Do Australian fund managers create value?

Author: Matthew Martineer
Student ID: 3279900

Supervisor: Prof. Richard Holden

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Economics (Econometrics) (Honours)

November 4, 2013
I declare that this thesis is my own work and that, to the best of my knowledge, it contains no material that has been published or written by another person(s) except where due acknowledgement has been made. This thesis has not been submitted for award of any other degree or diploma at the University of New South Wales or at any other educational institution. I declare that the intellectual content of this thesis is the product of my own work except to the extent that assistance from others is acknowledged.

Matthew Martineer

November 4, 2013
Acknowledgements

First and foremost, I would like to acknowledge my supervisor Professor Richard Holden whose devotion, belief in my research topic and extraordinary depth of economic knowledge has been of great inspiration and assistance in the writing of this thesis.

Second, to the donors of the ASB Honours Scholarship who have provided financial support throughout my honours year, and to Professor Denise Doiron for providing the recommendation and joining me for dinner during the semester.

Finally, I would like to thank Professor Jerry Parwada, Dr David Simmonds and Mr Eric Tan for providing assistance with locating the numerous datasets used in this study and for providing guidance in the early stages of the thesis.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaration</td>
<td>i</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>2</td>
</tr>
<tr>
<td>2 Literature Review</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Performance evaluation using a single market factor</td>
<td>4</td>
</tr>
<tr>
<td>2.2 Performance evaluation using multi-factor models</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Australian literature</td>
<td>18</td>
</tr>
<tr>
<td>3 Carhart’s (1997) four-factor model</td>
<td>21</td>
</tr>
<tr>
<td>3.1 The four factors</td>
<td>21</td>
</tr>
<tr>
<td>3.1.1 Constructing the four factor portfolios for the Australian market</td>
<td>22</td>
</tr>
<tr>
<td>3.1.2 The Australian factor portfolios</td>
<td>27</td>
</tr>
<tr>
<td>3.2 Mutual Fund Data</td>
<td>31</td>
</tr>
<tr>
<td>3.3 Standard estimation procedure</td>
<td>33</td>
</tr>
<tr>
<td>4 Accounting for Luck</td>
<td>40</td>
</tr>
<tr>
<td>4.1 Motivation for the bootstrap</td>
<td>40</td>
</tr>
<tr>
<td>4.2 Residual Resampling</td>
<td>43</td>
</tr>
<tr>
<td>4.3 Analysis of individual funds</td>
<td>45</td>
</tr>
<tr>
<td>4.4 Analysis of the entire industry</td>
<td>48</td>
</tr>
<tr>
<td>5 Further Analysis</td>
<td>56</td>
</tr>
<tr>
<td>5.1 Industry Value Added</td>
<td>56</td>
</tr>
</tbody>
</table>
5.2 Performance by investment style .............................................. 58
5.3 Performance by other observable characteristics ...................... 64

6 Conclusion ............................................................................. 67

A Additional Figures ................................................................. 69
List of Figures

| 3.1 | The evolution of the four factor portfolios | 28 |
| 3.2 | Histograms of the four factors | 29 |
| 3.3 | Scatterplot correlation matrix of the four factors | 30 |
| 3.4 | Growth in Australian equity mutual funds | 33 |
| 3.5 | Cross-section of estimated alphas and $t$-stats | 37 |
| 4.1 | Bootstrapped Alpha Distributions at various percentiles | 49 |
| 4.2 | Bootstrapped Alpha $t$-Statistic Distributions at various percentiles | 50 |
| 4.3 | Density and cumulative distributions of the actual and simulated alphas | 51 |
| 4.4 | Actual vs. simulated annualised alphas | 52 |
| 4.5 | Actual vs. simulated alpha $t$-statistics | 55 |
| 5.1 | Actual and simulated alphas at each percentile | 57 |
| 5.2 | Probability density functions of alphas and $t$-statistics by size tilt | 60 |
| 5.3 | Cumulative distribution functions of alphas and $t$-statistics by size tilt | 61 |
| 5.4 | PDF’s of alphas and $t$-statistics by book-to-market tilt | 62 |
| 5.5 | CDF’s of alphas and $t$-statistics by book-to-market tilt | 62 |
| 5.6 | Histograms of management fees and net asset values | 64 |
| 5.7 | Morningstar analyst ratings vs. alpha and alpha $t$-statistics | 66 |
| A.1 | Management fee vs. estimated alpha by investment style | 69 |
| A.2 | Management fee vs. alpha $t$-statistic by investment style | 69 |
| A.3 | Log net asset value vs. estimated alpha by investment style | 70 |
| A.4 | Log net asset value vs. alpha $t$-statistic by investment style | 70 |
| A.5 | PDF’s of alphas and $t$-statistics by management fee | 71 |
| A.6 | CDF’s of alphas and $t$-statistics by management fee | 71 |
| A.7 | PDF’s of alphas and $t$-statistics by net asset value | 72 |
| A.8 | CDF’s of alphas and $t$-statistics by net asset value | 72 |
List of Tables

