortino Ratio | 1.000 | 0.966 | 0.692 | 0.435 | 0.380 | 0.395 | 0.444 | 0.663 |

### Rank Correlation Small (AIM) Firms Middle Quartiles

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=126** | | | | | | | | |
| SPI | 0.999 | 0.999 | 0.992 | 0.968 | 0.925 | 0.921 | 0.916 | 1.000 |
| TSR | 0.325 | 0.300 | 0.419 | 0.448 | 0.898 | 0.956 | 1.000 | 0.317 |
| Calmar Ratio | 0.850 | 0.849 | 0.855 | 0.803 | 0.996 | 1.000 | 0.996 | 0.850 |
| Sterling Ratio | 0.516 | 0.489 | 0.575 | 0.632 | 1.000 | 0.961 | 0.918 | 0.497 |
| Information Ratio | 0.973 | 0.972 | 0.972 | 1.000 | 0.833 | 0.829 | 0.822 | 0.972 |
| ROPS | 0.992 | 0.990 | 1.000 | 0.975 | 0.953 | 0.950 | 0.945 | 0.990 |
| ROAS | 0.999 | 1.000 | 0.993 | 0.976 | 0.933 | 0.929 | 0.925 | 0.999 |
| Sortino Ratio | 1.000 | 0.999 | 0.993 | 0.974 | 0.932 | 0.928 | 0.923 | 0.999 |

### Rank Correlation Small (AIM) Firms Last Quartile

| | SORTINO RATIO | ROAS | ROPS | INFORM. RATIO | STERLING RATIO | CALMAR RATIO | TSR | SPI |
|---|---|---|---|---|---|---|---|---|
| **n=64** | | | | | | | | |
| SPI | 0.928 | 0.933 | 0.114 | 0.756 | 0.291 | 0.024 | -0.022 | 1.000 |
| TSR | 0.024 | 0.048 | 0.245 | 0.057 | 0.763 | 0.981 | 1.000 | -0.035 |
| Calmar Ratio | 0.010 | 0.033 | 0.267 | 0.041 | 0.749 | 1.000 | 0.980 | -0.054 |
| Sterling Ratio | 0.130 | 0.180 | 0.243 | 0.094 | 1.000 | 0.706 | 0.683 | 0.065 |
| Information Ratio | 0.845 | 0.821 | 0.320 | 1.000 | 0.500 | 0.207 | 0.133 | 0.853 |
| ROPS | 0.566 | 0.521 | 1.000 | 0.546 | 0.616 | 0.665 | 0.633 | 0.492 |
| ROAS | 0.881 | 1.000 | 0.135 | 0.683 | 0.412 | 0.135 | 0.095 | 0.911 |
| Sortino Ratio | 1.000 | 0.897 | 0.294 | 0.792 | 0.460 | 0.210 | 0.165 | 0.913 |

## *ENDNOTES*

1. We would like to thank Prof. Francis X. Diebold, Anna Kiukas, Dr. Stephen E. Satchell, and Richard Thornton for comments, and Jay Haverty, Arvid Krönmark and Chen Leo for helping with the analysis.

2. We would like to thank Frank Diebold for pointing this out to us.

3. The precise requirement is that the joint returns are spherically symmetric (see Chamberlain (1983)), but a widely accepted interpretation is the assumption that individual return are normally distributed

4. The CAPM has been the subject of intense debate and scrutiny in the financed community since its inception. Since it is not the central focus of this article, we refer to the extensive and excellent survey in Jagannathan and McGratten (1995) for the key arguments.

5. The Central Limit Theorem states that in the limit, the average or sum of an increasing number of observations is normally distributed. Hence, if returns are recorded hourly, daily returns are an average of 8 observations, weekly 40 observations and monthly about 180 observations.

6. The key argument is that convex feasible sets cannot be shown to always exist, so tangency portfolios may not be unique. For details, the reader should consult Harlow and Rao (1989).

7. Data is monthly returns in the period 01/02/98-01/03/03. Average sample size is 168, smallest sample sizes: 13 (hedge funds) and 16 (emerging market country indices).

8. See Oliver, Wyman & Company's report *State of the Financial Services Industry (2003)* for details; currently the results cover the ten year period from 1993-2003, and the most recent index covers the period of January 1998 to December 2002.

