Investment performance ranking

of superannuation firms

by

Wilson Sy and Kevin Liu[1]

Australian Prudential Regulation Authority
400 George Street
Sydney NSW 2000
Email: Wilson.Sy@apra.gov.au
Phone: 0612 9210 3507

23 June 2009

(To be presented to the 17th Australian Colloquium of Superannuation Researchers 6-7 July 2009 at the University of New South Wales)

The views and analysis expressed in this paper are those of the authors and do not necessarily reflect those of APRA or other staff. The data samples and methods used in this paper are selected for the specific research purposes of this paper and may differ from those used in other APRA publications.

APRA does not accept any responsibility for the accuracy, completeness or the currency of the material included in this Publication, and will not be liable for any loss or damage arising out of the use of, or reliance on, this Publication.

[1] The authors thank Charles Littrell, Katrina Ellis, James Cummings, Kevin Dent and Irene Shiu for comments on earlier drafts of this paper. Neil Esho and Chris Inman have participated in the design and collection of the investment performance data used in this paper.

Abstract

Investment performance studies of pension or mutual funds have overall been too statistically inconclusive to create definitive rankings to help investors make fund selection decisions. This paper presents an alternative approach based on comparing the pension or mutual fund firms themselves which are highly diversified composite portfolios aggregated over all individual portfolios managed by the firms. Performance data of managed fund firms are more useful for investors because we show here that their statistics are more stable and predictable than those of individual portfolios which are subject to more random influences such as selection bias. This paper overcomes the pitfalls of the Sharpe ratio in ex-post performance ranking by using a new metric for the performance of the composite portfolio of a firm. The metric is risk-adjusted by the volatility of the firm's benchmark, defined by the aggregate asset allocation of the firm. In an empirical study using unique data, we measure the performance of 115 major Australian superannuation firms. The results show that using the new metric the investment performance ranking of the firms is persistent. We identify that higher operational costs are correlated significantly with lower net investment performance. We indicate how the performance ranking of the firms can help investors make fund selection decisions.

1. Introduction

Investment performance of pension funds and mutual funds has been studied formally for several decades, starting with methods of calculating returns (Dietz, 1966), investment performance measurement, attribution and comparison (Treynor, 1965; Sharpe, 1966). Despite a voluminous literature[2] built up over the years, there has been limited progress in translating performance comparisons into guidance for investment selection for ordinary investors, who are advised to heed the disclaimer: "Past performance is not a reliable indicator of future performance".

Most published research applies a variety of performance measures and regression models to publicly available return data of mostly equity mutual funds to detect any statistical regularity such as persistent correlation of returns to various factors. The general motivation appears to be to find out whether there is empirical evidence to show that professional managers possess investment skills. There appears to be no consensus on whether or not statistical evidence shows that the average manager has investment skills. This inconclusiveness on manager skills is taken by some researchers (Malkiel, 2005) as evidence to support the efficient market hypothesis.

So far inconclusive research has not led to a definitive method of ranking funds to help investors select funds. Perhaps the question about the existence of investment manager skills or otherwise is not the best question to ask or to try to answer for investors. A better question for investors would be: how should one go about ranking managed funds for investment selection? A manager might have investment skills, but this is irrelevant if the manager extracts most of the excess returns through visible and invisible fees, commissions or through cost shifting. What matters (Bogle, 2005) is what investors get in net returns for the risks they take.

In researching for new approaches to performance analysis which is useful for investors we have identified past and present deficiencies with data, methodology and their inter-relationships. They may explain why empirical studies of individual portfolio or fund performance are not likely to lead to clear conclusions which are useful for investors. The broad reasons for the inconclusiveness of past research come from factors including data quality, expected small differences in persistent performance and missing information about investment mandates. These factors combined with unavoidable selection bias, discussed in the next section, create material uncertainty about the significance of published statistical results.

In section 3, we suggest a new approach where instead of comparing investment performance of individual funds or portfolios, we compare the investment performance of management firms or their composite portfolios. The main advantages of this new approach are firstly, we can apply a consistent methodology to calculate returns using audited accounting data for all firms and secondly, we avoid selection bias by making a weighted-average assessment of all portfolios managed by the firm. Thirdly, the investment performance of a managed firm is a quantitative measure of the effectiveness of its corporate governance. This firm characteristic is likely to be statistically more stable and persistent than that of individual funds which could be idiosyncratic or volatile, depending on their managers or their markets.

In section 4, we discuss in detail what we want to achieve with investment performance comparisons and propose a new metric that quantitatively differentiates management

[2] For references to earlier literature and for more recent studies, see e.g. Carhart (1997); Kosowski et al. (2006); Haslem et al. (2007); Fama and French (2008); French (2008); Busse et al. (2006), Bauer and Frehen (2008). Opinions are divided over whether empirical data has revealed anything systematic or statistically significant for investors.

firms based on their implicit management mandates. We explain why performance analysis should be a two-step process, rather than the standard single-step ones, because performance measurement as a first step should be separated from performance attribution as the next step. We use simple illustrative examples to justify our new risk adjusted value added (RAVA) metric.

Sections 5 and 6 are given to an empirical application and assessment of our new performance metric. In section 5, we discuss how we have collected unique and high quality data and describe our methodology in calculating gross, pre-tax and net benchmarks, which are needed for computing our performance metric values. Section 6 has three sub-sections. After the first sub-section providing a general statistical overview of the data, in the next sub-section, we apply the new performance metric to the collected data to calculate metric correlations, rank correlations, rank-transitions and to assess performance persistence over time. The last sub-section provides a discussion of performance attribution, where we show that one of the causes of observed persistence is the cost structures of the firms.

In the final section, we review our contributions to the theory and practice of investment performance measurement and discuss their implications for research and the superannuation industry. In order for investors to use our approach, we need to have quarterly asset allocation data of the total fund of regulated superannuation funds. With performance analysis made possible by this new data, we suggest a new and more reliable way for individual investors to make investment selections for their retirement savings.

2. Theory and practice of performance comparison

In order to justify a different approach to performance analysis, we have to identify the problems which make current approaches less than useful for producing the sorts of analysis which are helpful for investors. The issues we discuss include data quality, methodology and industry practices which impact on statistical significance of performance results.

Typically the raw performance data are submitted voluntarily to asset consultants, research houses, magazines or newspapers by the pension or mutual fund firms, largely for the purpose of advertising selected managed funds (Jain and Wu, 2000). There are no laws governing the creation, destruction, naming, classifying or performance recording of managed funds. It is well-known that there are more US mutual funds than there are stocks on the New York Stock Exchange. There are also more Australian retail unit trusts and managed investment schemes than there are stocks on the Australian Stock Exchange.

At any given time, investment management firms are creating new fund “track records” either through “proof-of-principle” computer simulations or through real-time performance with small amounts of seed money. If marketing of some of these funds with promising “track records” are successful, then they are registered with a regulator to accept real client moneys. Only when these funds continue to perform well and grow to respectable sizes of fund under management will they be selected for advertising in performance comparisons. The many failed or dead funds are quietly buried if they are unregistered. If they are registered either they are wound up with money being returned to investors or they are merged with other funds of the same family. Survivorship bias and selection bias already exist (Rekenthaler, 2003) in commercial data due normal business practices, even before they appear in public databases.

But of course there are many other sources of data errors, such as unit pricing errors due to non-synchronous prices in international mutual funds (Goetzmann et al., 2001; Jares and

Lavin, 2004) and stale prices in illiquid assets such as real estate (Redding, 2006) in mutual funds. In some cases, the fund pricing errors have been systematically exploited by executives in well-known (McCallum, 2004) mutual fund market timing scandals, where large cash flows from switching further distorted unit prices. Frauds, such as the recently exposed Madoff fraud, show that many of the mutual funds of funds could have misleading fund prices over long periods of time.

The data submitted to publishers are rates of return where neither the raw accounting data nor methodologies of calculation are audited. Publication of performance tables in the media serves to advertise the presence of consulting or research firms, reduce the cost of searching for investors and improve the flow of funds into the marketed funds (Sirri and Tufano, 1998).

Practitioners who collect the data for client advice work do not place undue importance on them, because they are aware of their limitations. In recommending managers for their pension fund clients, the asset consultants rarely rely on more sophisticated statistical analysis of the performance data. Instead they approach the selected management firms (or more accurately their key persons) to assess the knowledge, philosophy and skills of the managers involved and to understand the quality and stability of their investment processes.

Performance data are collected into publicly available databases such as CRSP (Center for Research in Security Prices). While electronic data submission has increased substantially over the years, older data from CRSP were collected from printed sources in magazines and handbooks (CRSP, 2008). Moreover, the diverse origin of the data creates enormous difficulties just for classifying fund types. In his recent research on equity mutual funds using CRSP, French (2008, p.1565) wrote: “The major challenge is identifying US equity funds”.

Even when classification of funds for investment return comparison is obvious, there is a significant problem, where equity funds (say) being compared could have different levels of growth assets, different investment styles and different asset sizes. In Australia, for diversified funds, a “balanced” fund can have anywhere from 16% to 85% in portfolio weight in high risk assets. This raises a question about the direct comparability of the returns of “balanced” funds in Australia.

Methods of performance analysis which require long data periods (say more than ten years) are disadvantaged by the poorer quality of older data and by the volatility of the financial services industry itself. For example, the famous Fidelity Magellan fund in the US is not a fixed product, but a sequence of different products with several different managers with different styles over the life of the fund. The most recent 20 years of performance history of the Magellan fund have come from five different managers, including the famous Peter Lynch who managed the fund from 1977 to 1990. This type of product mutation due to manager turnover (with usually less than ten-year tenures) is typical of the managed fund industry and its implications should not be ignored in longer horizon empirical studies.

Even when we have accurate data for two identical investment products, the statistical noise from market price volatility required a large quantity of data to be able to detect small but persistent return differences over a long time. In the presence of even the tamest form of white noise from price fluctuations, there is still a significant chance for short-period comparisons that what we are seeing is random noise. For example (Sy, 2009), assume Fund A returns 2% higher than Fund B over the long-term due to greater skill or lower cost and their return volatility is 7% per year, then there is a 29% probability that Fund B will still do better than Fund A over a given three-year period due to chance. This random chance in masking skill in producing a spurious ordering has still nearly 24% probability over a five-year period.