3.1 Cross-correlation matrix of the four factors .......................... 27
3.2 Australian vs. US factor portfolios, 1986–2013 .......................... 29
3.3 Australian vs. US factor portfolios, 1986–2013 .......................... 31
3.4 Loadings on the four factors ................................................. 35
3.5 Diagnostics of the four-factor models ................................. 39
3.6 Evidence of skill at the 5 and 10 per cent levels ......................... 39

4.1 Testing for skill vs. luck at various percentiles in the cross-section ... 45
4.2 Inferring skill using parametric and bootstrap approaches ............... 47
4.3 Comparison of observed annualised alphas with distribution under luck alone ................................................. 53

5.1 Analysis of results by investment style .................................. 63
5.2 Performance by management fee and net asset value ..................... 65
Abstract

I apply a cross-sectional bootstrap resampling technique to explicitly control for luck and the common risk factors that influence the returns of Australian equity mutual funds. I find strong evidence that a small proportion of predominantly small-cap funds exhibit strong performance that cannot be explained by chance or exposure to risk. At the other end of the distribution, I find that the poorest funds perform ‘less worse’ than what is predicted by chance alone, but identify a large number of mediocre funds (e.g. those who perform marginally above their benchmarks) who perform worse than what would be expected solely by chance. I generalise these results to develop a new method to approximate the value created by managed funds, which for the funds included in this study I estimate at $95 million a month, compared to the aggregate net assets under management of $150 billion. The study also calls into question the findings of previous Australian studies—all of which have used a parametric approach to evaluate mutual fund performance leading to a negatively biased number of both outperforming and underperforming funds.

The code for all of the empirical work performed in this study was written by Matthew Martineer on Stata. The 113K .do file, as well the three .dta files containing the main results can be downloaded from [http://research.economics.unsw.edu.au/richardholden/miscellania/martineer.zip](http://research.economics.unsw.edu.au/richardholden/miscellania/martineer.zip)
The extraordinary performance of a small group of Australian equity mutual funds raises the question of whether such performance provides evidence that the managers of these funds have the ability to consistently identify mispriced stocks, or whether they are merely the luckiest of a large group of fund managers. This paper conducts the first comprehensive investigation of the ex-post performance of individual Australian mutual funds that explicitly controls for luck without imposing an ex-ante parametric distribution from which the abnormal returns of these funds are assumed to be drawn. Specifically, the returns of each Australian mutual fund are additively decomposed into five portions, mostly in accordance with Carhart’s (1997) four-factor model: (i) the portion that is a result of the manager’s active (stock-selection) decisions, or the abnormal return (ii) a market component, (iii) style components (i.e. size and book-to-market factors), (iv) the portion due to one-year momentum in stock returns, and (v) the portion unexplained by the model, or the idiosyncratic return. A technique involving non-parametric bootstrap simulations of the idiosyncratic and total returns is then used to identify the funds which generate high (or low) abnormal returns simply due to luck (as a result of the randomisation inherent in the returns of stock prices), as opposed to actual stock-selection ability.

