Working Paper

Toward a national default option for low cost superannuation

Wilson Sy – August 2008
Toward a national default option for low cost superannuation

WILSON SY

Australian Prudential Regulation Authority
400 George Street
Sydney NSW 2000

Email: Wilson.Sy@apra.gov.au
Phone: 0612 9210 3507

18 August 2008

(An earlier version was presented to the 16th Australian Colloquium of Superannuation Researchers 3-4 July 2008 at the University of New South Wales)

The views and analysis expressed in this paper are those of the author 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 I dedicate this paper to Jack Bogle, who has long championed tirelessly the basic truths that inspire this paper. I thank Jack Bogle, Keith Ambachtsheer, Anthony Asher, Natalie Gallery, Charles Littrell, Katrina Ellis, Belinda Tracey and the participants of the 16th Australian Colloquium of Superannuation Researchers for helpful comments.
Abstract

We propose an approach to design a national default option to maximize retirement savings in defined contribution superannuation. We prove that the proposed proportionate shareholding approach minimizes total cost of investing and is a collective optimal strategy. The strategy is more flexible than standard indexing, and is fully-scalable and available to all investors. In an inefficient market, with information asymmetry and professional manager conflicts of interest, we argue that the strategy is also an optimal strategy for individuals. The national default option, which we suggest to be offered by authorised professional fund managers, will not only create a simple and understandable alternative for most workers but will also provide a performance benchmark to encourage the development of a more competitive and efficient superannuation market.
1. Introduction

Many countries around the world face similar problems of funding the retirement of aging populations, now increasingly through defined contribution schemes. Defined contribution plans are pension schemes which are essentially self-funded through personal savings encouraged by tax incentives. It is important that the national savings are maximized through efficient pension investments.

The recent trend towards individual investment choice in competitive markets for pension products has given rise to a complex web of financial intermediaries, including trustees, consultants, fund managers, stockbrokers, financial planners and other service providers. The unbundling of investment services traditionally provided by the pension trustees and now often separately charged has increased the overall cost of the pension system (Coleman et al. 2006; Sy, 2008b) so that the growth of national pension wealth is now sub-optimal.

It is arguable whether the increased flexibility and complexity of market offerings justify the increased cost to most members of the pension system. For the bulk of workers typically before the age of 55 years, their financial objectives are simply to accumulated sufficient wealth so that they may be in a financial position to consider more sophisticated alternatives through financial advisors. It is imperative that some simple, low-cost investment choices be understood by, and made available to, all workers.

The observed inertia (Cameron and Gibbs, 2005) or lack of interest in making investment choices is economically rational, given the low account balances (less than $100,000) of most workers. In the unlikely event that a worker can make a better investment choice and beats the averages, the benefits of better gross returns would be swamped by the additional costs involved in making a better decision, due to the lack of scale, leading ultimately to lower net returns. Brown et al. (2002) have suggested the establishment of a government-regulated universal default fund to cater for members who are unable or unwilling to make informed choices. Similar expressions of the need and suggestions for national approaches for most workers have been made in other countries such as in the U.K. (Turner et al., 2006) and in Canada (Ambachtsheer, 2008). We will investigate a particular non-coercive national default option in this paper.

Member investment choice, as distinct from “choice of funds” provided by superannuation choice legislation in Australia, has led to a higher cost structure for the pension industry without necessarily maximizing savings for most workers. The recent regulatory effort by ASIC (2007) to improve product disclosure has served to show that the pension industry has evolved a complex cost structure (Chant, 2008) which is difficult for most investors to understand and quantify.

Regardless of whether the trustee is a profit-making organisation or not, most of the costs of the superannuation fund come from service providers which are profit-making entities. A profit-driven market often fails to deliver simple, low-cost products, as markets generally abhor simplicity. Simple products compete only on price, leading to a process of commoditization which eliminates abnormal profits.
Hence the expected benefit of market competition driving down the cost of superannuation may not have been realised or is being realised only very slowly. Competition is less effective when there are complexities and there are few simple low-cost products which can serve as benchmarks for comparison.

Faced with alternatives of “more choices” or “fewer choices”, almost everyone would want “more choices”, without knowing the potential problems of too many choices (Benartzi and Thaler, 2002; Brown et al., 2002). But when faced with alternatives of “many choices at high costs” or “few choices at low costs”, almost everyone should prefer the latter if the goal is to maximize savings for retirement, as we will show below. Everyone should given the choice to decide whether they want “no choice”, “few choices” or “many choices”, whereas many workers are forced currently to face many choices. We believe (Thaler and Sunstein, 2003) governments have the responsibility to help unsophisticated individuals and guide them to save in simple and effective ways so that aggregate national savings are optimised.

The purpose of this paper is to propose an approach to create a national default option for low-cost superannuation. We are motivated to help workers who are mostly in the wealth accumulation phase and have little or no other assets apart from their superannuation assets and perhaps the houses they live in. This would be the case for the bulk of the working population under the age of 55. We show that when we take a balanced view of all theoretical arguments and empirical evidence accumulated over the past thirty years in finance and economics, there is an optimal investment strategy that should be used to create default superannuation options for most workers. We propose such a non-coercive choice to capture the best for most workers based on what we regard as the soundest parts of current knowledge and experience of the investment industry.

In the next section, we describe a truly passive approach to investment called the proportionate shareholding approach which is simpler and more flexible than the standard capitalisation weighted index approach. We also explain why our approach should not be called “indexing”.

In section 3, we prove that the collective optimal investment strategy is to use the proportionate shareholding approach to minimize the cost of investing in the nation’s corporations for all investors. In doing this we have converted the “Cost Matters Hypothesis” (Bogle, 2005a) to a “Cost Matters Theorem”. Also a similar statement has been put forward before by Sharpe (1991), using a different verbal argument based on comparing active and passive investments relative a market benchmark.

There are several reasons why we prove formally what may appear to be obvious. Verbal arguments (Sharpe, 1991) lack the clarity and certainty of formal proofs, which are particularly important in clarifying the assumptions made. Important theorems in mathematics are often proved in several ways, because different methods of proof provide different insights. We show from our proof that it is almost free of assumptions according to “relentless rules of humble arithmetic” (Bogle, 2005a) and that important insights about the nature of passive investing will be applied in new ways.
Our insights are particularly important in resolving many conflicting assertions and opinions which currently exist in the investment markets, particularly between academics and practitioners. In section 4, we introduce four different definitions of “efficiency” and argue that securities markets are generally inefficient on every definition. The market inefficiencies can be and have been exploited to “beat the market” by selected individuals. But we suggest that if most individuals cannot easily exploit those market inefficiencies then the proportionate shareholding strategy will also be an individual optimal strategy.

In section 5, we show why finding that rare talented manager who can “beat the market” for us is only a faint hope, because empirical studies and performance comparison tables routinely published in journals and the media suffer from a variety of sampling errors. Even if we succeed in finding a talented manager, in investing there is “nothing that fails like success” (Bogle, 1999), due to what we would call a law of conservation of capital. The phenomenon of boutique managers is seen to be related to managers trying to escape the performance limitations of asset size. We also show from a quantitative model of the business asymmetric incentive faced by an active investment manager that the prospects of beating the market is not good, because under many circumstances it is rational for the active manager to adopt a passive strategy, thus providing an explanation for the phenomenon of “closet indexing” commonly observed. We conclude that it is difficult for most individuals to find managers who could out-perform a passive approach in the long term.

In section 6, we discuss the rationale for government intervention and the creation of a national default option for most workers to have low-cost superannuation. We discuss how the national default option might be implemented in the context of existing practices in the superannuation industry. We address the “free rider” arguments often raised against passive investing in relation to liquidity and price-efficiency. Instead of trading we indicate that real wealth creation or improvements in social welfare can be achieved through better corporate governance and that the proportionate shareholding approach can be used to facilitate the creation of a registry for efficiently transferring voting rights to a new class of proxy capitalists, in a new form of representative capitalism.