There are rational theoretical arguments (Sy, 2009) based on asymmetric information, fund size and cost, and asymmetric business incentives to suggest that any persistent value added by professional investment managers would be small. These small incremental returns, even if persistent, are easily masked statistically by data noise and market noise from price fluctuations. It could typically take more than 20 years for a 1% or 2% return advantage to become statistically significant, even if very clean data were used in standard performance analysis.

Standard performance analysis typically uses general metrics and methodologies, which treat both performance measurement and performance attribution as a single-step process. We argue that the analysis needs to be a two-step process, where the first step is performance measurement taking into account investment mandates which determine which potential value-adding factors are and which are not under the control of the manager. The ex-post single-period performance analysis in a multi-factor approach is expressed generally an econometric problem:

$$r = F(x, \theta, \varepsilon) \ , \qquad (1)$$

where $F(.)$ is a function to be determined, r is the rate of return of the fund, x is a vector of factors which are outside the control of the fund but can determine the ex-post return, θ is a vector of factors for adding value controlled by the manager and ε is the residual noise. In a typical linear regression model, equation (1) can be written as

$$r = \alpha(\theta) + c(\theta)x + \varepsilon \ , \qquad (2)$$

where $\alpha(\theta)$ is the intercept of the linear regression, if positive, indicates value added and would lead to performance persistence. The scalar product of the vector of regression coefficients $c(\theta)$ and the vector of their corresponding factors x controls for all influences outside the jurisdiction of the fund manager.

Estimation of equation (2) should be all that is required in performance measurement, as a first step in performance analysis. The next step in performance attribution may be to separately analyse $\alpha(\theta)$ in a regression analysis, with factors θ which may explain $\alpha(\theta)$, in performance attribution. In practice, these two steps are merged into one single regression, thus confusing performance measurement with performance attribution.

As we discover all factors θ that make $\alpha(\theta)$ positive and control for those, we could reach a stage where

$$r = c'\theta + c''x + \varepsilon \ , \qquad (15)$$

where ε is a stochastic normal variable with a zero mean and typically $\alpha(\theta) \neq c'\theta$, because $c(\theta)x \neq c''x$. Unless we clearly identify θ as being under the discretion of the investment manager to add value, it would appear the intercept of the regression is zero. For equity mutual funds, it appears that four factors (Carhart, 1997; Fama and French, 2008) are sufficient to achieve an intercept which is approximately zero on average[3]. However, there is no indication that five, six or more factors might not do a still better

[3] Many empirical studies "assume theory to prove theory". They assume validity of the capital asset pricing model (CAPM) or the arbitrage pricing (multi-factor) models such as the Fama-French models in their analysis of empirical data to prove the market is efficient according to those models.

job, in a statistical sense.

Many studies have tested numerous metrics and different econometric models to “explain” investment returns. Instead, we are concerned in this paper with performance measurement which is to quantify the value added by managers given certain freedoms implied by their investment mandates. Hence it is important to know and understand the investment mandate of a given manager, before any investment performance of the manager can be assessed and interpreted as presence or absence of skills. Most empirical studies ignore this presumably because soft data such as mandates are rarely publicly available and therefore cannot be included to enable a two-step analysis.

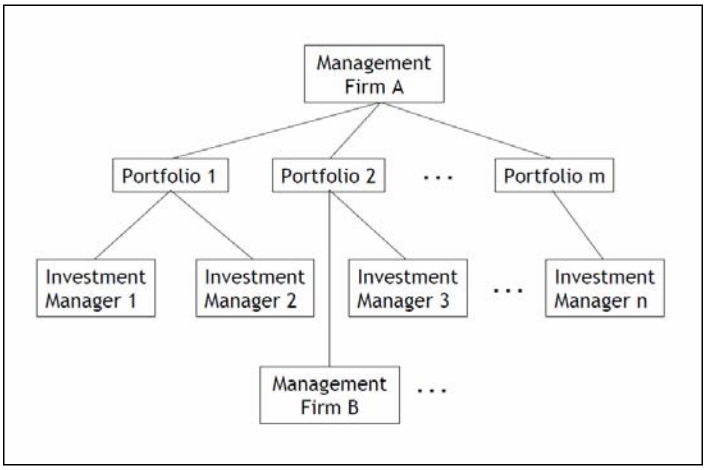
It is vitally important in measuring value added by a manager not to control for factors for which the manager has total discretion. For example, we will discuss below how common performance metrics such as the Sharpe ratio have been inappropriately applied in many cases.

3. Measuring the management firms

For the many reasons discussed in the previous section, measuring the investment performance of individual funds in a statistically meaningful way is very difficult, because the limited power of general statistical methods requires high quality data over long periods of time, making data demands which are difficult to meet due to the way the industry operates. In this paper we suggest and develop a new approach which may solve many of the problems identified.

Typically, the managers of a pension or mutual fund firm are organisers or coordinators of investment portfolio products, which are often known as investment options in Australian superannuation firms, and are known as mutual funds in US mutual fund firms. In Figure 1, the management firm *A* has any number *m* of portfolios or funds where each fund is managed by one or more of any number *n* of investment managers, who may be internal (typical for mutual fund firms) or external (typical for pension fund firms). Some of the portfolios or funds (such as Portfolio 2) of firm *A* may be partly or wholly composed of products from other management firm (such as firm *B*).

Figure 1: Schematic investment structure of a pension or mutual fund management firm



The number of portfolios in a typical management firm is potentially very large. Many would be small, most would be registered and a selected subset of relative large portfolios would be publicly recognised and available to external investors. In the case of a mutual fund firm, some of the portfolios may be identical except for differences in fund size and the fees that are charged (Rekenthaler, 2003). The managers or directors of the firm are responsible through their settings of governance policies and practices and through their portfolio design and selection of investment managers.

An accurate assessment of the investment ability of a firm would require examining the performances of all funds or investment options under management of the firm. Imagine if we had performance data of all funds or investment options under management of the firm, then how would we evaluate its performance? Comparing all investment options of a large firm against products from other firms would be a daunting, if not impossible, task when comparisons are done separately for each of the funds or investment options. A pension or mutual fund firm is really a manager of a composite portfolio, which is defined as the aggregation of the portfolios of all investment options managed by the management firm.

In the Appendix, we have proved that the return of the composite portfolio or the total fund of the firm is equal to the average return of all its funds (portfolios), provided they are asset weighted correctly to take into account external cash flows. We suggest the performance of the composite portfolio or the total fund provides the most reasonable and accurate assessment of the investment performance of a pension or mutual fund firm, because it reflects the overall efficiency of the firm as an independent operating entity.

The most important advantages in using total fund data are that the data are audited and reported to the regulator, including items such as starting total assets, investment earnings, contribution and benefit payment flows, operating expenses, tax expenses and ending total assets. The data for the total fund are not subject to selection bias, and the performance analysis using the data avoids many of the problems related to data quality described in the previous section. For example, we apply the same methodology for calculating the total fund returns for all firms (Coleman et al., 2006). The costs as expenses and taxes are clearly defined and measured, rather than estimated and reported. The calculations of the returns of the total fund are more accurate, because the underlying cash flows tend to have less impact on them than on the returns of its individual constituent investment options (see Appendix).

The main disadvantage in using the performance analysis of total funds of management firms is that it does not translate directly into investment decisions, since it is not normally possible to buy investment units in the total fund of a firm. Rather, selecting the best firm can only the first step, in a sound investment decision process, because as we suggest in this paper the relative qualities of firms are more likely to be stable and persistent over time. But comparing the investment performance of firms is more complicated than simply comparing investment returns of the firms as the following example indicates.

Consider two similar-sized firms A and B each managing two funds (or investment options) 1 and 2 with identical benchmarks. Suppose there are no external cash flows for the managers over the period and their investment rates of returns are given by the following table. In Table 1, Firm A performs better than Firm B in both funds and yet on a total fund comparison, the relative ranking is reversed.

Table 1: An example of individual fund versus total fund comparison

	Fund 1 Return (%)	*Fund 2 Return (%)*	*Fund 1 Asset Weight (%)*	*Fund 2 Asset Weight (%)*	*Firm return (%)*
Firm A	25	5	20	80	9
Firm B	20	0	80	20	16

This simple example shows that there may be no correlation between the rankings of individual funds and the total fund. An investor should always choose Firm A over Firm B, because of better relative performance of all funds. This shows that a simple comparison of returns of firms would be misleading. Of course, this apparent paradox is due to the impact of asset allocation.

In this example, if the firms have control of the asset allocation then the relative performance of the total funds is the more appropriate measure of the firms. But for our interest in pension or mutual fund firms, asset allocation of the total fund is outside the control of the firms. For mutual fund firms or defined contribution pension firms, the firm asset allocations are determined largely by investor choice and for defined benefit pension firms the asset allocations are determined by asset and liability management or other factors outside the control of the firms.

Hence in most situations of interest to us, direct comparisons of the returns of the total fund of the firms is not appropriate, as the asset allocation which is a strong determinant of investment returns (Brinson et al., 1986 and 1991) is outside the control of the firms. While a DC pension firm may set the asset allocations of individual fund or investment option, investors freely choosing those portfolios or sub-funds lead effectively to an uncontrolled asset allocation for the total fund. On the other hand, comparing only individual funds or a subset of funds of firms may also be unsatisfactory due to possible selection bias and resulting in misleading assessments of the performance of the firms.

4. Performance metric for firms

The discussion of the previous section shows that in assessing total fund returns, we need to adjust for asset allocation which is mostly outside the control of the managers of the firm. This leads naturally to the concept of value added relative to a benchmark for the total fund, which is typically a highly diversified portfolio. The value added can be defined both from gross returns or net returns. The net investment return of a pension or mutual fund firm can be decomposed by definition into two components:

$$\textit{Net Return} = \textit{Net Benchmark Return} + \textit{Net Value Added} \tag{1}$$

The net benchmark return is the net return of the benchmark portfolio, which is determined by the aggregate asset allocation of the composite portfolio of the total fund, performing according to benchmark indices for various asset classes, such as the MSCI World indices for international equity and the All Ordinaries Index for domestic equities in the case of an Australian pension firm.