The issue of mutual fund performance evaluation is of principal importance to investors and major plan sponsors (such as pension plans, endowments and foundations) and is central to the debate over the efficacy of Eugene Fama’s (1970) Efficient Markets Hypothesis. The phenomenon that many funds continue to attract
new investors whilst consistently generating returns below their benchmarks is well-documented and is an important issue in the field of behavioural finance. This issue is also explored in reference to Sharpe’s (1991) accounting identity, which implies that active investment management is a negative sum game after costs, and that if the aggregate Australian active mutual fund market generates positive excess returns (before costs), it is at the expense of active investments elsewhere.

Despite the fact that the Australian mutual fund market manages more than AUD2 trillion worth of assets as of March 2013, making it the fourth largest mutual fund market in the world, the performance of Australian funds on an individual level has received surprisingly little attention. In addition, many past studies have made errors in controlling for the common risk factors in stock returns, resulting in a biased measurement of abnormal returns, which in turn gives a biased measure of how many funds exhibit superior performance. These two factors, as well as the simulation technique which has far greater power properties than standard inference procedures and doesn’t make any assumptions about the ex-post distribution of excess returns, are the primary motivations for this study.

By applying Carhart’s (1997) four-factor model, the study finds that the main drivers of the total returns of Australian mutual funds are the return of the market, one-year momentum in stock returns and their exposure to value stocks relative to growth stocks, rather than true superior (or inferior) performance. While the excess returns of almost all funds are clustered around zero, there are six funds that consistently generate excess returns in excess of 10% on an annualised basis, including one which consistently generates returns larger than 20% per year. And whilst Monte Carlo simulations show that the distribution of abnormal returns that would be generated under luck alone is statistically different from the actual distribution, the practical significance of the difference is debateable.
CHAPTER 2

Literature Review

2.1 PERFORMANCE EVALUATION USING A SINGLE MARKET FACTOR

The first methods of evaluating the risk-adjusted performance of mutual funds using mean-variance criteria were based on the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Litner (1965b). The CAPM was developed by extending Markowitz’s (1952, 1959) pioneering work on Modern Portfolio Theory, which showed that the variance of the rate of return was a reasonable measure of a portfolio’s risk under a reasonable set of assumptions. This was used to develop a model to identify the efficient (mean-variance optimising) set of portfolios, known as the efficient frontier of risky assets, by identifying the set of portfolios which offered the highest expected return for each level of risk.

Tobin (1958) showed that there is a single risky portfolio that lies on the efficient frontier that is optimal for all investors, irrespective of their degree of risk aversion. This result assumed perfect competition, no capital gains taxes or transaction costs, the existence of a risk-free asset and that all of the assumptions of the Markowitz model hold. This single risky portfolio is then combined with the risk free asset, with more risk-averse investors putting a higher weight on the risk-free asset, and less risk-averse investors putting a higher weight on the risky portfolio. This may

---

1These assumptions are: Investors base their decisions solely on expected return and the variance of returns (i.e., are mean-variance optimisers), are risk-averse, and have a single-period time horizon.
include a negative weight on the risk-free asset in order to borrow money to leverage the purchasing of the risky asset. Tobin (1958) referred to this as a Separation Theorem (now often referred to as the one-fund theorem), since the selection of an optimal portfolio consists of two independent tasks: (1) Identifying the optimal risky portfolio, and (2) Choosing the weight on the risk-free asset given the investors’ level of risk aversion. It is also assumed (see Litner 1965b for a more detailed discussion) that the investor is able to borrow unlimited amounts at the risk-free rate, which is equal to the rate on savings deposits.

Sharpe’s (1964) derivation of the CAPM directly follows from this result. By varying the weight allocated to the risk-free asset, the relationship between the expected return and the standard deviation of the single risky portfolio can be derived (see Sharpe 1966 or Cochrane 2005 for details). It is assumed that all investors correctly identify the single optimal portfolio, so its expected return, \( E(r_p) \), can be expressed in the form \( r_f + b\sigma_p \), where \( r_f \) is the risk-free rate, \( \sigma_p \) is standard deviation of returns, and \( b \) is the risk premium from holding the risky asset. It can be easily found that \( b \) is equal to \( E(r_i) - r_f / \sigma_i \), which is always positive since investors are risk averse (see Sharpe 1964). Since the risky portfolio must be equal to the market portfolio (because this is the only fund that investors invest in), the CAPM describes the behaviour of asset prices in equilibrium, and the expected return of the combined portfolio may be written as follows:

\[
E(r_p) = r_f + \frac{E(r_M) - r_f}{\sigma_M} \times \sigma_p
\]  (2.1)