In the final section, we summarise the main propositions proved and arguments raised in this paper leading to the proposal of a concept of national default option for low-cost superannuation.

2. Proportionate Shareholding

There is no precise universal definition for the market, which could include any traded securities. To make our discussion simpler and more concrete, we restrict ourselves initially to stocks. We define a stock market as consisting of a couple of objects: a set of listed company stocks and the set of all investors and traders who deal in those stocks. So depending on different sets of selected company stocks, there are different stock markets. After we have defined a particular stock market, we might subsequently refer to it as the stock market for convenience. The portfolio which has all stocks of the market is defined as the market portfolio. The market portfolio is owned collectively by all investors.
Consider a stock market with \( N \) listed companies, the market capitalisation \( C_i \) at time \( t \) is defined as

\[
C_i = \sum_{i=1}^{N} n_i p_u
\]

(1)

where \( n_i \) and \( p_u \) are respectively the number of outstanding shares in \( i \)-th stock and its stock price at time \( t \).

Ignoring dividends for the moment for simplicity, the market return \( R_i \) for one period starting at time \( t \) is defined by

\[
R_i = (C_{t+1} - C_t) / C_t.
\]

(2)

Consider an investor who holds a constant fraction \( \pi \) of the outstanding shares of every stock, then the number of the \( i \)-th stock held is \( \pi n_i \) and the portfolio value \( v_i \) at time \( t \) is given by

\[
v_i = \sum_{i=1}^{N} \pi n_i p_u = \pi C_i.
\]

(3)

Clearly by holding a constant fraction \( \pi \) of the outstanding shares of every stock, the investor holds the same fraction \( \pi = v_i / C_i \) of total market capital. The return of such a portfolio \( \rho_i \) for the period starting at time \( t \) is equal to the market return \( R_i \), since from equations (2) and (3):

\[
\rho_i = \frac{\pi C_{t+1} - \pi C_t}{\pi C_t} = R_i.
\]

(4)

This is the proportionate shareholding approach to obtaining exactly the market return, which can obviously include as well dividend income over the period. It is important to note that the strategy consists simply of holding the same proportion \( \pi \) of the outstanding shares of every stock in the market.

This strategy is independent of individual company prices or capitalisations. To mention prices, capitalisation and portfolio weights is to make a simple idea, unnecessarily complicated in the first instance. We only need to know the total number of outstanding shares for each stock to easily construct such a portfolio. Consider a two stock portfolio consisting of BRKA and ABC: BRKA has 1.55 million shares with a capitalisation of $281 billion and ABC has 161 million shares with a capitalisation of $6.2 billion on a given day. Under the proportionate shareholding approach, a portfolio with a 10% constant proportion of the market would hold 155 thousand BRKA shares and 16.1 million ABC shares. The initial cost of acquiring such a portfolio obviously depends on individual stock prices, which may vary over the period of construction. In the case of quick construction, we pay approximately $28.1 billion for BRKA and $620 million for ABC. The same fractional ownership of companies implies the same fractions of shares and capital of the companies.
There is no subsequent maintenance cost of the portfolio, which has no need for rebalancing due to price fluctuations. Portfolio adjustments are needed rarely only when the numbers of outstanding shares change due to corporate actions such as new issuances or buy-backs. Obviously, for a given set of stocks, our approach is equivalent to a capitalisation weighted approach, for which we provide a formal proof in the next section. However, the emphasis on capital weighting is an unnecessary distraction from the basic ownership concept, because weighting is a portfolio construction concept due to starting from a fixed number of stocks. We will expose a popular misconception that passive management is synonymous with indexing. We will argue that not all indexing is passive and that our passive approach is not indexing. We deliberately disassociate our concept from the common “index” concept for several important reasons.

It is fundamental to our approach that it is truly passive in every sense of the word. The adjective “active” as opposed to “passive” has two important meanings in portfolio management. The more common meaning refers to actively trading stocks leading to turnover, as in market timing or portfolio rebalancing. The other meaning is having a different asset allocation from a benchmark portfolio, as in active stock selection. The proportionate shareholding approach is guaranteed to be passive in both senses of the word.

Indexing as the term is now used refers to many different, and sometimes conflicting, investment approaches, with the only common denominator being perhaps a largely or completely rule-based approach to investing. Some indexing approaches based on multi-factor models, such as the recent Research Affiliates Fundamental Index (Arnott et al. 2005) are active in terms of having turnover due to regular trading required to rebalance the portfolio. Even some capitalisation weighted index funds can be quite active, if portfolio simulation including only a small number of stocks is used to replicate the performance of the index.

An index generally refers to a fixed definition for portfolio construction, such as the Dow-Jones Industrial Average (DJIA) created in 1896 or Standard and Poors (S&P) indices: S&P90 created in 1923 and S&P500 created in 1957. These indices are useful for a variety of purposes including investment comparison and financial research and we expect they will continue to perform their useful roles. However, they have inherent limitations for purely investment purposes due largely to pre-determine universes of stocks, as we will discuss below.

The selection of stocks to include in an index is based on criteria adopted by the indexer, which might not coincide with those of the investors. For example, the number of stocks (too few or too many), the types of stocks (eg for ethical investment) or stock weight adjustments (for liquidity reasons) might not suit some investors. Hence if we adopt a particular index we are forced to invest in ways which are dictated by some outside agency, who does not take into account our particular investment objectives.

In this paper, we demonstrate an understanding of why certain index approaches have been so successful and apply their underlying principles rather than follow known indices slavishly to create an investment strategy which is more flexible than standard index approaches. For example, we could create a fully passive and highly diversified portfolio which excludes certain stocks which may be regarded as
unethical or socially irresponsible or which are redundant to our approach such as exchange traded funds (ETF) or some listed investment companies (LIC). There is also no need to set arbitrary numerical upper or lower limits to the number of stocks the portfolio can hold, causing unnecessary trading when stocks enter or exit the portfolio with a fixed number of included stocks. The portfolio structure can vary with the number and selection of stocks over time, depending e.g. on cash flows or corporate action, but at all times maintaining the principle of passivity.

To make the difference between our approach and standard cap-weighted indexing clearer, let us consider a simple example. Suppose a large-cap stock with about 3% weight of the S&P500 went bankrupt, having its share price written down to near zero. This would induce the index to sell down fractions of holdings of 499 stocks and buy a new stock, incurring transaction costs. For a $100 billion fund with 0.1% brokerage, the cost would be about $3 million. In our approach, it does not matter whether we happen to have 385 or 733 or any (large) number of stocks in our portfolio, we need to do nothing, thus saving $3 million in transaction costs.

For now, we note the proportionate shareholding approach has a fully neutral preference to all stocks in the portfolio, in the sense that it has the same relative stake in every stock, without favouritism. The portfolio is totally passive and fully scalable from a relatively small asset size all the way to the whole market (scalable in the full range: $0 \leq \pi \leq 1$) and it is available to all investors, at least through a collective arrangement if individual asset sizes are small. Note that unless the relative stakes are the same, with a constant proportion for every stock, the approach will not be fully scalable (e.g. DJIA, which is not cap-weighted) and the return will not equal the market return as defined by equation (2). The proportionate share portfolio has an identical structure to the market portfolio. Even though we have emphasised full scalability, we do not anticipate every investor to adopt the approach, as we will explain below.