If w_i are the asset allocation weights $(i = 1, 2, ...n)$ of the composite portfolio with n asset classes and r_i are the benchmark index returns net of all costs, which includes all fees and taxes where applicable. An example on the calculation of such net benchmark returns can be found from information provided on mutual funds managed by the Vanguard Group. If R_B is the net benchmark return of the total fund, then

$$R_B = \sum_{i=1}^{n} w_i r_i \ . \tag{2}$$

The net benchmark return R_B is a real, realisable, low cost, passive alternative which a pension or mutual fund firm can potentially adopt to construct all the individual portfolios managed by the firm.

Nearly all firms employ some or all active investment managers in various asset classes in an endeavour to add value by beating their respective benchmarks. The value added in individual asset classes and in aggregate depends generally on three factors:

$$Value\ Added = Tactical\ asset\ allocation + Securities\ selection - Active\ Cost \quad (3)$$

The delegated investment managers are hired to exploit market inefficiencies to add incremental returns through market timing or tactical asset allocation and through selecting or over-weighting better-performing securities, while trying to minimize trading costs. In general, we assume pension and mutual fund firms actively seek to deviate from their benchmarks through taking different levels and types of investment risks.

Clearly performance comparisons based simply on value added or the excess return relative to the benchmark (or simply benchmark excess return), is not adequate, without accounting for the risk or the variability implied by the benchmark asset allocation. For example, if a manager added 1% relative to a conservative cash-like benchmark, then the achievement is much more significant than if the manager added 1% relative to an equity benchmark, which is much more volatile. Hence our performance metric based on value added needs to incorporate risk adjustments.

We propose a generic concept of risk-adjusted value added (RAVA) for our performance metric:

$$Performance\ Metric = Value\ Added\ /\ Risk\ Measure. \quad (4)$$

Such a generic concept of “reward to risk” ratio is well-known in the investment performance literature. We will differ from other performance metrics only in the details of how the risk measure is defined. Examples of such “reward to risk” ratios include the Sharpe ratio (Sharpe, 1966), Treynor ratio (1965), information ratio, as well as many which use more sophisticated measures of risk, such as downside risk or value-at-risk. In our case, as we will discuss below, our choice of risk measure is critical and vital in avoiding some known pitfalls in other measures.

The Sharpe ratio is probably the most widely used and abused investment performance metric and serves as a popular point of reference. It was developed initially for ex-ante ranking of investment opportunities for portfolio decision making. Strictly speaking then, the quantities defining the Sharpe ratio are expectations, which are assumed generally to take values leading to positive Sharpe ratios. However, in ex-post performance measurement applications, actual values of the Sharpe ratio, particularly when they are negative, can lead to contradictions or counter-intuitive results.

An oft-cited (Israelsen, 2005; Scholz and Wilkens, 2008) problem with the Sharpe ratio is when its values are negative. Suppose we are comparing the investment performance of a share and a bond manager. Suppose the managers have largely kept to their respective benchmark risk levels, with the results shown in Table 2.

Table 2: Example on Sharpe ratio

$r_f = 5\%$	*Return (%)*	*Volatility (%)*	*Sharpe Ratio*
Share Manager	2	15	-0.2
Bond Manager	2	5	-0.6

Examples such as this one are considered to invalidate the Sharpe ratio as a performance measure, because despite achieving identical returns of 3% less than cash rate, the share manager having suffered higher volatility, has still a higher performance rating than the bond manager who has lower volatility. However, this apparent counter-intuition is not a paradox, because there is no real difficulty in interpreting this apparent contradiction, even though the Sharpe ratios are negative. Shares being more volatile, a 3% under-performance relative to cash is less statistically significant than the same under-performance for a less volatile asset such as bonds. Hence the rank ordering by the Sharpe ratio makes sense in this case, because the managers have no control over the volatility of their benchmarks.

However, the Sharpe ratio is inappropriate in situations where the manager has control of the volatility or is encouraged to assume different levels of risk from the benchmark. Consider the situation of a down market with negative Sharpe ratios in Table 3.

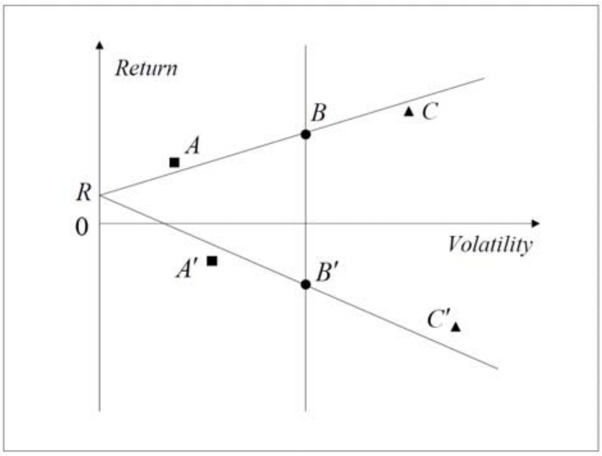
Table 3: Example with negative Sharpe ratios

$r_f = 5\%$	*Return (%)*	*Volatility (%)*	*Sharpe Ratio*
Benchmark	-2	10	-0.7
Manager A	-5	20	-0.5
Manager B	+1	5	-0.8

Manager A did not see the down-market coming and took twice the risk, as measured by ex-post volatility. As a result, the performance of the manager was -5%, less than the return of the benchmark and yet its Sharpe ratio is higher than the benchmark. On the other hand, Manager B did see the down-market coming and took risk management action to reduce its ex-post volatility to half that of the benchmark. As a result, Manager B avoided a negative return, but its Sharpe ratio at -0.8 is worse than that of either benchmark or Manager A. This is counter-intuitive.

A similar example is graphically illustrated in Figure 2, where the risk-free rate is represented by the point R, B' is the benchmark and the slope of the line RB' is the Sharpe ratio of the benchmark in a down market. In can be seen that a manager performance A' with a higher return and lower volatility has a more negative slope RA' and hence a lower Sharpe ratio. On the other hand, a manager performance C' has a lower return and a higher volatility has a less negative slope RC' and a higher Sharpe ratio.

Figure 2: Illustrations of the Sharpe ratio



In a recent paper, Scholz and Wilkens (2008) called this phenomenon a "market climate bias", where "a fund with constant fund-specific characteristics that outperforms the Sharpe ratio of the market index in a declining market will not necessarily have a superior Sharpe ratio in a normal market period." They discovered from empirical data that market climate has a considerable impact on fund rankings.

This paradox originates from the fact that Sharpe ratio merely measures the efficiency of risk-taking: how much excess return per unit risk taken. The Sharpe ratio does not reward or measure the decision of the manager to select the appropriate level of risk. The inappropriateness of the Sharpe ratio when one wants to include value added through risk level selection extends to up markets with positive Sharpe ratios as well. So it is not just a "market climate" phenomenon, where the paradox is simply more noticeable with negative, down-market Sharpe ratios, as the following up-market example in Table 4 illustrates.

Table 4: Example with positive Sharpe ratios

$r_f = 5\%$	*Return (%)*	*Volatility (%)*	*Sharpe Ratio*
Benchmark	8	10	0.3
Manager A	10	25	0.2
Manager B	7	5	0.4

Recognising an up-market, Manager A increased risk and was rewarded with an out-performance of 2% versus the benchmark, but its Sharpe ratio is lower than that of the benchmark. In contrast, Manager B took an overly defensive stance and underperformed the benchmark by 1%, but has the highest Sharpe ratio. Clearly, what the Sharpe ratio is measuring is only the efficiency of risk-taking and not the effectiveness of risk level selection. A corresponding graphical illustration is given in Figure 2 by upwardly sloping line with comparisons of the slopes of the lines RA, RB and RC.

The Sharpe ratio has been widely used among both academics and practitioners and its appropriateness has rarely been questioned, except when pathologies from negative values emerged (Scholz and Wilkens, 2008). Our examples show that the Sharpe ratio as ex-post performance metric may not be universally appropriate. If risk level selection is one of the means of adding value which we expect the pension or mutual fund firm to attempt to do, then it is inappropriate to control for this and eliminate its effects from the performance metric. On the other hand, the Sharpe ratio may be appropriate if the investment mandate discourages ad-hoc risk variations or if it requires the manager to be efficient at risk-taking.

Many measures have been proposed and used to gauge the performance of managed funds, typically without reference to investment mandates of those funds. Using inappropriate metrics by asset consultants to measure and compare the investment performance of managers can lead to distorted incentives (Ingersoll et al., 2007) when the measures can be manipulated. For example, if the Sharpe ratio was used to measure the performance of a mutual fund, then the example in Table 4 shows that there may be a strong business incentive for the fund not to take defensive action even if the manager anticipated correctly a down market. This may explain why few if any managed funds appear to have avoided substantial losses in the recent global financial crisis. This may also be another reason, in addition to the ones discussed by Sy (2009), for "benchmark hugging" or "closet indexing" commonly observed for actively managed funds.

Hence our performance metric should not serve as a disincentive for managed fund firms to adjust the overall level of risk-taking in changing market conditions. Most of the other existing performance measures such as the Jensen's alpha, Modigliani and Modigliani risk-adjusted measure and other variations (Scholz and Wilkens, 2005), have originated from static, equilibrium assumptions (Sharpe, 1964) about investment markets, whereas our performance metric needs to be robust against such assumptions, because our world may be dynamic and in disequilibrium, as the occurrence of the recent global financial crisis has made self-evident. In order to avoid the problems illustrated in Table 3 and Table 4, our risk measure will not be based on the ex-post volatility of the composite portfolio of the managed fund firm, which would be the case if we follow the logic of the Sharpe ratio.

Many of the performance metrics mentioned above have alternative risk measures, which often rely on theoretical assumptions and may be open to manipulation (Ingersoll et al., 2007). To avoid those problems identified, we propose an objective risk measure for the firm be based on the ex-post volatility of the corresponding benchmark portfolio.

If R_A is the total fund return of the firm, R_B is the benchmark return σ_B is the benchmark volatility, then we propose a performance metric ρ defined by

$$\rho = \frac{R_A - R_B}{\sigma_B} \quad . \qquad (5)$$

We have called this metric: a risk-adjusted value added (RAVA) metric, which requires asset allocation data as well as the benchmark return data for various asset classes to calculate the metric. We need to assume the metric is universally monotonic. That is, for two metric values ρ_1 and ρ_2, if $\rho_1 > \rho_2$ implies $U(\rho_1) > U(\rho_2)$ for all utility functions of the users of the metric. This metric has an implicit assumption that there is a linear trade-off between value-added and the risk measure used and that the benchmark return distribution is sufficiently "normal" for its volatility to be an adequate measure of risk. While none of these assumptions can be justified a priori, they form only a very small subset of the standard assumptions routinely made in other metrics and theories. Ultimately, we need empirical studies to see whether or not the metric is useful for discriminating investment performance between firms.