Equation (2.1) is known as the capital market line. \( E(r_M) \) and \( \sigma_M \) represent the expected return and variance of the market portfolio, and \( r_f \) is the risk-free rate. It is often shown diagrammatically with \( E(r_p) \) on the vertical axis and \( \sigma_p \) on the horizontal axis, with the coefficient on \( \sigma_p \) bring the slope of the capital market line.

\(^2\)See Sharpe 1964 for a theoretical justification of this assumption, or Fama 1970 for a more extensive treatment. The assumption is equivalent to that all investors agree on the probabilistic structure of assets.
In well diversified portfolios, the only relevant risk from a mean-variance standpoint is the exposure to macroeconomic (or market) conditions, which is known as systematic risk. Fama (1976) and Roll (1977) argue that this is the most important implication of the CAPM—that that the value-weighted market portfolio is mean-variance efficient in market equilibrium. Non-systematic risk will tend to zero as a portfolio becomes more diversified. The beta of a portfolio (or single asset) measures the extent to which its returns and the returns of the market move together. Since the investor is holding the market portfolio in (1), the only dimensions that need be of concern are expected return and beta, so $\sigma_p$ may be replaced with $\beta_p$ and $\sigma_M$ may be replaced with $\beta_M$, which is equal to 1. Making these substitutions and re-arranging the equation yields the CAPM:

$$E(r_p) - r_f = \beta_p [E(r_M) - r_f]$$

(2.2)

The equation states that the risk premium on a well-diversified portfolio is proportional to the risk premium, $E(r_M) - r_f$, on the market portfolio, and the beta coefficient of the portfolio. The implication is that sensitivity to market volatility is the only measure of risk that is needed to explain average returns. Using the fact that (1) passes through the risk-free rate and is tangential to the efficient frontier of risky assets, which Sharpe (1964) referred to as the tangency condition, it can be derived using first-order conditions that $\beta_p$ in (2) is equal to $\sigma_{p,M}/\sigma_M^2$, where $\sigma_{p,M}$ is the covariance between the returns of the risky portfolio and the market portfolio, and $\sigma_M^2$ is the variance of the market portfolio. Far more technical treatments of this derivation are provided in; inter alia, Mossin (1966) and Cochrane (2005).

Cochrane (2005) interprets the CAPM as a single-factor model, where the market factor acts as a proxy for marginal utility growth. The model is derived by

---

3 The CAPM may also be derived in reference to an individual security: $E(r_i) - r_f = \beta_i [E(r_M) - r_f]$, with subscript $i$ denoting an individual security, but this is not relevant for the purposes of this study.

4 Or equivalently, that the slope of the curve of the efficient frontier of risky assets is equal to the slope of the capital market line at the point $M$, where $M$ is the market portfolio.
modelling investors using a utility function defined over current and future values of consumption, that is, \( U(c_t, c_{t+1}) = u(c_t) + bE_t[u(c_{t+1})] \), where \( u(c_t) \) is consumption at date \( t \). The CAPM is then derived by observing the link between asset returns and consumption (see Lucas [1978]), and by treating the CAPM as a specialisation of a consumption-based model where an investor wishes to maximise his expected discounted utility by taking first-order conditions. Using this method, factor pricing models look for variables that are reasonable proxies for aggregate marginal utility growth. That is, variables where (a) marginal utility growth is a linear combination of some number of factors, and (b) the approximation is reasonable and can be interpreted economically. Cochrane (2005) argues that the market portfolio provides a reasonable interpretation under any of the following assumptions: (i) two-period quadratic utility, (ii) two periods, exponential utility and normal returns, (iii) infinite horizon quadratic utility and i.i.d. returns, or (iv) log utility; and provides the derivation under each case. Under this interpretation, the CAPM is a general equilibrium model with linear technologies and rates of return.