In summary, the proportionate shareholding approach is a passive investment method which emphasises collective ownership of companies. It is not a portfolio investment product like an index, which is a well-defined portfolio for investment performance replication. The proportionate shareholding approach is flexible in terms of the number or selection of securities so that its portfolio structure cannot be easily anticipated or replicated. Some residual uncertainty is important in practice as it prevents gaming by the rest of the market which could otherwise take advantage of the trading activities of what could be an enormous portfolio. The concept of index tracking error is also inappropriate for our proportionate shareholding approach, as we are unconcerned by short-term differences in performance against any particular benchmark index.

### 3. Collective Optimal Strategy

We will prove that the proportionate shareholding approach is the optimal investment strategy for all investor collectively, in a real world with transaction costs and taxes. We will argue that it is also an optimal investment strategy for individual investors in a game-theoretic sense in an imperfect world in the following section.
We show firstly that the market return defined in equation (2) is the capitalisation weighted average return of individual stocks. The capital weight \( w_i \) of \( i \)-th stock in the market portfolio at time \( t \) is defined by

\[
    w_i = \frac{n_i p_i}{C_t}
\]

(5)

and its return \( r_i \) for one period starting at time \( t \) is defined by

\[
    r_i = \frac{(p_{i(t+1)} - p_i)}{p_i} \cdot
\]

(6)

Combining (1) and (2) then using definitions (5) and (6), we find

\[
    R_t = \sum_{i=1}^{N} n_i \frac{p_i}{C_i} \left( \frac{p_{i(t+1)}}{p_i} - 1 \right) = \sum_{i=1}^{N} w_i r_i.
\]

(7)

The market return is by definition the return of the market portfolio; it is also equal to the capital weighted average return of individual stocks. It is important to note that the market return defined here has nothing to do with any conception of market equilibrium or the market return of the capital asset pricing model (CAPM). The CAPM market portfolio (Sharpe, 1964; Lintner, 1965) is an abstractly defined concept which has never been precisely specified in reality (in terms securities, investment horizon and statistical estimates etc) and therefore no statement about this portfolio can be directly tested through empirical observations.

The market return of CAPM requires a portfolio optimisation of its mean return versus its variance, an optimisation concept which we have not used or needed to use here. There is a widespread, but mistaken assumption that the capital weighted portfolio is a good proxy for the market portfolio of CAPM. It has been proven (Hsu, 2004) that the capital weighted portfolio is mean-variance sub-optimal and hence cannot be the CAPM market portfolio, which by definition must be mean-variance optimal. Capital weighted indexing cannot be advocated on the basis of mean-variance optimality.

Next we show that the market return given by equation (7) equals the asset weighted average return of all investors, regardless of their investment strategies. An equivalent statement was made by Sharpe (1991) as “the return on the actively managed dollar will equal the return on the average passively managed dollar”. However, this statement on its own can be misinterpreted and is untrue generally as it depends on the benchmark against which active or passive is defined. For example, the statement would be untrue, if the benchmark is the Dow Jones Industrial Average index. The statement would be nearly true if the benchmark is the S&P 500 capitalisation weighted index, which was what Sharpe used in his statement. Hence it is worthwhile to provide a rigorous formal proof of the general statement with few assumptions and without referring explicitly to any benchmark or investment style.

We provide two formal proofs below: the longer proof which is perhaps more intuitive, requires the notion of stock returns, whereas the shorter one does not. Consider the stock market having a total of \( M \) investors. Let the \( j \)-th investor own
number of the $i$-th stock at time $t$. The value of the investor’s portfolio $v_{jt}$ at time $t$ is defined by

$$v_{jt} = \sum_{i=1}^{N} u_{it} p_{it} \tag{8}$$

which represents a fraction $\phi_{jt}$ of the total market given by

$$\phi_{jt} = \frac{v_{jt}}{C_t} \cdot \tag{9}$$

Since the sum of all investor holdings of a particular stock must equal to the number of outstanding shares in that stock, we have

$$n_i = \sum_{j=1}^{M} u_{it} \tag{10}$$

Equations (8) and (10) together imply the obvious mathematical identity that the sum of all investor capital $C_t$ equals the total market capital

$$C_i = \frac{\overline{C_t}}{C_t} = \sum_{j=1}^{M} v_{jt} \tag{11}$$

where the right hand equality is a definition of total investor capital.

The $j$-th investor’s return $\rho_{jt}$ for the period starting at time $t$ can be shown to be given by

$$\rho_{jt} = \frac{(v_{j(t+1)} - v_{jt})}{v_{jt}} = \sum_{i=1}^{N} (u_{it} p_{it} / v_{jt}) r_t \tag{12}$$

The coefficient $(u_{it} p_{it} / v_{jt})$ ahead of the stock return $r_t$ is the capital weight of the stock relative to the investor’s total portfolio value. For now, we can think of prices changing due to trading at the end of the period, which normally requires the existence at least some active investors. In general, for an active investor, these capital weights will not be equal to the corresponding capital weights of the market portfolio. The individual investor returns will be different from the market return given by equation (7).

However, the asset weighted average return $\overline{R}_t$ of all investors for the period starting at time $t$ is given by

$$\overline{R}_t = \sum_{j=1}^{M} \phi_{jt} \rho_{jt} \tag{13}$$

where the quantities under the summation are defined by (9) and (12).
Substituting the definitions in (13) and interchanging summations, we find

\[
\overline{R}_t = \sum_{i=1}^{N} \sum_{j=1}^{M} \left( u_{ij} p_{it}/C_i \right) r_{it} = \sum_{i=1}^{N} \left( n_i p_{it}/C_i \right) r_{it} = \sum_{i=1}^{N} w_i r_{it}
\]  

(14)

where we have used the identity (10).

Comparing (7) and (14), we have proved the asset weighted average return \( \overline{R}_t \) of all investors (without needing the concept of active or passive investing) is exactly equal to the market return \( R_t \) of the market portfolio: \( \overline{R}_t = R_t \), assuming no trading between investors over the period. Again this result is a mathematical identity and is valid independent of assumptions about stock price volatility, investor risk preferences and market efficiency. It is important to note that the result is valid not only for the whole market (containing all stocks), but it is valid for any subset of stocks which we might like to define as “the market”.

For a shorter proof, we note that in a frictionless market, without costs, market capital is conserved despite changes in shareholdings among investors in the sense that from equation (11),

\[
C_{t+1} - C_t = \sum_{j=1}^{M} \left( v_{j(t+1)} - v_{jt} \right),
\]  

(15)

which provides another way of proving the equality: \( R_t = \overline{R}_t \). To see this, we divide both sides of (15) by \( C_t \) to obtain from (9) and (12),

\[
\frac{C_{t+1} - C_t}{C_t} = \sum_{j=1}^{M} \frac{v_{j(t+1)} - v_{jt}}{C_t} = \sum_{j=1}^{M} \phi_{jt} P_{jt}.
\]  

(16)

This equation leads directly to \( R_t = \overline{R}_t \) on applying equations (2) and (13). Essentially, the statement is a direct consequence of the law of conservation of capital. Trading or active management does not create capital but merely redistributes it.

The conservation of capital holds only if there is no trading or if the trading is costless over the period. Otherwise, if investors traded stocks inter-period or over the period, then for the market as a whole, capital gains and losses for market timing or stock selection cancel, leaving investors in aggregate with a net loss from brokerages, other trading costs and taxes from realising capital gains. The asset weighted average return \( \overline{R}_t \) of all investors for the period \( t \) is given by

\[
\overline{R}_t = R_t - \tau_t.
\]  

(17)

The market return is \( R_t \) and \( \tau_t \) denotes the impact of transaction costs and capital gain taxes, leading to a loss of capital of \( \tau_t C_t \) over the period. Note this is the actual total net market return for all investors when price changes from trading only occur between periods. The result also holds when there is inter-period
trading (see Appendix). All quantities in (17) would vary with time. But since we expect \( \tau > 0 \) in any period, we would conclude \( R_t < R \) for all time periods. The total capital gain (and dividend income) of all investors must be less than the capital gain (and dividend income) of the total market, when costs are included.