The advantages of the RAVA metric (5) are firstly, avoidance of the contradictions experienced (Scholz and Wilkens, 2008) in ex-post applications of the Sharpe ratio; secondly, easier estimation of the volatility of the benchmark than that of the total fund due to greater data availability; and thirdly, the "manipulation proof" risk measure of our metric. The disadvantage of the RAVA metric is that it does not measure the efficiency of risk-taking relative to return volatility. But in reality, return volatility is only one, among other risks such as operational risk faced by pension or mutual fund firms and therefore an over-emphasis of just one type of risk may not be desirable.

5. Data and methodology

For the remainder of this paper, we apply our new RAVA performance metric to actual empirical data, which APRA Research has collected (without commercial bias) in order to maximize data quality. The objective of this exercise is to see whether the new performance metric is able to provide new and statistically significant information which will help us in selecting managed fund firms.

Using its legislated powers to collect data from regulated superannuation funds, the Australian Prudential Regulation Authority (APRA) which administers the provisions of the *Superannuation Industry (Supervision) Act 1993* and associated regulations, conducted two extensive surveys in 2006: a pension fund governance survey to study trustee policies and practices and an investment performance survey of the largest funds representing 85% of institutional pension assets in Australia. A comprehensive report on the statistical findings on pension fund governance has been published (Sy et al, 2008) and an interpretation has been provided (Sy, 2008).

The investment performance data of the survey consist of yearly financial performance and quarterly asset allocation data for 5 years from financial years June 2001 to June 2006, for both the default investment option and the total fund. A report on the investment performances of the default investment options available to fund members has been published (Ellis et al. 2008). For the reasons stated above, we focus here on the total fund data to show its significance in assessing pension fund trustee performance.

The survey data were collected electronically with data validation before the data were allowed to be submitted. Moreover, the data are financial accounting items which were checked against audited data reported annually to APRA. Of the 187 largest Australian superannuation funds which responded to the survey, only 115 funds, with total assets of $343 billion in 2006, met our strict criteria of quality and completeness. Many submissions were rejected because they did not have all five years of data required, due mostly to industry consolidation associated with the licensing of superannuation trustees by 2006. We then applied a consistent methodology (e.g. equation (A2) in the Appendix) in calculating directly the rates of return of total fund under management from accounting data, which include explicit and audited statements of operational expenses and taxes.

The quarterly data between June 2001 and June 2006 were provided in the ten asset classes shown in Table 5, with the benchmark indices we have selected to construct the benchmark portfolio indices of each pension firm. The passive management fees have been estimated from Vanguard retail products and the asset allocations are asset-weighted averages of all firms over all time periods, shown for indicative purposes.

Table 5: Asset classes, benchmark indices and Passive management fees and asset allocation (%)

Asset Class	*Benchmark Index*	*Manager Fee (%)*	*Allocation Average (%)*
Cash	UBS bank bill index	0.15	8.6
Australian fixed interest	UBS Australian composite bond index	0.29	12.6
International fixed interest (unhedged)	JP Morgan world ex-Australia government bond Index in AUD (converted from USD)	0.31	1.4
International fixed interest (hedged)	JP Morgan world ex-Australia government bond index hedged in AUD	0.31	4.2
Australian shares	S&P/ASX 200 merged accumulation index	0.34	33.5
International shares (unhedged)	MSCI world ex-Australia total net return index in AUD (converted from USD)	0.34	13.8
International shares (hedged)	MSCI world ex-Australia total net return index in local currency	0.36	9.1
Australian listed property	S&P/ASX 200 property merged accumulation index	0.36	3.7
Australian direct property	Australian Mercer unlisted property funds index	0.40	5.3
Other	Australian Mercer unlisted property funds index	0.40	7.8

The asset allocation data for the total fund of the firms were new data items in the sense that APRA has not previously collected the data according to those specific asset classes. Hence the data which are expressed in local currency amounts could only be checked for internal consistency against other quantities, such as total assets. The "Other" asset class potentially includes investments in hedge funds, infrastructure and private equity. In principle, it would be better to use more appropriate benchmark indices such as the HFRX global hedge fund index for hedge funds. There are several reasons why we have not developed a more sophisticated benchmark for "Other".

Alternative investments before 2006 were still relatively small, as over the data period and across all firms the asset allocation to alternatives averaged arithmetically to less than 7% of total assets. Indices for alternative investments have only been developed in the last few years and the available historical data typically do not extend as far back as 2001. Moreover, we did not collect detail data on allocations to each of the alternative investment sub-classes. In Australia, when alternative investments are large for the total fund, they have tended to be infrastructure investments for which the direct property index is probably the best available proxy benchmark.

Gross benchmark returns for the total fund were calculated for each pension firm by multiplying asset allocations averaged over the quarter by gross benchmark index returns. Net benchmark returns, after all fees and taxes, are calculated similarly, except we need to construct net benchmark index returns for each asset class. The construction of net indices was done monthly using accrual accounting methods.

Pro-rata manager fee shown in Table 5 was deducted every month to create after-fee before-tax indices. The assumed manager fees, including administration, custody and manager expense ratios, are obtained or estimated from Vanguard index funds available to Australian retail investors. They are likely to be over-estimates for pension fund firms which should be able to negotiate lower wholesale fees. The gross benchmark returns less manager fees are the pre-tax benchmark returns, which on subtracting tax are the net benchmark returns. Gross, pre-tax and net benchmark returns are calculated for each benchmark index and the returns of the indices are combined together with the asset allocation of a firm to calculate gross, pre-tax and net benchmark returns for each individual firm.

To construct the net benchmark indices we subtracted 15% per annum tax for income and dividends. We estimated that the dividend yield averaged over the five year period was 3.9% for Australian shares, 2% for international shares and 8% for Australian property. For Australian shares, we assume an average franking rate of 80% for dividend imputation and add back the franking credit to offset other tax payments. Finally we calculate capital gains by subtracting income from total returns and apply a 10% capital gains tax. Capital losses are assumed to be used to offset other tax liabilities.

To estimate the benchmark portfolio volatilities of the pension firms, we interpolate the quarterly asset allocation data linearly to produce monthly asset allocation data. From these monthly asset allocations and benchmark index returns, we estimate individual monthly benchmark returns for each firm and calculate the annual return volatilities of the composite portfolios of the firms. We have calculated the actual annual total fund returns (2002-2006) for each firm and their corresponding benchmark returns and volatilities on a gross, pre-tax and net basis. This completes the methodology needed to estimate all quantities which are needed to calculate the performance metric (5) also on a gross, pre-tax and net basis.

6. Analysis of the performance data

The purpose of this section is to analyse the performance data, to see whether the collected empirical data and the computed performance metric data are useful in helping us to understand the investment performance of Australian superannuation firms. Note that the performance of a firm is equal to the asset-weighted performance of all its investment options (see Appendix). The investment performance data for the total funds can be analysed according to various classifications such as sector or operational structure types. But we will present the results of such a study elsewhere; but for now, we focus on a study of the whole dataset to see whether our new metric has adequate power to discriminate in a statistically significant way the performances of individual firms.

6.1 Data overview

Table 6 provides the summary statistics of the five years of annual data for 115 superannuation firms, showing annual firm return, annual benchmark return, benchmark volatility and performance metric, on both gross and net basis. Unless stated otherwise, for most of the tables and figures below the sample size is 575.

The average cost can be defined by the difference between gross and net returns. For the firms the mean and median costs are 1.7% and 2.0% per year respectively. These are likely to be under-estimates because the sample does not include two large retail conglomerates which tend to have expensive operations. The corresponding mean and median costs for the benchmarks are 1.0% and 1.6% per year respectively. The benchmark costs are probably over-estimates because we have used retail expenses to calculate the net passive benchmarks. The average firm in this dataset has under-performed its gross benchmark by 0.2% and has under-performed its net benchmark by 0.9%.

Table 6: Average of annual values 2002-2006

Quantity	Firm Return (%)		Benchmark Return (%)		Benchmark Volatility (%)		Performance Metric	
	Gross	Net	Gross	Net	Gross	Net	Gross	Net
Mean	8.0	6.3	8.2	7.2	5.6	5.0	-0.05	-0.18
Median	12.5	10.5	13.0	11.4	5.4	4.8	-0.04	-0.14
St Dev	8.2	8.0	8.4	7.5	1.6	1.4	0.45	0.53
Minimum	-9.8	-10.7	-8.9	-8.2	1.4	1.2	-2.51	-3.17
Maximum	19.9	18.0	19.5	17.5	10.6	9.5	2.43	2.12

Based on an analysis of the arithmetic mean (see Table 7), the approximate total cost of running the average superannuation firm is 1.2% pre-tax and 1.9% post-tax, of which 1% is due to the cost of the passive benchmark (with 0.3% from fees and 0.7% from taxes).

Table 7: Analysis of the total cost (p.a.) of the average firm

Cost Type	*Calculation*	*Cost (%)*
Passive fee	Gross benchmark return – Pre-tax benchmark return	0.3
Passive taxes	Pre-tax benchmark return – Net benchmark return	0.7
Active cost: visible	Total cost difference between firm and benchmark	0.7
Active cost : invisible	Gross benchmark return – Gross firm return	0.2
Total	Gross benchmark return – Net firm return	1.9

The cost of active investing is 0.9% with 0.7% being due to visible costs and 0.2% being invisible costs incurred while getting the gross return. The cost of active investing for Australian superannuation funds at 0.9% is higher than 0.67% estimated (French, 2008) for US equity mutual funds. Invisible active investing costs come from sources such as investment returns declared net of costs to the firm (e.g. hedge funds or fund of funds), soft dollar arrangements and possibly other transfer payments between conglomerate subsidiaries.

Given the tax concessions on earnings, the average superannuation firm has an effective tax rate of about 9%, which is 60% of the full nominal rate of 15%. The cost of the average superannuation firm is nearly 2% with about half of it due to active management costs, at least one third due to taxes and remaining sixth due to administration cost, which we identify as most of passive fees.