## *REFERENCES*


Artzner, P., Delbaen, F., Eber, J. and Heath, D. (1997), *"Thinking Coherently"*, Risk 10, 68-71.

Balzer, L. (Fall 1994), *"Measuring investment risk: A review"*, The Journal of Investing, 47-58.

Bawa, V. and Lindenberg, E. (1977), *"Capital market equilibrium in a mean-lower partial moment framework"*, Journal of Financial Economics 5, 189-200.

Chamberlain, G. (1983), *"A characterisation of the distributions that imply mean-variance utility functions"*, Journal of Economic Theory 29, 185-201.

Chow, K., Denning, K. (1994), *"On variance and lower partial moment betas and the equivalence of systematic risk measures,* Journal of Business, Finance and Accounting, 231-241.

Damant, D. and Satchell, S. (1996), *"Downside Risk: Modern Theories; Stop – and Start Again"*, Professional Investor, 12-18.

Damant, D. Hwang, S. and Satchell, S. (2000), *"Using a Model of Integrated Risk to Assess U.K. Asset Allocation"*, Applied Mathematical Finance 7(2), 127-152.

Dimson, E. and Jackson, A. (2001), *"High Frequency Performance Monitoring"*, Journal of Portfolio Management 28(1), 33-43.

Dimson, E. and Marsh, P. (1998), *"The Hoare Govett smaller companies index 1955-1997"*, Technical report, ABN Amro Hoare Govett.

Eftekhari, B, Pedersen, C. and Satchell, S, (1998), *“On the Volatility of Measures of Financial Risk: An Investigation Using Returns From European Markets”*, The European Journal of Finance 6, 1- 38.

Eftekhari, B. and Satchell, S. (1996), *“Non-normality of returns in emerging markets”*, Research in International Business and Finance (Supplement 1), 267-277.

Estrada, J. (2000), *''The Cost of Equity in Emerging Markets: A Downside Risk Approach''*, Emerging Markets Quarterly 4(3), 19-30.

Fishburn, P. (1980), *“Foundations of risk Measurement: I. Risk as a probability of loss”*, Management Science 30, 396-406.

Fishburn, P. (1981), *“Foundations of risk Measurement: II. Effects of gains on risk”*, Journal of Mathematical Psychology 25, 226-242.

Fishburn, P. and Kochenberger, G. (1979), *“Two-piece, Von Neumann – Morgernstern Utility Functions”*, Decision Sciences 10, 503-518.

Harlow, W. and Rao, R. (1989), *“Asset-pricing in a generalised mean-lower partial moment framework: Theory and Evidence”*, Journal of Financial and Quantitative Analysis 24(3), 285-311.

Harvey, C. (Fall 1995), *''Predictable Risk and Returns in Emerging Markets''*, Review of Financial Studies, 773-816.

Harvey, C. (2000), *“Drivers of expected returns in international markets”*, Emerging Markets Quarterly 4(3), 32-48.

Henriksson, R. (1984), *''Market timing and mutual fund performance: An empirical investigation'*, Journal of Business 57, 73-96.

Henriksson, R. Merton, R (1981), *''On market timing and investment performance. II statistical procedures for evaluating forecasting skills,* Journal of Business, 54, 513-533.

Hogan, W. and Warren, J. (1972), *"Computation of the Efficient Boundary in the E-S Portfolio Selection Model"*, Journal of Financial and Quantitative Analysis 7, 1881-1896.

Huang, T., Srivastava, V. and Raatz, S. (2001), *"Portfolio Optimisation with Options in the Foreign Exchange Market"*, Derivatives Use, Trading and Regulation 7 (1), 55-72.

Hwang, S. and Pedersen, C. (2002a), *"On Empirical Risk Measurement with Asymmetric Return Distributions"*, Financial Econometrics Research Centre Working Paper WP02-07, City University Business School, London.

Hwang, S. and Pedersen, C. (2002b), *"Do Asymmetric Risk Measures Matter When Modelling Emerging Market Equity Returns?"*, Emerging Markets Working Paper, City University Business School, London.

Jagannathan, R. and McGrattan, E. (1995), *''The CAPM debate''*, Federal Reserve Bank of Minneapolis Quarterly Review 19, 2-17.