Our result about costs is more than a hypothesis, as in Bogle’s “Cost Matters Hypothesis”, it is a mathematical certainty and hence it should be called a “Cost Matters Theorem” in accordance with the “relentless rules of humble arithmetic” (Bogle, 2005a; Sharpe, 1991). It is now clear why the proportionate shareholding approach is the optimal investment strategy for all investors collectively (but not necessarily individually), in a real world with transaction costs and taxes. As more and more investors adopt the strategy then \( \tau \to 0 \) and the market becomes more and more cost efficient. There is no other “buy and hold” strategy that can achieve equal returns at zero-cost for all investors, because of the full scalability of the approach. We can elaborate on this claim a little further here.

The proportionate shareholding approach is the only investment strategy that every investor can employ, though not all would as we will explain below. It is obvious that the minimum number of investors who can deviate from this strategy is two. In this case, as stock prices fluctuate, these two active investors would generally have to rebalance their portfolios by trading against each other, assuming it is feasible to do so. By having to pay transactions costs and capital gains taxes, these two investors together must under perform the rest the market, which does not need to trade by definition. Hence we have proved that the only investment strategy which maximizes the total return to all investors equally and collectively is the proportionate shareholding strategy. However, we do not anticipate that the market will ever likely to become totally passive with everyone adopting the proportionate shareholding strategy, for reasons discussed below.

Note we have not claimed that individual investors cannot “beat the market”, i.e. we have not shown that individual investor returns cannot exceed the market return. Some investors may be able to beat the market consistently even after costs, but only at the expense of other investors. Clearly, if individual investors can easily beat the market, then the case for adopting a passive investment strategy would not have broad appeal. However, if we can prove that individual investors cannot easily beat the market, then we would have a better case for advocating our approach as a universal investment strategy for almost everyone.

4. Individual Optimal Strategy

We will argue here that the collective optimal strategy is also an individual optimal strategy for almost everyone, in a fully scalable and equitable way. To make our argument compelling, we need to state our assumptions in the context of existing knowledge of investments and financial markets so that we do not leave any major questions unanswered.

It has been widely stated that individual investors cannot consistently beat the market, because the market is efficient. Indeed, if this were the case, the only rational investment strategy is passive. Some even assert (Malkiel, 2005) that “not being able to beat the market” is synonymous with “market efficiency”. However,
we take an almost diametrically opposite view that some individual investors can consistently beat the market, because the market is inefficient. But most individual investors, with a large amount of capital collectively, cannot consistently beat the market, because as we have shown above the market is inefficient due to costs. Clearly, much depends on how we define efficiency.

Efficiency when applied to the financial markets has several distinctly different meanings and in many discussions, they are often confused, as we will explain. We note at least four relevant meanings for finance where efficiency can be defined relative to price, information, mean-variance and cost. Each of these concepts of efficiency should be defined independently of each other, but they are often not clearly distinguished.

Price efficiency can be defined as a stable relationship existing between the cost of acquisition of a product and the benefit of acquisition. The consumer goods markets say for a litre of milk or for a motor vehicle are relatively price-efficient. We do not generally observe large price fluctuations over short periods of time, leading to great uncertainties about the cost of acquisition relative to its benefits. Financial markets are clearly not price-efficient over most time scales, sometimes even up to decades. The price of a stock (e.g. Bear Stearns recently) can be 15 times greater on a Friday than the following Monday. Given the number of stock market bubbles, the scandals of major stocks such as Enron and the write-downs in the prices of mortgage-backed securities in the recent credit crisis, it is difficult to sustain the idea that the market is price-efficient. Indeed many empirical studies (e.g. Shiller, 1981, 2005; Summers, 1986) formally support the observation that the financial markets are generally price-inefficient.

Information efficiency for a stock has been defined as the situation where the stock price fully reflecting all available information (Fama, 1970, 1991). The fact that “information” has not been clearly defined means we do not know precisely what is included in “information” and therefore cannot independently verify whether the market is information-efficient or not. Introducing the concept of different types of information still does not make the basic concept any more precise or testable. For example, if we spread a false rumour, say a company is being taken over, should the stock price reflect this information? If the price changes, then the market in reflecting this information is then price-inefficient. If the price does not change, then the market does not reflect all information. In many actual cases, the target stock price does rise, only to fall again later when the rumour is confirmed to be false. In the period between the price rise and fall the market is clearly price-inefficient.

Grossman and Stiglitz (1980) have shown that the cost of information gathering means financial markets cannot be information-efficient. This may explain why the market can be price-inefficient for long periods of time. If we assume that financial markets are price-inefficient, then whether or not it is information-efficient is not so important. In some discussions (Malkiel, 2005), the efficient market hypothesis often identifies information-efficiency with price-efficiency and sometimes even mean-variance efficiency in the equilibrium context.

Mean-variance efficiency for a portfolio is defined by the set of portfolios on the efficient frontier in modern portfolio theory (Markowitz, 1952). The portfolio
which is on the frontier and is still optimal allowing for borrowing or lending is called the market portfolio in CAPM. The market portfolio defined in section 2 above has not been optimized with respect to expected mean return and variance and hence it is not mean-variance efficient and it is distinct from the market portfolio of CAPM. Hence it should not be a surprise that our market portfolio is sub-optimal (Arnott et al., 2005; Hsu, 2006) in the sense that it is mean-variance inefficient.

Cost efficiency is the most common notion of efficiency in economics. We define the market as cost efficient if the total cost in investing for all investors is the minimum, where the total cost includes transaction costs, advisor fees, consultant fees, manager fees, operating costs, taxes and any other costs which are deducted from investment returns (not accounting for the full economic cost of research time and effort by individual active investors). That is the market is cost efficient if the total cost \( t \) in equation (17) is minimum. It is an objective of this paper to suggest an approach to improve the cost-efficiency of the market for the collective benefit of investors.

Clearly, with so many inefficiencies, the market can be beaten and some individual investors have done this systematically and consistently over long periods of time. The fact that these successful investors often adopt similar approaches (Buffett, 1984) indicates that we are not simply being “fooled by randomness” (Taleb, 2005). Most cited evidence in support of the notion of efficient markets comes from the investment performance statistics of professional investment managers. There is an implicit assumption (Malkiel, 2005) that if any group of investors should beat the market then it should be the professional investment managers. We will show below how this leads to the false conclusion that since the professional investment managers have not beaten the market, then the market must have been efficient.

An inefficient market argument has been advanced for passive investing by Thorley (2002), mainly on the basis that individual investors are likely losers as they have a competitive skill disadvantage compared to professional investment managers who are more likely to be winners. Our argument is very different. Even though the market is inefficient, individual investors might not be able to easily find professional investment managers to beat the market on their behalf, if there are structural inefficiencies in the process of financial intermediation. If this is the case then the collective optimal strategy proved above will also be shown to be the optimal strategy for all individuals, except for those who have the skills and resources to find good managers or are able to exploit market inefficiencies themselves. We will show the existence of problems of financial intermediation.

5. Financial Intermediation

Equilibrium theories of finance and economics assume that there are large numbers of price-taking individuals, all rationally optimizing their investment and consumption using complete and costless information, leading to efficient markets. When empirical evidence contradicts such theories, attempts have been made to develop theories where some of the assumptions are relaxed. Current alternative theories (Lo, 2005) are largely based on the notion of irrationality or bounded
rationality of individual behaviour. Even though there are merits in understanding individual behaviour, it is doubtful that irrationality (Shiller, 2005) can lead to a coherent understanding of observed regularities of markets. In this paper, we suggest another approach based on the behaviour of rational financial intermediaries to explain market inefficiencies and to support our suggestion to lower pension cost inefficiencies through passive investment.