We note that the dispersion in benchmark volatility in Table 6 suggests that adjusting for volatility is important in the performance metric (5). Figure 3 displays the distribution of annual volatility of the net benchmarks, showing the possibility of variations of volatility by multiples of two or more.

Figure 3: Net benchmark volatility distribution 2002-2006

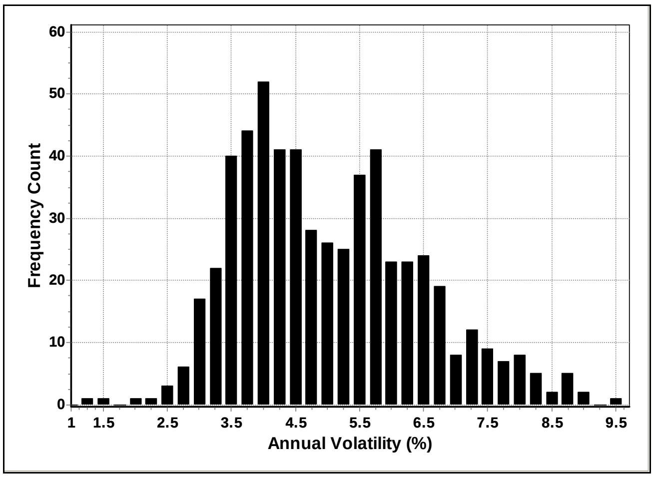
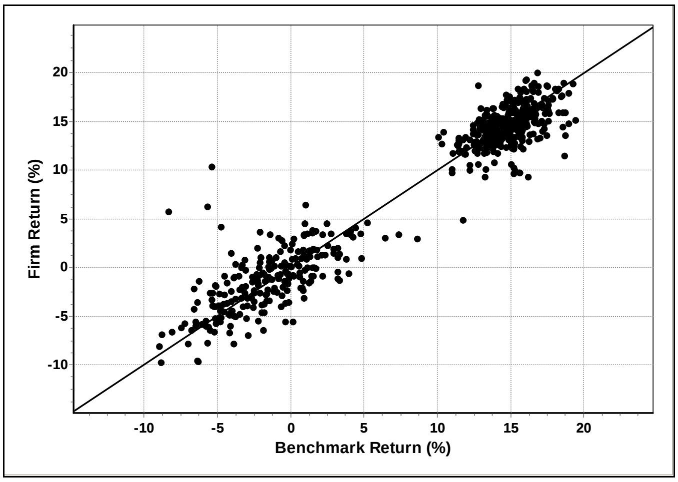


Table 6 also shows that the firms performed about 0.2% lower on average than their gross benchmarks over the five year period. Figure 4 shows the relationship between the gross returns of the firms and their benchmarks. The return distribution is bimodal due the 2001-2003 down-market period followed by a 2003-2004 transition period, then finally a 2004-2006 up-market period. The bi-modal distribution also explains the substantial differences (4.5% to 4.8%) seen in Table 6 between median gross and mean returns, because the medians happen to be located among the up-market returns.

Figure 4: Gross returns of benchmarks and firms 2002-2006



The diagonal line in Figure 4 represents the loci of all firm returns if they were equal exactly to their benchmark returns. Data points above (below) the line represent firms who have performed better (worse) than their benchmarks. Roughly equal numbers of points above and below the line are consistent with the summary statistics of gross returns in Table 6, showing the average firm performance being only slightly lower than their benchmarks. This suggests that averaged over all firms and the five-year period, any gains by the skills of the group of delegated investment managers are more than lost in paying invisible costs so that there were overall small losses relative to the gross benchmarks.

Figure 4 is also a visual empirical verification of the well-known observation (Brinson et al. 1986, 1991) that a substantial part (perhaps more than 90%) of the variability of investment performance is attributable to the asset allocation of the benchmark. It underlines the importance of taking into account asset allocation in comparing investment performance among firms and managers. For example, when two apparently similar "balanced" portfolios are compared, differences in performance may be merely due to, say, a 10% difference in allocations to growth assets and not due to any differences in skills or operational efficiencies of the firms. We will discuss these statements further below.

As indicated above, the average firm under-performed their net benchmark by 0.9% per year. Figure 5 provides some indications of how or when money is lost in Australian superannuation. Evidently, more data points tend to lie below the diagonal line for par performance during down-markets. This raises a question about the value of the active approach to risk management of investment portfolios and may support our doubt about the appropriateness of the Sharpe ratio in measuring performance as discussed in the previous section.

Figure 5: Net returns of benchmarks and firms 2002-2006

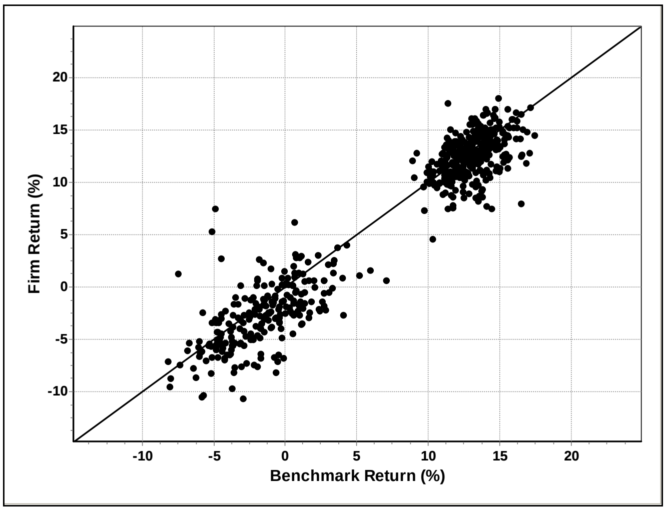
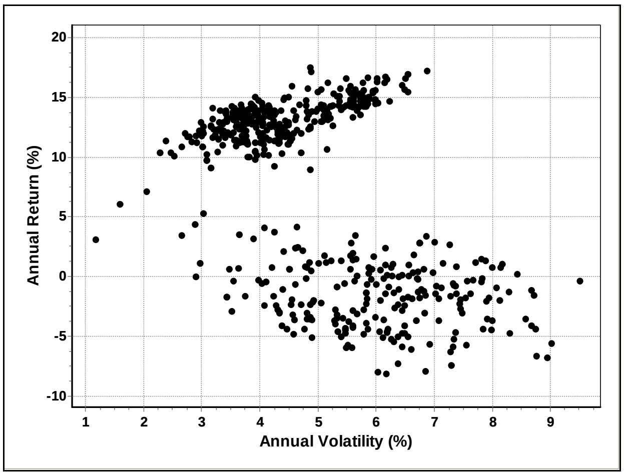


Figure 5 shows evidence of downward parallel shift of data points from the diagonal line and shows a greater downward shift in down markets. The net under-performance of the average firm appears more pronounced in down markets. This suggests either inactive risk management where investment managers appear to forego value adding opportunities in down markets or unsuccessful risk management in down markets due perhaps to costs. Figure 6 shows that market returns are more volatile in down markets. In down-markets, capital losses provide tax credits which mitigate the impact of negative net returns. The empirical data suggest that superannuation firms may be less efficient at using the tax credits from capital gains and losses than we have assumed in constructing the net benchmarks. For example, excessive share trading could forfeit capital gains tax concessions which are available after a 12 month holding period.

Figure 6: Net benchmark return and volatility 2002-2006



The significance of the asset allocation of the benchmark in explaining performance has been discussed numerous times since the early work of Brinson et al. (1986, 1991). That asset allocation explains (Ibbotson and Kaplan, 2000) less of cross-sectional variability (40%) than of time variability (90%) is substantially a statistical observation due to sampling limitations. We provide an example of period dependence of cross-sectional variability from our dataset in Table 8, where we compute the correlation coefficient or the R-squared of the total fund returns and their benchmark returns.

Table 8: Percentage R-squared of dependence of firm returns on their benchmarks

Period	*Gross*	*Net*
2002	43.5	38.5
2003	51.5	48.7
2004	61.3	58.7
2005	45.1	42.6
2006	66.7	55.7
2002-2006	96.3	95.8

It is seen that 38.5% to 66.7% of the cross-sectional annual returns of our dataset are explained by their benchmark asset allocations, depending on the period and depending on whether the comparison is gross or net of costs. Over the whole five-period, the cancellation of short-term noise leads to an R-squared of more than 95%. Our results are consistent with expectations from earlier research (Brinson et al., 1986, 1991; Ibbotson and Kaplan, 2000).

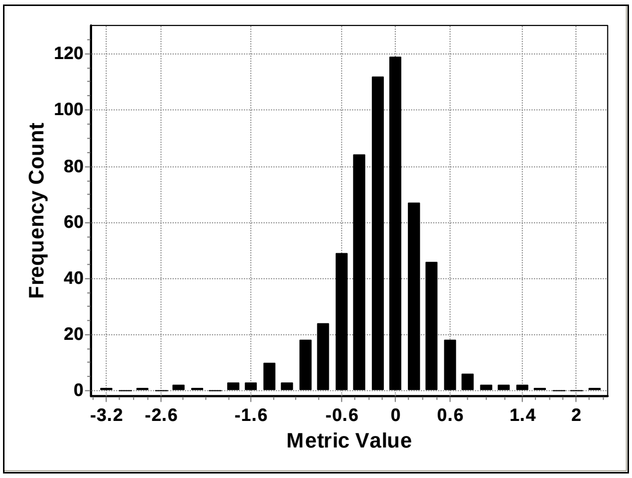
These results have very simple explanations. It is clear that the greater are the differences in returns to different asset classes, the more asset allocation explains performance. Over the short-term, the asset return differences may be insignificant and may be swamped by other short-term effects of active management; asset return differences explain less of the cross-sectional variability. Asset allocation explains more of time variability because over time the differences in returns of different asset classes become more statistically significant.

6.2 Analysis of the performance metric

A performance metric for investment analysis is analogous to a telescope for astronomy, which enables the observer to see more clearly than is possible with the naked-eye. With our RAVA performance metric we want to reliably discern important investment information which is not possible from simply comparing investment returns. As the first step in investment performance analysis, we measure firm performance using the new RAVA metric to demonstrate the potential of the metric to reveal reliable new information.

Reliability requires the performance metric to have certain stable properties. For example, it needs to have a stable range of possible values, even though averages might shift due to changing market conditions. The property is needed to interpret metric comparisons over time as well as for cross-sectional comparisons. Figure 7 shows the frequency distribution of the net performance metric over the five year period. The uni-modal distribution of the metric with an approximate range from -2 to +2 is also similarly stable for each of the five years, regardless of up-markets or down-markets.