Jensen, M. (1972), *"Capital Markets: Theory and Evidence"*, Bell Journal of Economics and Management Science 3, 357-398.

Kahnemann, D. and Tversky, A. (1979), *"Prospect Theory: An analysis of Decision under Risk"*, Econometrica 47, 263-291.

Kestner, L. (1996), *"Getting a handle on true performance*", Futures 25, No. 1.

Kijima, K. and Ohnishi, M. (1993), *"Mean-risk analysis of risk aversion and wealth effects on optimal portfolios with multiple investment opportunities"*, Annals of Operations Research 45, 147-163.

Leland, H. (January / February 1999), *"Beyond mean-variance: Performance measurement in a non-symmetric world"*, Financial Analysts Journal, 27-36.

Litner, J. (1965), *"The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets"*, Review of Economics and Statistics 47, 13-37.

Luce, R. (1980), *"Several possible measures of risk"*, Theory and Decision 12, 217-228.

Lux, H. (2002), *"Risk gets riskier"*, Institutional Investor Magazine, US edition, cover story

Markowitz, H. (1952), *"Portfolio Selection"*, Journal of Finance 7, 77-91.

Mossin, J. (1969), *"Security pricing and investment criteria in competitive markets*", American Economic Review 59, 749-756.

Oliver, Wyman & Company (2003), *"State of the financial services industry / 2003"*

Pedersen, C. (1998), "*Empirical tests for differences in equilibrium risk measures with application to downside risk small an large UK companies*", Cambridge Discussion Papers in Accounting and Finance, No. 41.

Pedersen, C. (1999), "*Separating risk and return in the CAPM: A general utility-based Approach*", European Journal of Operational Research 123 (2000), 28-639.

Pedersen, C. (2001), "*Derivatives and Downside Risk*", Derivatives Use, Trading and Regulation 7(3), 251-268.

Pedersen, C. and Satchell, S. (1998), "*An extended family of financial risk measures*", Geneva Papers on Risk and Insurance: Theory 23, 89-117.

Pedersen, C. and Satchell, S. (2000), "*Small sample analysis of performance measures in the asymmetric response model*", Journal of Financial and Quantitative Analysis 35 (3), 425-450.

Pedersen, C. and Satchell, S. (2002), "*On the Foundation of Performance Measures under Asymmetric Returns*", Quantitative Finance 3, 217-223.

Price, K., Price, B. and Nantell, T. (June 1982), "*Variance and lower partial moment measures of systematic risk: Some analytical and empirical results*", Journal of Finance 37, 843-855.

Rahl, L. (2000), "*Risk-budgeting – A new approach to investing*", Risk Books, London, U.K.

Roll, R. (1977), "*A Critique of the Asset Pricing Theory's Test: Part I*", Journal of Financial Economics 4, 129-176.

Roy, A. (1952), "Safety first and the holding of risky assets", Econometrica 20, 431-449.

Sarin, R. (1987), *“Extensions of Luce's measure of risk”*, Theory and Decision 22, 125-141.

Satchell, S. (1996), *“Lower partial moment capital asset pricing model: A re-examination”*, IFR Birkbeck College Discussion Paper, No. 20.

Sharpe, W. (1964), *“Capital Asset Prices: A theory of capital market equilibrium under conditions of risk”*, Journal of Finance 19, 425-442.

Sharpe, W. (1966), *“Mutual fund performance”*, Journal of Business, 119-138.

Sharpe, W. (2002), *“Budgeting and monitoring pension fund risk”*, Financial Analysts Journal 58(5), 74-86.

Sortino, F. and Price, L. (1994), *“Performance measurement in a downside risk framework”*, The Journal of Investing 3(3), 59-65.

Sortino, F. and Van der Meer, R. (1991), *“Downside risk”*, The Journal of Portfolio Management 17(4), 27-32.

Thaler, R. (1993), *“Advances in Behavioural Finance”*, Russell Sage Foundation.

Treynor, J. (1965), *”How to rate management of investment funds”*, Harvard Business Review, 63-75.

Young, T. (1991), *“Calmar Ratio: A smoother tool”*, Futures 20, No. 12.