It is self-evident that the behaviour of financial markets is determined largely by financial intermediaries and not by individuals. In Australia, about three quarters of total pension assets are managed by intermediaries in less than 1% (by number) of all superannuation funds and only one quarter of the pension assets are self-managed by individuals in more than 99% (by number) of funds (Sy, 2008a). Hence understanding the behaviour of financial intermediaries is more important (on an asset basis) than understanding individual investor behaviour in the superannuation market. Most financial intermediaries obviously try to understand their clients' objectives, but it is only practicable mostly on a collective rather than an individual basis. Financial advice based on any individual’s objectives has still to be implemented mostly collectively through other financial intermediaries.

Financial intermediation implies most individual objectives never lead directly to market transactions. This would contradict the fundamental assumption that the overall market behaviour is determined solely by the aggregate action of optimizing individuals. If the objectives of optimizing individuals are materially different from the objectives of financial intermediaries, as we will argue below, then the world described by equilibrium theories cannot be the same as the real world.

We will show that the market for the services of professional investment managers is structurally inefficient due to the presence of information asymmetry, where clients are not well informed about their investment products. This leads a resolution of the principal-agent conflict of interest which favours professional investment managers optimizing their own welfare ahead of optimizing the welfare of their clients. The consequence is that these professional investment managers mostly fail to beat the market and as a group, and they do not fully exploit market inefficiencies for the benefit of their clients.

The structural inefficiency of the market for investment products comes from several intrinsic and extrinsic factors, all of which create significant hindrances to investors making informed choices. Even if an investment manager has a sound and disciplined process to beat the market, that information is not available (e.g. due to commercial confidentiality) or understood by most unsophisticated investors. Moreover, market volatility or noise means that a superior investment strategy may take many years to emerge as being statistically significant. Markets can remain irrational longer than fund managers can remain solvent. Only investors like Warren Buffett with a strong conviction and a deep pocket can hold the course and exploit some of the longer-term market inefficiencies.

Chasing short-term opportunities to beat the market is the only viable business option for most professional investment managers, who are forced to become a class of professional speculators rather than real investors. Real investment, as defined for example in investments in plant and equipment, has a long-term objective of making earnings from operating cash flows, which provide stock dividends for the investor. The proper holding-period of an investment is long-term
or “forever”, according to Warren Buffett (Graham, 1973). Speculation, on the other hand, seeks capital gains by buying or selling securities for short-term profit. High turnover from speculation increases transactions costs which substantially reduce long-term returns, as we will prove below.

Apart from short-termism dictated by market realities, professional investment managers are also hampered by combinations of factors including asset size, transactions costs and their own business incentives. We will show how these factors operate individually and together to make beating the market by their investment managers, an unlikely outcome for most investors.

### Information Asymmetry

We have argued that the financial intermediation process has itself led to an information asymmetry, where the end investor does not fully understand the investment product. This reliance on intermediaries and the existence of uncertainties related to this process have led a short-term focus on investment performance as a way of monitoring the investor decisions. We will show that short-term investment performance monitoring provides little useful information for the investor, because of a combination of market noise and sampling errors over short periods of time, which creates a case of being “fooled by randomness”.

To illustrate the information problem from statistical sampling, we consider the simple example of a constant market return $R$ over $n$ periods. An initial total market capital of $C_0$ would accumulate over $n$ periods to $C_n$ given by

$$C_n = C_0 (1 + R)^n.$$  

(18)

A corresponding expression for the active market capital $C\overline{n}$ with a constant cost $\tau$ is given by

$$C\overline{n} = C_0 (1 + R - \tau)^n.$$  

(19)

The impact of costs which leads to capital leaving the market through lower share prices is to result in a lower total market capital than it otherwise would be. The reduction ratio is given from (18) and (19) as

$$\frac{C\overline{n}}{C_n} = \left(1 - \frac{\tau}{1 + R}\right)^n.$$  

(20)

A 2% active investment cost $\tau$ with a market return $R$ of 10% p.a. would lead to more than 30% reduction in market capital in 20 years. The numbers in this example are typical of those found in many empirical studies (Bogle, 1999). Note that apart from the constant parameter assumptions used, these results are mathematical certainties.

It is a mathematical certainty that the asset weighted average return net of costs must be less than the market return in any single period or in any number of
periods, provided the whole population of investors are included in the average. In empirical studies, we have to resort to sampling the population and therefore subject to sampling errors. However, the greater the total assets of the sample and the longer the time period of investment return comparison, the greater is the likelihood that the empirical data will confirm the result of (20).

To estimate how the probability of stochastic dominance of the market return over active returns scales empirically with asset size of the sample and the time period of return in random sampling comparisons, we use a continuous time model to simplify the mathematics of compounding returns.

If $S_t$ is the capital value of a random sample of portfolios of unskilled investors at time $t$ and $S_t^*$ is the capital value of a corresponding proportionate market portfolio (3), which is defined by equal initial capital values: $S_0^* = S_0$, then

$$x_t = \frac{S_t}{S_t^*}$$

(21)

is a stochastic variable, which we assume for illustrative simplicity to be a Gaussian variable diffusing under Brownian motion. This simplifying assumption is for the sake of convenience of structuring the argument and for making rough estimates, rather than for quantitative accuracy. It is well known that the stochastic variable evolves in time according to

$$z_t = \ln(x_t) = \frac{\ln(x_0) - (\mu + \frac{1}{2} \sigma^2) t}{\sigma \sqrt{t}}.$$  

(22)

The variable is a unit normal variate, where the standard deviation $\sigma$ quantifies the uncertainty introduced by random sampling and the drift parameter $\mu$ quantifies the non-random effect of active investment costs. From (20), we approximate the drift parameter for costs by

$$\mu = \tau/(1 + R).$$

(23)

The probability that the asset weighted average return of the random sample of portfolios is greater than the market return is given by the probability that $x_t > 1$ or $z_t > 0$, which has the value $N(z_t)$, given by the cumulative normal probability function.

It is useful to note the asymptotic property that $N(z_t) \to 0$ as $\sigma \to 0$ or $t \to \infty$, that is, larger samples and longer time periods lead to more certain results. We note that as the total asset of random sample of portfolios approaches the total asset of the market $\sigma \to 0$. Also the longer the timeframe of return comparison, the less the error from random sampling will matter in demonstrating the stochastic dominance of the market return.

In Table 1, we provide numerical examples of a few typical cases, where the drift parameter for costs is 2% or 3% and the dispersion of one-year returns of the random sample has a standard deviation of 5% or 7%. It can be seen that the
smaller the sample (the larger the standard deviation) and the lower the transaction cost, the higher the probability that the random sample will be seen to beat the market.

Table 1: Probability (%) of the Average Return of The Random Sample Exceeding the Market Return

<table>
<thead>
<tr>
<th>Years</th>
<th>Cost (top label)</th>
<th>Return Dispersion (next label)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>1</td>
<td>33.5</td>
<td>37.4</td>
</tr>
<tr>
<td>3</td>
<td>23.1</td>
<td>28.9</td>
</tr>
<tr>
<td>5</td>
<td>17.1</td>
<td>23.7</td>
</tr>
<tr>
<td>10</td>
<td>8.9</td>
<td>15.5</td>
</tr>
<tr>
<td>15</td>
<td>5.0</td>
<td>10.7</td>
</tr>
<tr>
<td>20</td>
<td>2.9</td>
<td>7.6</td>
</tr>
<tr>
<td>30</td>
<td>1.0</td>
<td>3.9</td>
</tr>
<tr>
<td>50</td>
<td>0.1</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Hence sampling errors in empirical studies can potentially give a misleading, but statistically inconclusive, impression that the group of selected professional investment managers has beaten the market, at least in the short term. However, in the longer term, say 30 years or more, the compound effect of costs overpowers any statistical error from random sampling and the probability is very small that any group of investors will be seen to beat the market. These observations and explanations are supported by empirical studies (Bogle, 1999).