Figure 7: Net performance metric distribution 2002-2006



Another desirable property for a performance metric is persistence, because unless there is some positive serial correlation, the performance metric provides little useful information to indicate future performance of the pension fund firms. Figure 8 (next page) provides a visual display of year-to-year correlations, which show highly significant positive correlation, apart from the 2003/2004 correlation which is less significant. This might be due to 2003-2004 being a transition year from down-market to up-market.

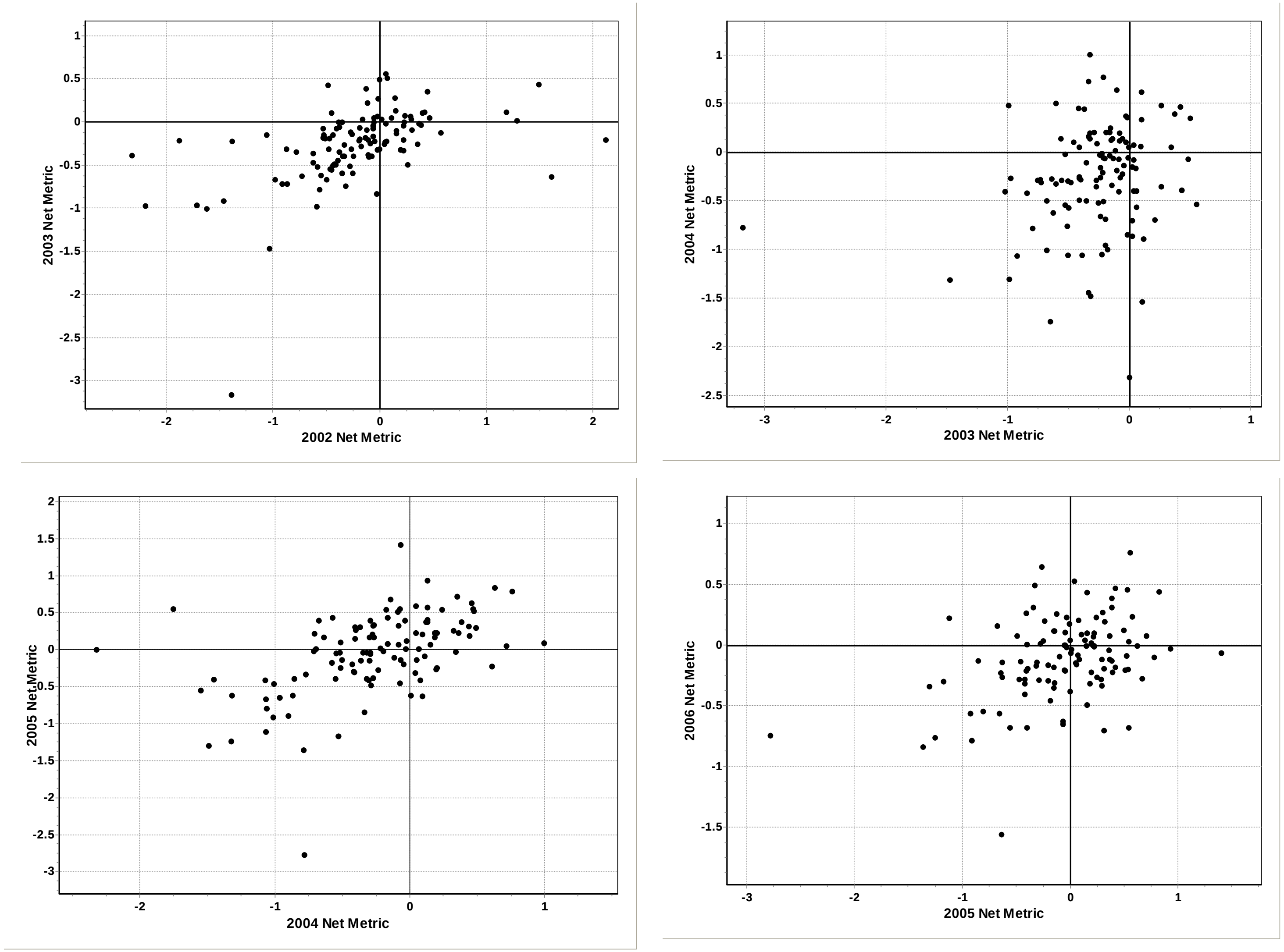
Table 9 shows the correlation coefficients of the net performance metric from one year to subsequent years. There is some evidence of persistence over two-year periods e.g. 2003-2005 and 2004-2006.

Table 9: Net metric correlations from one year to subsequent years

z-values (in brackets) are significant to 5% shown with (*), 1%(**) and 0.1% (***)

Correlation (%)	2003	2004	2005	2006
2002	48.8 (5.64)***	-1.6 (-0.17)	23.9 (2.58)**	23.4 (2.52)*
2003		23.8 (2.56)*	52.1 (6.11)***	24.5 (2.65)**
2004			49.9 (5.80)***	39.2 (4.39)***
2005				41.0 (4.61)***

Figure 8: Net performance metric year-to-year transition 2002-2006



Firm rankings based on the net performance metric are also persistent, but the evidence is more muted because rankings reduce the impact of large or small metric values in the calculation of correlation. Spearman rank correlation results corresponding to those of the metric correlations of are shown in Table 10, where longer-term correlations appear less significant.

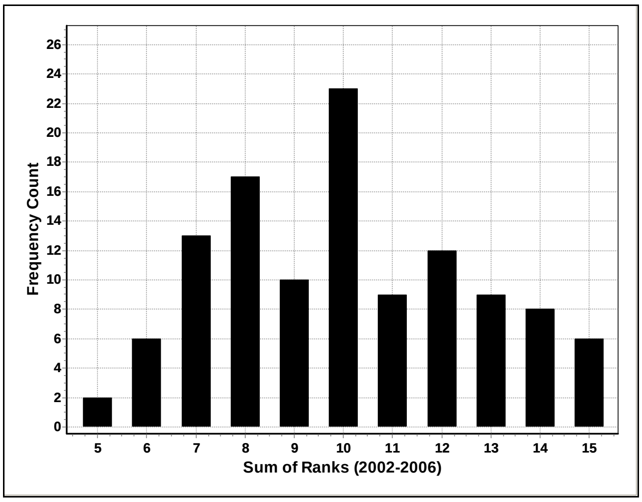
Table 10: Spearman rank correlations from one year to subsequent years
t-values (in brackets) are significant to 5% shown with (*), 1%(**) and 0.1% (***)

Correlation (%)	2003	2004	2005	2006
2002	57.9 (7.55)***	6.2 (0.66)	21.8 (2.38)*	12.9 (1.38)
2003		23.1 (2.53)*	32.5 (3.65)***	12.7 (1.37)
2004			54.4 (6.88)***	34.0 (3.85)***
2005				35.4 (4.02)***

From our studies we found that high quantile rank correlations do not appear to be persistent over the longer-term. For example, firms in the top decile in 2002 are not more likely statistically than other deciles to be in the top decile in 2006. However, at lower quantile rankings, where more intra-quantile migrations are naturally included in the statistics, measured persistence improves. We tested quintile and quartile ranking persistence, but still did not find sufficiently high statistical significance until we reached tertile or 3-quantile ranking.

An indication of tertile rank persistence can be seen from how the ranks are distributed over the period. Figure 9 shows the frequency distribution of the sum of tertile ranks over the five-year period and provides an indication of tertile rank persistence. If the rankings were random, we would expect to see a sharp peak at the value of ten, being the average sum for a random sample. Instead we have a broad distribution, with two cases of top quantile ranking and six cases of bottom quantile ranking, persisting for all five years. From one year to the next, 48.9% of firms have the same rank and a Pearson Chi-square test rejects the random hypothesis to a probability of less than 0.2%.

Figure 9: Sum of tertile ranks by net performance metric 2002-2006



A further analysis of the differences in persistence for different quantile ranks is provided in Table 11, which shows rank transitions of firms in the top and bottom quantiles have 50% and 57% probability of remaining in the same quantile from one year to the next. That is bottom quantile firms have the highest probability of remaining in the bottom quantile, from one year to the next. The firms in mid quantile with only 40% of remaining in the same quantile are only marginally persistent in their performance, indicating a greater tendency to migrate up or down quantiles. The transition matrix is diagonally dominant indicating persistence over the five year period.

Table 11: Transition matrix of tertile rankings 2002-2006

From/To (%)	*Q1*	*Q2*	*Q3*
Q1	50	31	19
Q2	34	40	26
Q3	15	28	57

How likely is a particular quantile ranking persisting over the longer-term? Table 12 gives an indication of longer-term persistence, by examining what happened since 2002 to the 115 firm tertile rankings in subsequent years. Of the top quantile firms in 2002, 61% and 47% were still top quantile firms in 2003 and in 2005, but by 2006 only 32% of those remained, indicating declining longer term persistence. Slightly improved persistent results are observed for bottom quantile firms and more mixed results are seen for mid quantile firms.

Table 12: Tertile rank persistence since 2002

(%)	*2003*	*2004*	*2005*	*2006*
Q1	61	37	47	32
Q2	40	37	47	37
Q3	64	39	49	39

The results show a decline in rank correlation over time. But while persistence is not strong in the longer-term, it is nevertheless sufficiently significant to swing the odds in favour of picking the top quantile firms.

6.3 Causes of persistence

In the first step of performance analysis, we have shown empirically that the RAVA performance metric possesses stable properties such as persistence attributes, which suggest that it may be a reliable tool for performance measurement. In the second step of performance analysis, we undertake performance attribution by exploring some causes of persistence, where we carry out some preliminary analyses to explain the data obtained from the performance metric.

Evidently from equations (3), (4) and (5), which define the RAVA performance metric, there are mainly three sources of manager activity which determine the investment performances of the pension mangers. All these sources are theoretically available to delegated investment managers of a pension fund firm to influence the investment outcome of the whole pension firm.

Firstly, the pension fund firm can change the overall level of risk of the composite portfolio through tactical asset allocation usually implemented via overlays by external specialist managers on individual portfolios.