In conclusion, we have shown that over short periods, say less than 5 years or less, groups of unskilled investors could by chance beat the market due to errors from random sampling. It could take up to 30 years, given our parametric assumptions, for systematic effects to overwhelm statistical noise. Hence it is difficult to distinguish a good investment manager from a bad one based on short-term statistical comparisons alone.

Size and Cost

Another common source of misleading statistical information for the investors comes from performance comparison tables which do not normally take into account fund sizes. A fund with $10 billion that beats the market by 1% is much more significant than another fund which achieves the same feat but with only $1 billion. A below market return by a very large-size fund can provide a lot of lost capital for many small-size funds to get above market returns. Consider a $10 billion fund under-performing by 1% and 10 $1 billion funds out-performing by 1%, before costs. This creates the false impression that active management has succeeded, since more than 90% of managers in the sample have beaten the market. Such misleading comparisons are common, widespread and practically universal in the industry.
To illustrate how size and cost in active management create a daunting task for professional investment managers, consider the following analysis. In any period $t$, there are winners whose returns beat the market return and there are losers whose returns are beaten by the market return. If we define capitals and average active excess returns net of costs of the group of winners and the group of losers by $C_w, \alpha_w, C_L, \alpha_L$ respectively then we have a mathematical identity for capital changes

$$C(R - \tau) = C_w (R + \alpha_w) + C_L (R + \alpha_L)$$  \hspace{1cm} (24)

on account of (2) and (17) while suppressing the time subscript. On noting that $C = C_w + C_L$ we can simplify (24) and write

$$\alpha_w = -\tau \frac{C}{C_w} - \alpha_L \frac{C_L}{C_w}.$$ \hspace{1cm} (25)

This result shows that due to the conservation of capital within the market and presence of transaction costs, for there to be any winners at all ($\alpha_w > 0$), we need to have losers, with negative excess return given by

$$-\alpha_L > \tau \frac{C}{C_L}.$$ \hspace{1cm} (26)

The ratio on the right of (26) is greater than one and $\alpha_L < 0$, by definition. For there to be any winners at all, the losers have to bear all the transactions costs (including those of the winners), say typically 2% p.a., with their own below market returns. As winning capital increases, the losses sustained by losing capital have to increase rapidly. For example, if two thirds of the active managers perform marginally above the market and if costs are 2% for the period, then the losers have to perform on average 6% below the market, from (26). As the pool of loser capital dwindles, the task of beating the market by the winners becomes eventually impossible.

This provides a rigorous explanation for the observation by Bogle (1999) that in professional investing: “nothing fails like success”. As a fund manager succeeds, more and more capital flows to the fund until the combination of fund size and transaction costs eventually causes it to fail. Even if performance leagues tables of fund managers were accurate, everyone will switch their money to top performing funds leading to increases in fund sizes to such levels that they will eventually fail. Picking winners from well-publicized performance leagues tables is unlikely to be a sound long-term investment strategy, because of the costs in switching to chase yesterday’s winners, who are unlikely to persist.

For example, the recent fundamental indexation approach (Arnott et al., 2005) or some hedge fund trading strategy may work, so long as they do not become too popular, with large amounts of capital using their methods. A well-known historical example (Gross, 2005) is the Fidelity Magellan fund, which had a great investment performance record in its earlier years, but grew to an enormous size of more than $100$ billion. Poor performances in the 1990s forced it to close to new
investors in 1997 and today the Magellan fund is only about half of its peak size.

A similar explanation can also be advanced for the emergence of boutique fund managers in Australia and elsewhere. It has often been observed in the Australian and other markets that an investment manager starting with limited investor capital establishes a record of good returns for several years. This attracts substantial fund inflows until large amounts of capital are being managed, with correspondingly large rewards to the company in taking more asset-based fees which are only partly passed onto the manager as performance bonuses. Sometimes these firms are also bought out by conglomerates, further increasing their size and scale of operation. As investment performance falters due to asset size and cost effects discussed above, the manager realises the difficulty of the task, resigns and sets up a boutique investment firm, which has now a much smaller amount of investor capital with which to beat the market. The manager has now an easier task and is more fully compensated by taking a larger share of the investment management fees.

Asymmetric Incentive

Even before an active fund manager reaches a “too large not to fail” status, it is often already hampered by a business asymmetric incentive, leading to a defensive strategy called “closet indexing”. Empirical studies exist (see e.g. Chevalier and Ellison, 1997) to suggest that the level of risk-taking by mutual funds changes in response to incentives. We will supplement such studies with an analytical model, which assumes that it is rational for a large fund manager to have as its primary goal: to stay in business by retaining funds under management. It is only a secondary goal to increase funds under management. This creates an asymmetric incentive, which is stronger for larger fund managers.

If a manager performs poorly in the sense of returning less than the market, the investors will withdraw funds from the manager leading to a threat to the primary goal of retaining funds. On the other hand, if a manager performs well in the sense of returning more than the market, the investors will make additional contributions leading to achievement of the secondary goal. Funds under management from existing investors are generally much greater than their future contributions or new fund flows. As primary goal is more important than the secondary goal, the asymmetric incentive leads to risk aversion on the part of manager relative to benchmark performance and the phenomenon of “closet indexing”. This problem occurs whether the manager has real skills or not and it is more pronounced for large fund managers.

To illustrate potential investment manager behaviour arising from asymmetric incentive, we consider a simple model shown in Figure 1, where the reward for performance relative to the benchmark is a linear function of the net active return. The reward function $r(x)$ is define by

$$
\begin{align*}
  r(x) &= \beta_1 x & x > 0 \\
  &= \beta_2 x & x < 0
\end{align*}
$$

(27)

where the gradients are different but are both positive. A reward symmetry ratio
can be defined by $\rho = \beta_+ / \beta_-$, where $\rho < 1$ and the smaller the ratio, the less symmetric the reward incentive.

Figure 1: Model of Active Return and Asymmetric Reward

The payoff $P$ for an investment manager faced with risk taking under such an incentive structure is given from (27) by

$$P = \int_{-\infty}^{\infty} r(x)f(x)dx = \int_{-\infty}^{0} \beta_-xf(x)dx + \int_{0}^{\infty} \beta_+xf(x)dx$$  \hspace{1cm} (28)$$

where $f(x)$ is a probability density function of the outcome from active risk taking by the investment manager. As our general conclusions will be robust against the details of $f(x)$, we assume for simplicity a shifted normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}s} \exp \left\{ -\frac{1}{2} \left( \frac{x-e}{s} \right)^2 \right\}$$  \hspace{1cm} (29)$$

where $e$ is the expected net active return (due to skills or confidence) of the manager and $s$ is the volatility or deviation from expectation from risk taking relative to the market benchmark. It is straightforward to evaluate the integral in (28) given (29) to find the ratio of the payoff to the risk taken as
\[
\frac{P}{s} = \beta kN(k) + \beta kN(-k) + (\beta_+ - \beta_-) \frac{\exp(-\frac{1}{2}k^2)}{\sqrt{2\pi}}
\]  

(30)

where \( N(k) \) is the unit cumulative probability function and we have introduced \( k = e / s \), which is the ratio of expected net return to the risk taken.

If the manager has negative expectations such that \( e \leq 0 \), an expected excess net return which is not positive, the first two terms in (30) are non positive \( (k \leq 0) \) and the last term is negative \( (\beta_+ < \beta_-) \). Hence the manager’s optimal strategy is to take no risk relative to the benchmark to avoid a negative payoff.

Even if the manager has positive expectations in the sense that \( e > 0 \), it may still rational for the manager not to take risk against the benchmark because of reward asymmetry. Using the symmetry ratio \( \rho = \beta_+ / \beta_- \) and the mathematical identity \( N(-k) = 1 - N(k) \), we can write equation (30) also as

\[
\frac{P}{\beta_- s} = k + (\rho - 1) \left[ kN(k) + \frac{\exp(-\frac{1}{2}k^2)}{\sqrt{2\pi}} \right]
\]  

(31)

The first term on the right hand side is positive \( k > 0 \) since \( e > 0 \) on assuming positive expectation. The second term is negative since \( \rho < 1 \) due to asymmetric incentive. Whether the manager should take risk \( s > 0 \) to beat the benchmark depends on whether \( e \) is large enough to make the right hand side of (31) positive. Let us assume the manager takes risk efficiently so that \( s = \tau + e \), where \( \tau \) is the cost of active management. For given symmetry ratio \( \rho \) and cost \( \tau \), we can solve for \( e \) which makes the right hand side (31) greater than zero and obtain the excess net and gross return hurdles which make active investment a rational proposition.

A sample of the results is shown in Table 2, where we see that high cost \( \tau > 2\% \) and low reward symmetry \( \rho < 0.4 \) would make active management unattractive.

### Table 2: Excess Gross Return Hurdle to Overcome Cost and Asymmetric Incentive

<table>
<thead>
<tr>
<th>Reward Symmetry Ratio</th>
<th>1%</th>
<th>2%</th>
<th>3%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>2.8</td>
<td>5.5</td>
<td>8.2</td>
</tr>
<tr>
<td>0.4</td>
<td>1.6</td>
<td>3.2</td>
<td>4.7</td>
</tr>
<tr>
<td>0.6</td>
<td>1.3</td>
<td>2.5</td>
<td>3.8</td>
</tr>
<tr>
<td>0.8</td>
<td>1.1</td>
<td>2.2</td>
<td>3.3</td>
</tr>
</tbody>
</table>

When a fund is newly established and asset size is low the reward symmetry ratio is closer to one, because the reward of fund flow due to investment performance is more symmetrical. As a fund becomes well established with a larger asset size, the symmetry ratio would become substantially smaller and the risk of money out flow due to under-performance is not adequately compensated by the reward of money.
inflow from out-performance. Hence as asset size becomes large, symmetric incentive decreases and the symmetry ratio falls, it is logical to reduce cost, target a lower excess return and take a lower level of risk.

The tendency to “hug benchmarks” has been widely observed in the industry, leading to the phenomenon called “closet indexing”, where active management fees (French, 2008) have been charged for largely passive management. Since the portfolios of many active investment managers are de-facto substantially passive relative to their benchmarks, the cost of residual active management is very high (Miller, 2007). Indeed, the market may already be de-facto substantially passively managed, but paying substantially higher active fees. The attempt to “separate alpha” by pension funds where active fees are paid for only active management and lower passive fees are paid for passive management, has led to the rapid growth of hedge funds in recent years. The national default option proposed in this paper saves paying any active fees, which is unnecessary collectively for most workers.

It is clear that maximizing operating profits of investment managers through increasing fees from funds under management is different from attempting to beat the market. The assumption that active investment managers have the primary objective of beating the market does not agree with general observations in the industry as well as our alternative assumption proposed here that active investment managers have the primary objective of maximizing operating profits. Since active investment managers are not primarily attempting to beat the market, the fact that they have not been beating the market cannot therefore be a valid reason for asserting that the market is efficient. Relatively poor performance of professional active managers has been a fundamental argument advanced in favour of the “efficient market hypothesis” (Malkiel, 2005).

In this section, we have shown the process of financial intermediation in professional investment management has inherent conflicts of interest such as the conflict between short-term business survival and long-term performance of client portfolios. The information asymmetry that exists due to the low information content of short-term performance statistics has led to a relative absence of vigorous competition in the market for the services of professional investment managers. The structural inefficiencies of the market make the identification of superior investment managers difficult for most investors.

In conclusion, we advocate the proportionate shareholding approach, not because we believe the market is efficient in any sense, but because it is the individual optimal strategy for most investors who have no special advantage in extracting better returns from the market. Because the market is inefficient with respect to price, information, risk and cost, there will be individuals who can exploit these inefficiencies for abnormal gains. Hence not everyone will adopt the totally passive strategy proposed here. The cost $\tau_r$ in (17) will never be zero, but can be reduced when more investors adopt our passive approach. The non-zero cost $\tau_r$ can be viewed as the cost that we have to pay to limit the extent of market inefficiencies through active investment management.
6. Toward a National Default Option

We have proved that the totally passive proportionate shareholding approach is a collective optimal strategy for all investors, not because the market is in any sense efficient, but because the strategy seeks to minimize the total cost inefficiency in the market. Even though the strategy is not necessarily optimal for every individual investor due to the assumed presence of market inefficiencies, it has been argued nevertheless that the strategy is optimal for most individuals who need to use directly or indirectly (through collective investments) professional investment managers as financial intermediaries.

The structure of the financial markets is such that most investors are forced to play a negative-sum game through the active investments of their financial intermediaries. The optimal game-theoretic strategy in such a situation is the maxi-min strategy, which is to maximize the minimum return of the investor and in this case it is the market return provided by the proportionate shareholding strategy. The principal-agent conflict of interest prevents the optimal solution to be adopted for most investors and the information asymmetry in an information-inefficient market prevents the situation from being widely recognised. The classic market failure may have led many individual investors who have sufficient resources to withdraw from the market to do it themselves in self-managed superannuation, which is the fastest growing sector of the industry. A recent study (Sy, 2008a) suggests the superior performance of self-managed small APRA funds is partly due to most small funds having more cost-efficient structures.

These reasons provide a rationale for government intervention to improve the cost-efficiency and competitiveness of the superannuation industry. There already exists precedence for government intervening in the pension product market in the creation of the Retirement Savings Account (RSA). This is an individual capital guaranteed account operated by authorised service providers such as banks. The main differences between an RSA and an ordinary bank account are the differences in tax treatment and additional pension regulation imposed on withdrawals. Key advantages of this product are certainty, simplicity and portability. Investors can change RSA service providers without cost.

The RSA product, which should earn close to the risk-free interest rate, could be supplemented by a Retirement Growth Account (RGA), which is a national growth fund managed on the proportionate shareholding strategy, providing a low-cost, high growth product, universally available to all individuals for superannuation. If adopted this fund would form the cornerstone of a national default option scheme available to all employees, as an alternative to any default options offered by their own employers. Like the RSA, we expect the RGA will be operated by authorised service providers under prescribed mandates.

The approach can potentially save all investors collectively about $20 billion per year (depending on the uptake), at current total asset sizes of superannuation and of the equity market. It could potentially double the terminal wealth accumulated by a worker through superannuation over a working life. Even though our discussions have been confined implicitly to the domestic stock market, they apply equally to other asset markets as well, such as the fixed interest market or international stock markets. The approach can be extended to create a fully diversified fund including other asset classes for the RGA or the RSA, without the
knowledge or calculation of capitalization weights.