Secondly, the pension management firm can, directly or indirectly through delegated investment managers, deviate from the benchmark indices of the different asset classes, by over-weighting or under-weighting various risk factors, such as for example having more or less investment in small growth companies than in the equity market indices.

Thirdly, the pension firm can reduce overall operational cost by negotiating lower fees with service providers, such as administrators, custodians and delegated investment managers. Or, the pension firm can avoid paying high active management fee by using passive management for a larger proportion of its assets.

Figure 4 and Table 6 show that taken over the five-year period and averaged over the 115 firms, about 0.2% per year of value was subtracted from active management, on a gross basis before costs. This does not imply necessarily that investment managers have no skills. Table 13 shows some potential evidence to suggest that there may be some firms may have selected investment managers with skills.

Table 13: Gross metric correlations from one year to subsequent years

z-values (in brackets) are significant to 5% shown with (*), 1%(**) and 0.1% (***)

Correlation (%)	2003	2004	2005	2006
2002	42.4 (4.79)***	-11.3 (-1.21)	15.1 (1.61)	-2.7 (-0.29)
2003		24.6 (2.66)**	48.5 (5.61)***	-5.9 (-0.63)
2004			42.0 (4.74)***	24.3 (2.63)**
2005				18.3 (1.96)*

The weaker results show that if they do have skills and consistently out-performed the markets, some of the additional returns may be eroded by invisible costs, such as e.g. transfer payments between financial intermediaries of conglomerate subsidiaries. Despite this, there may still be a few firms which might have consistently added value from the available active sources over the period, where the data points for a particular firm may be above the diagonal line in Figure 2 and have higher gross metric values in Table 13. However, the probability of this happening is substantially lower when costs are taken into account (see Figure 3). From Table 6, the average firm under-performs its net benchmark by 0.9% per year, which is our estimate for the cost of active management for our dataset.

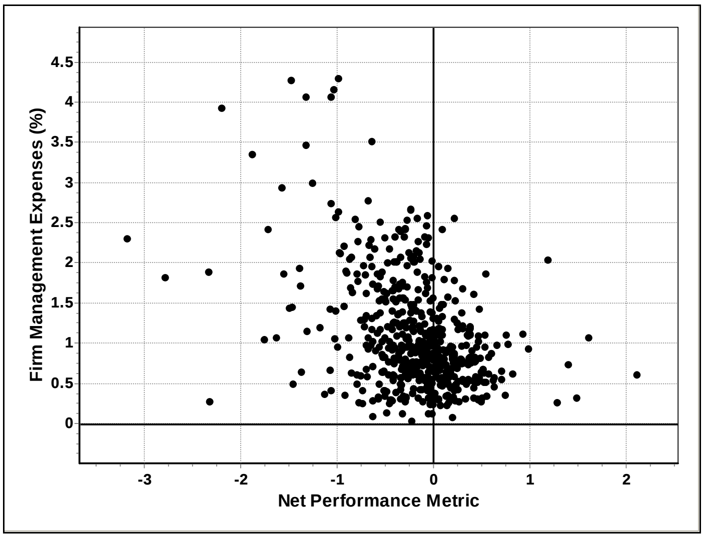
In an analysis of the factors determining potential value added in equation (3), both tactical asset allocation and securities selection are expected typically to be volatile, with investment gains or losses depending on market conditions. However, costs composing of manager fees, operational costs and taxes nearly always subtract from investment returns. Since cost structures of pension fund firms are typically stable, not changing as frequently as market prices, we expect relatively small but stable differences in costs between pension firms may cause persistent performance differences.

We have studied the relationship between total cost and the net performance metric for the dataset. The correlation coefficient is -26.8% between the net performance metric and the cost, which is defined as the difference between gross and net returns and therefore includes all fees, operational costs and taxes. The z-value for the correlation is

-6.57, giving a probability (p-value) of zero correlation or randomness of less than 0.01%. Generally lower costs lead to statistically significantly better risk-adjusted value adding performance.

As tax is under less control than total firm expenses, we also studied the relationship between total firm expenses and the net performance metric for the dataset, as shown in Figure 10.

Figure 10: Relationship between total firm expenses and net performance metric 2002-2006



The correlation coefficient is -41.0% between the net performance metric and the total firm expenses, which is defined as the difference between gross and pre-tax returns and therefore includes all fees and operational costs. The z-value for the correlation is -10.4, giving a probability (p-value) of zero correlation or randomness of less than 0.01%. Evidently lower management expenses of the firms lead to statistically significantly better risk-adjusted value added as measured by the RAVA performance metric. Figure 10 indicates that for those firms which added value or performed above their benchmarks, they have expenses (fees and operational costs) usually less than 1% per year.

Our study shows that low costs and low expenses are clearly important for better investment performance of Australian superannuation firms. Whatever other values firms may add through their policies of active management, on average, they may not overcome costs and expenses as significant factors for net investment performance.

A study of the relationship between the gross performance metric and total firm expenses for the dataset shows a correlation coefficient of -8.5%, z-value of -2.0 and p-value of 0.0213, indicating a small but statistically significant correlation. Since the gross performance metric has no costs or expenses included in its definition, this result suggests firms with higher visible costs may also have a tendency to have higher invisible costs which subtract from their gross investment performance. Invisible costs may be related to corporate governance policy and practice of superannuation firms (Sy et al., 2008).

7. Summary and discussion

In this paper, we have shown by both theoretical arguments and empirical verification that the investment performance of management firms are predictable using high quality data and the new performance metric developed in this paper. In essence, firms have relatively stable operational and governance characteristics which influence their investment performance in a more predictable way than the performance of their individual investment options or funds.

We have analysed the intrinsic difficulties and uncertainties in directly using publicly available performance data on pension fund options and mutual funds to create definitive ranking and help investors make investment selection decisions. In this paper, we propose an alternative indirect approach based on comparing the investment performance of pension and mutual fund firms. In this endeavour, we have discussed why the risk adjustment implicit in the Sharpe ratio is inappropriate where managers are permitted or even encouraged to actively adjust risk levels in a dynamic market environment. We have constructed a new risk adjusted value added (RAVA) performance metric which depends on few theoretical assumptions and is robust against manipulation. The important role in performance analysis played by the benchmark (Brinson et al., 1986, 1991; Ibbotson and Kaplan, 2000) used in our metric has also been independently verified using our unique empirical data.

We have argued in this paper that calculation of the performance metric is the first step which must be separated from the second step in a two-step investment performance analysis. We need to rank the investment performance of firms first based on their mandates before we investigate the properties of the rankings and explain possible causes for them. We have computed metric values for the highly diversified composite portfolios of 115 Australian superannuation firms (2002-2006), using high quality data which we have collected using the legislative power of a prudential regulator. A statistical analysis of the calculated performance metric data shows stable and persistent properties, providing confidence that the metric may provide reliable, new information from performance attribution analysis.

Attribution analysis shows a statistically significant inverse correlation between the net performance metric values and management expenses of the firms. Our estimate on fees and expenses for the average superannuation fund at 1.2% per year is broadly in agreement with other estimates (Rice, 2008) and we note 0.9% is due to the cost of active investing which is higher than 0.67% found recently (French, 2008) for US equity mutual funds. Tax subtracts about 0.7% from the average superannuation fund return and the effective tax rate on superannuation investment earnings was found empirically to be only about 9%.

As management expenses subtract mathematically from net investment performance, we have therefore shown from our dataset that higher management expenses leads to poorer net investment performance of the firms. That is, on average, value adding from active management appears statistically to be unable to overcome higher costs associated with attempts to exploit market inefficiencies, as shown empirically here and elsewhere (Bogle, 2005). Theoretical explanations based on agency conflicts of interest (Drew and Stanford, 2003) and on analytic modelling of asymmetric business incentives, fund size effects, trading costs and asymmetric information have been proposed (Sy, 2009).

Inability of investors to reliably compare individual funds or investment options, as we have argued in this paper, has created an information asymmetry in the market. This together with inadequate fund disclosures (Gallery and Gallery, 2006) has led to a form of market failure, where given the necessary resources many investors may chose to minimize their dependence on the managed fund market. This may explain the recent rapid growth of self-managed funds to become the largest superannuation sector in Australia (APRA, 2008). However, for the bulk of workers who have insufficient assets or other resources, the self-managed alternative is not economically justifiable and they remain captives of the institutional market.

The Australian Government (Sherry, 2008) has taken the initiative to improve market efficiency by requesting APRA to collect and publish disaggregated investment performance data at individual firm or individual investment option level to help superannuation investors to make more informed choices. In this paper, we conclude that to calculate our RAVA performance metric the key additional data which we need from regulated superannuation entities are quarterly asset allocation data of the total fund of the firms.

The usual investment approach for many investors is firstly to select an investment strategy or asset allocation based on the age, income and risk preference of the individual and then secondly to select the investment product from among those offered by competing firms to implement their strategy with the help of financial advisers. We have argued that this approach is unreliable due to information asymmetry discussed in this paper.

Rather, we suggest a new alternative approach where the investors select, as a first step, several top firms as ranked by the RAVA performance metric proposed in this paper. Then as a second step the investors should compare product offerings, taking into account whatever other services which are relevant to the individual such as insurance, financial advice and other client service qualities to select the vehicles to best match their investment strategies and other requirements.

Most pension firms today in various ways can already cater for nearly every conceivable asset allocation requirements of most individuals, whereas their ancillary services may differ. Investment selection is rarely as simple as merely comparing one set of numbers. Our approach should serve at least as a reliable initial filter to reduce the universe of firms under consideration for the investor. It is much easier and cheaper, to avoid paper work and possible entry and exit fees, to switch portfolio or change asset allocation within a chosen firm in our view than it is to have to switch firms when asset allocation changes in the usual approach. Depending on individual circumstances, our approach may provide a more practical alternative.

We suggest that the impact of this new approach to fund selection may improve the efficiency and competitiveness of the Australian superannuation system, with substantially lower cost and increased wealth accumulation for the average worker at retirement. A 1% per year greater cost than is necessary can have seriously deleterious impact over long-term wealth. For example, an Australian worker contributing 9% of wages for 40 years and receives 5% real rate of return could have (Bateman, 2001) final accumulated retirement savings reduced by 22%, if costs were to increase by 1% per year.