The simplest version of a national default option would be for employers to direct the super guarantee levy on behalf of their employees as contributions to both the RSA and RGA. The contribution for each employee could be split with \((115 - \text{age})\) per cent directed to a RGA and the balance to a RSA, based on the investment life-cycle assumption. Other rule-based variations in asset allocation are also possible. The total cost to the employee could be targeted at around 0.1% of assets per annum as the typical management expense ratio, with potential entry and exit fees to cover and discourage trading in the accounts.

Such a national default option scheme would provide another important benefit to the industry, by creating an actual performance benchmark (net of fees and taxes) against which other offerings could be compared to facilitate proper functioning of a competitive market. The introduction of such a benchmark, which is simple for most investors to understand both in concept and performance, would provide a yardstick for all other pension fund products. More complex and sophisticated products will continue to exist as financial intermediaries continue to seek abnormal profits through innovative products. But the existence of a benchmark such as the national default option will encourage market forces to operate more competitively and efficiently.

To fully justify our proposal for a totally passive approach, we need to address the “free rider” argument often raised against passive investing in relation to liquidity and price-efficiency. The general argument is that active investing increases liquidity helping the price discovery process and therefore improves price-efficiency of the market. As passive investing does not contribute to this process it is therefore a “free rider”, obtaining an unpaid-for benefit. The implication that more passive investing would lead to greater price-inefficiency has little theoretical or empirical support.

The empirical evidence is just the opposite. One only needs to observe that during various phases of the last market bubble in technology stocks, turnover and liquidity reached historically high levels and yet as events subsequently proved the market was extremely price-inefficient. Theoretically, both information-inefficiency and price-inefficiency could rationally induce the “herd” instinct evident in the chartist’s momentum trading, which could lead to market instabilities (see e.g. Chiarella et al., 2008).

Price volatility and liquidity are important risk factors for speculators, particularly those who are highly leveraged. The risk factors are of little concern to the long-term investor who is seeking the rewards of business earnings and dividends over a long period of time. We could hardly express the matter better than Keynes (1936):

*Of the maxims of orthodox finance none, surely, is more anti-social than the fetish of liquidity, the doctrine that it is a positive virtue on the part of investment institutions to concentrate their resources upon the holding of ‘liquid’ securities. It forgets that there is no such thing as liquidity of investment for the community as a whole. The social objective of skilled investment should be to defeat the dark forces of time and ignorance which envelop our future.*
No amount of liquidity and active trading will change the business earnings of the market in any given period. Speculation merely re-distributes those earnings, in ways which some would regard as anti-social. Indeed, one could question (Burtless, 2003) whether a greater disparity of outcomes in individual retirement savings than is necessary, is desirable from the point of view of national interest, since more people falling in the lower tail of a more broadly spread outcome distribution would require more public assistance (e.g. social security or age pension) than is necessary. The outcome disparity further increases public costs in social safety nets.

Instead, better long-term investment returns can really be obtained ultimately from greater business profits, through better business operations and investments. Shareholders can increase their chance of real wealth creation through their demand for better corporate governance (Bogle, 2005b). We note in passing that the stable shareholdings implied by the proportionate shareholding approach can be used to facilitate the creation of a registry for the efficient transfer of voting rights to a new class of proxy capitalists, leading to a new form of representative capitalism.

7. Summary

In this paper, we have proved Bogle’s “Cost Matters Hypothesis” (Bogle, 2005a) and show that it is in fact a “Cost Matters Theorem”, because minimizing the total cost of investing leads to a collective optimal investment strategy for all investors, independent of assumptions about market efficiency. The strategy is a totally passive, proportionate shareholding approach which is fully scalable and available to all investors. Unlike an index approach, the proportionate shareholding approach can have more flexible portfolio structures to further lower transaction costs.

We argue that financial intermediaries, not considered by existing equilibrium theories of finance and economics, play an important role in understanding the overall behaviour of markets. The fact that professional investment managers have not beaten the market is not a valid support for the efficient market hypothesis (Malkiel, 2005), because a professional investment manager’s objective is to maximize business profits which is not necessarily the same as the assumed objective of trying to beat the market.

In an information-inefficient market with significant statistical noise (Bogle, 1999; Taleb, 2005), we suggest that a structural inefficiency persists with misaligned incentives of professional investment managers which leads to the presence of inferior investment products, not correctly priced. The complexity of most active investment processes is such that the regulatory demand for better disclosure in fees and costs may be difficult to comply with and therefore may not by itself solve the problem of information asymmetry. In an environment where information is costly (Grossman and Stiglitz, 1980), we suggest that the collective optimal strategy has also become the individual optimal strategy for most investors, because they should seek to maximize their minimum game-theoretic return.

Because we assume the market is inefficient, there will always be active investors (Buffett, 1984) who will try to extract arbitrage profits and in so doing limit the
extent of market inefficiencies. The overall cost of investing in the market will never be zero, but it can be substantially reduced by having a larger share of passive investing proposed in this paper. Instead of only around 10% of managed investments paying passive fees in Australia (Vanguard, 2008), the percentage paying lower fees can easily be several times higher, potentially saving the superannuation system as much as $20 billion per year, which is about the same order as current government payments in the age pension.

A passive hub with active spokes is often recommended by asset consultants in a “hub and spoke” strategy for large superannuation funds. The national default option suggested in this paper can be seen as a cost-efficient way to implement the hub in a “hub and spoke” strategy for superannuation on a national scale. The national default option is suggested to be operated by the Government, but instead be operated by selected authorised private-sector service providers under prescribed mandates, with many details to achieve a low-cost structure yet to be stipulated.

The fact that the market is not fully competitive and efficient suggests that there is an opportunity for government intervention to promote a more competitive and efficient market. The approach we have proposed removes layers of financial intermediation costs and agency risks and leads to a national default option based on a simple proportionate shareholding strategy. A national default option is envisaged provide a simple and easy-to-understand alternative, in addition to existing alternatives, available to all individuals, particularly those in the early stages of wealth accumulation for retirement. It would be particularly useful for employers who do not want to nominate a default fund or who prefer to use a ready-made choice for their employees. It provides a choice for people who do not want to make choices. The option will be seen also to provide a performance benchmark to promote competitiveness and efficiency of the superannuation industry.

8. References


Chiarella, C., Xue-Zhong He and Min Zheng (2008), Heterogenous expectations and exchange rate dynamics. Quantitative Finance research Papers, University of technology, Sydney.


Appendix

We provide a proof of equation (17) when there is inter-period trading. The relationship (17) is obviously correct when trading occurs only between periods, when some stocks are assumed to be traded and portfolios are adjusted at the end of one period and held fixed until the end of the next period.

To show that the result is true also when there is inter-period trading, we assume that for a given period $t$ in (17) there are $n$ sub-periods, which need not be regularly spaced, when there is no trading inter-period and trading only occurs between sub-periods, then (17) holds (including dividends) for each sub-period:

$$\bar{R}_i = R_i - r_i \quad (i = 1, 2, \ldots, n) \quad (A1)$$

The net return of all investors $\bar{R}_i$ for the period $i$ is defined by

$$\bar{R}_i = \prod_{i=1}^{n} (1 + R_i) - 1 \quad (A2)$$

Substituting (A2) into (A1) and using a similar definition as (A2) for $R_i$, we have proved equation (17) holds with inter-period trading, except that the transaction cost $r_i$ has to be determined by costs incurred between sub-periods:

$$r_i = \prod_{i=1}^{n} (1 + R_i) - \prod_{i=1}^{n} (1 + R_i - r_i) \quad (A3)$$

To leading orders, assuming small sub-period transaction costs $r_i$ $(i = 1, 2, \ldots, n)$ the total transaction cost $r_t$ of the period $t$ is given by:

$$r_t = \sum_{i=1}^{n} r_i (1 - \sum_{j=1, j \neq i}^{n} R_j) \quad (A4)$$