Pension and mutual fund firms now manage collectively a significant fraction of global assets and managed funds have become important parts of many economies around the world. There is now a strong global motivation (Antolin, 2008) to assess investment performances and monitor governance practices of pension and mutual fund firms. Our new RAVA performance metric provides a reliable quantitative measure to evaluate the effectiveness of different pension governance policies and practices. We will address this topic in forthcoming studies.

In this paper, costs (agency fees and taxes) are critical for understanding our empirical data of the superannuation industry which is a major part of the Australian economy. Our observations are inconsistent with neoclassical equilibrium economics, including capital asset pricing model and the efficient market hypothesis, where market friction is negligible. Instead, our empirical observations support the cost theory in general (Coarse, 1998; Bogle, 2005) and the agency theory[4] in particular in the understanding of the world of finance and economics.

[4] See for example, Jensen and Meckling (1976), Drew and Stanford (2003), Ambachtsheer (2005), Coleman et al. (2006) and Sy (2009).

8. References

Ambachtsheer, K. (2005), Beyond portfolio theory: the next frontier, *Financial Analysts Journal* (January-February), pp. 29-33.

Antolin P. (2008), Pension fund performance, OECD Working Papers on Insurance and Private Pensions, No. 20, OECD publishing.

APRA (2008), Quarterly superannuation performance: December 2008, *APRA Statistics*, available at:
http://www.apra.gov.au/Statistics/Quarterly-Superannuation-Performance.cfm

Bateman, H. (2001), Disclosure of superannuation fees and charges, Discussion Paper 03/04, Centre for Pensions and Superannuation UNSW, available at:
http://www.business.unsw.edu.au/nps/servlet/portalservice?GI_ID=System.LoggedOutInheritableArea&maxWnd=_CPS_DiscussionPapers_2003

Bogle, J.C. (2005), The relentless rules of humble arithmetic, *Financial Analysts Journal*, Vol. 61 No. 6, pp.22-35.

Bauer R. and Frehen, R. (2008), The performance of US pension funds, ICPM Working Paper, (January 28, 2008), available at SSRN: http://ssrn.com/abstract=965388.

Brinson G. P., Hood L. R. and Beebower, G. L. (1986), Determinants of Portfolio Performance, *Financial Analysts Journal* (July-August), pp. 39-48.

Brinson G. P., Singer B. D. and Beebower, G. L. (1991), Determinants of Portfolio Performance II: An Update", *Financial Analysts Journal* (May-June), pp. 40-48.

Busse, J.A., Goyal A. and S. Wahal (2006), Performance and persistence in institutional investment management, July 2006, EFA 2006 Zurich Meetings Paper, available at SSRN: http://ssrn.com/abstract=890319

Carhart, M. M. (1997), On persistence in mutual fund performance, *The Journal of Finance*, Vol. LII, No. 1, pp. 57-82.

CFA Institute (2005), Global investment performance standards, *Chartered Financial Analyst Institute publications*, February 2005.

Coase R. (1998), The institutional economics, *The American Economic Review*, Vol. 88 (2), pp. 72-74.

Coleman, A., N. Esho and M. Wong (2006), The impact of agency costs on the investment performance of Australian pension funds, *Journal of Pension, Economics and Finance*, 5(3): 299-324.

CRSP (2008), CRSP Survivor-Bias-Free US Mutual Fund Database Guide, Center for Research in Security Prices, Chicago Booth.

Dietz, P.O. (1966), Pension Fund Investment Performance – What Method to Use When, *Financial Analysts Journal* (January-February), pp. 83-86.

Drew, M., & Jon Stanford (2003), Principal and agent problems in superannuation funds, Discussion paper no 142, Queensland University of Technology.

Ellis, K., Tobin, A. and Tracey, B. (2008), Investment performance, asset allocation and expenses of large superannuation funds, *APRA Insight,* 3, pp. 2-20.

Fama, E.F. and French, K.R. (2008), Mutual fund performance, Working Paper, University of Chicago, available at SSRN: http://ssrn.com/abstract=1153715.

French, K.R. (2008), Presidential address: the cost of active investing, *The Journal of Finance*, Vol. LXIII No. 4, pp. 1537-1573.

Gallery, G. and Gallery, N. (2006), Accounting impediments to better superannuation fund governance, *Accounting, Accountability & Performance*, Vol. 12 No. 2, pp. 33-57.

Goetzmann, W.N., Ivkovic, Z., Rouwenhorst, K.G. (2001), Day trading international mutual funds: evidence and policy solutions, *Journal Financial and Quantitative Analysis,* Vol. 36, No. 3, pp. 287-309.

Haslem, J.A., Baker, H.K. and Smith, D.M. (2007), Identification and performance of equity mutual funds with high management fees and expense ratios, *The Journal of Investing*, Summer 2007, pp.32-51.

Ibbotson, R.G. and Kaplan, P.D. (2000), Does asset allocation policy explain 40, 90, or 100 percent of performance?, *Financial Analysts Journal*, (January/February), pp. 26-33.

Ingersoll, J., Spiegel, M., Goetzmann, W. and Welch, I. (2007), Portfolio performance manipulation and manipulation-proof performance measures, *The Review of Financial Studies,* Vol. 20, No. 5, pp.1504-1546.

Israelsen C.L. (2005), A refinement to the Sharpe ratio and information ratio, *Journal of Asset Management*, 5, pp. 423-427.

Jain, P.C. and Wu, J.S. (2000), Truth in mutual fund advertising: evidence on future performance and fund flows, *The Journal of Finance,* Vol. LV, No. 2, pp. 937-958.

Jares, T.E. and Lavin, A.M. (2004), Predictable pricing errors and fair value pricing of US-based international mutual funds, *Journal of Financial Regulation and Compliance,* Vol. 12, No. 2, pp.132-150.

Jensen, M. C. and Meckling, W.H. (1976), Theory of the firm, managerial behavior, agency costs, and ownership structure, *Journal of Financial Economics*, Vol. 3, pp. 305-360.

Kosowski, R., Timmermann, A., Wermers, R. and White, H. (2006), Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *The Journal of Finance,* Vol. LXI, No. 6, pp. 2551-2595.

Malkiel, B. (2005), Reflections on the efficient market hypothesis: 30 years later, *The Financial Review*, Vol. 40, pp. 1-9.

McCallum, J.B. (2004), Mutual fund market timing: a tale of systemic abuse and executive malfeasance, *Journal of Financial Regulation and Compliance,* Vol. 12, No. 2, pp. 170-177.

Redding, L.S. (2006), Persistent mispricing in mutual funds: the case of real estate, *Journal of Real Estate Portfolio Management*, Vol. 12, No. 3, pp. 223-232.

Rekenthaler, J. (2003), When mutual funds die companies bury their mistakes, distorting returns, Money Magazine, April 1, 2003, available at:
http://money.cnn.com/magazines/moneymag/moneymag_archive/2003/04/01/339717/in

dex.htm

Rice, M. (2008), Superannuation fee report 2008: market segment analysis, *IFSA Report*, Rice Warner Actuaries Pty Ltd., 30 June 2008.

Scholz, H. and and Wilkens M. (2005), A Jigsaw puzzle of basic risk-adjusted performance measures, *The Journal of Performance Measurement*, Spring 2005, pp. 57-64.

Scholz H. and Wilkens M. (2008), Interpreting Sharpe ratios – The market climate bias, to appear in *Finance Letters*, also available at:
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=524842

Sharpe, W.F. (1964), Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, T*he Journal of Finance*, Vol. XIX, pp. 425-442.

Sharpe, W.F. (1966), Mutual fund performance, *The Journal of Business*, 39, pp. 119-138.

Sherry, N. (2008), Choosing not to Choose, Speech to launch the Australia Institute/ Industry Super Network Paper, 16 December 2008, Sydney.

Sirri, E.R. and Tufano, P. (1998), Costly search and mutual fund flows, T*he Journal of Finance*, Vol. LIII, No. 5, pp. 1589-1622.

Sy, W. (2008), Pension governance in Australia: an anatomy and an interpretation, *Rotman International Journal of Pension Management*, Fall 2008, pp. 30-37.

Sy, W. (2009), Toward a national default option for low cost superannuation, to appear in *Accounting Research Journal,* also available at:
http://www.apra.gov.au/Research/Working-papers.cfm

Sy, W., Inman, C., Esho, N. and Sane, R. (2008), Superannuation fund governance: trustee policies and practices, *APRA Insight,* 1, pp. 1-17.

Treynor J.L. (1965), How to rate management of investment funds, *Harvard Business Review*, XLIII (January-February), pp. 63-75.

Appendix

We prove here that the investment return of the firm is equal to the average return of all its portfolios (funds), provided they are asset weighted correctly to take into account external cash flows.

According to section 2.A.3 of the Global Investment Performance Standard (GIPS) published by the CFA Institute (2005), "composite returns must be calculated by asset weighting the individual portfolio returns using beginning-of-period values or a method that reflects both beginning-of-period values and external cash flows". External cash flows are important in the calculation because individual funds of pension or mutual fund firms could have potentially substantial cash flows due to contributions, payments or redemptions. We will prove here that the return of the composite portfolio is same as the return of the total fund of the firm provided we use "a method that reflects both the beginning-of-period values and external cash flows."

Suppose for a given period, n funds or investment options of a firm have beginning-of-period assets A_{0i}, earnings E_i and cash flows C_i, for $i = 1, 2, ...n$, then by definition for the total fund:

$$A_0 = \sum_{i=1}^{n} A_{0i} , \qquad E = \sum_{i=1}^{n} E_i , \qquad C = \sum_{i=1}^{n} C_i . \tag{A1}$$

The rates of return of the total fund and individual investment options, money-weighted by external cash flows, are given approximately by

$$R = \frac{E}{A_0 + C/2} , \tag{A2}$$

and

$$R_i = \frac{E_i}{A_{0i} + C_i/2} , \qquad (i = 1, 2, ...n) . \tag{A3}$$

We define the asset weights w_i by

$$w_i = \frac{A_{0i} + C_i/2}{A_0 + C/2} , \qquad (i = 1, 2, ...n) . \tag{A4}$$

We can verify that

$$R = \sum_{i=1}^{n} w_i R_i . \tag{A5}$$

We have proved that the rate of return to the total fund of a firm is equal the asset weighted rate of return of all funds or investment options, provided the correct asset weightings (A4) are used. The method that is consistent with the approximations (A2) and (A3) and reflects beginning-of-period values and external cash flows is given by the asset weights defined in equation (A4).