Sharpe (1966), Treynor (1966) and Jensen (1968) developed different procedures to evaluate risk-adjusted mutual fund performance using mean-variance criteria shortly after the CAPM was developed. Each of these performance measures require a long history of consistent management and investment style, and a representative sample of investment environments (e.g. both bull and bear markets) to be reliable.

Sharpe (1966) measured mutual fund returns per unit of risk and is known as the Sharpe ratio. The previous assumption that asset prices are in equilibrium (i.e. that all investors hold the same portfolio of risky assets) is dropped. To compute the Sharpe ratio (SR) of a given fund, the average excess annual return \((\bar{r}_p - \bar{r}_f)\) is divided by its standard deviation of annual returns \((\sigma_p)\) which gives a ‘reward-to-variability’ ratio, with a higher SR denoting a higher level of performance:

\[
SR = \frac{\bar{r}_p - \bar{r}_f}{\sigma_p}
\]  
(2.3)
A theoretical justification for using the SR as a measure of mutual fund performance is given as follows. Assuming that investors can borrow and lend freely at the risk-free rate, it is possible to attain any point along the line (on $E(r_p)$ vs. $\sigma_p$ space) from the risk-free rate (with expected return $r_f$ and standard deviation zero) to point A, where A represents the expected return and standard deviation of some mutual fund. If the line drawn from $r_f$ to fund A has a higher gradient than the line drawn from $r_f$ to fund B, the rational investor will strictly prefer investing in mutual fund A to mutual fund B. This is due to the assumption that the investor can borrow and lend freely at the risk-free rate, which allows combinations of fund A and the risk-free asset to provide a higher return at all levels of risk, compared with combinations of the risk-free asset and fund B.\footnote{The derivation is similar to that of (2.1).} The gradient of the line from the risk-free rate to any portfolio (again, in $E(r_p)$ vs. $\sigma_p$ space) is the Sharpe Ratio. Note that in equilibrium, when all investors hold the same risky portfolio, the market portfolio has the highest Sharpe ratio. \text{Treynor’s (1966)} measure is computed in the same way, but uses systemic risk ($\beta_p$) rather than total risk to compute excess return per unit of risk.

\text{Jensen’s (1968)} measure is derived from a direct application of the theoretical results of the CAPM. Since the CAPM will hold when estimating the systematic risk of any individual security or unmanaged portfolio, the constrained time-series regression estimate of $\beta_p$ in (2.2) can be used as an efficient estimate of a portfolio’s exposure to systematic risk. If a mutual fund manager is a superior performer, the manager’s portfolio will earn more than the risk premium implied by the CAPM given its level of risk. Since the allowance of such performance can be made by simply not constraining the regression equation to pass through the origin, we can allow for the possible existence of a non-zero constant by estimating the following regression:

$$r_{p,t} - r_{f,t} = \alpha_p + \beta_p [r_{m,t} - r_{f,t}] + \varepsilon_t$$  \hspace{1cm} (2.4)
In Equation (3.2), \( r_p \) represents the time series of the realised returns of a given fund (e.g. quarterly or monthly), \( r_f \) represents the risk-free rate, and \( r_{m,t} \) is the return of the market proxy (Jensen used the returns on the S&P 500 price index reinvesting dividends—similar measures continue to be used as market proxies). The estimated market beta parameter \( \hat{\beta}_p \) is the estimated market beta parameter and is assumed to be constant through time. Jensen (1968) showed that this assumption is empirically valid on the returns of US mutual funds (see Ferson and Schadt, 1996 for an extensive discussion of this issue). Recall that in Sharpe’s (1964) theoretical derivation of the CAPM, \( \beta_p \) was defined to equal \( \sigma_{p,M}/\sigma_M^2 \), which is precisely how it is estimated using simple linear regression. \( \varepsilon_t \) is the residual series (which by construction, are equal to zero in expectation), and were assumed to be serially uncorrelated on theoretical grounds.

However the parameter that is of primary interest in the evaluation of mutual fund performance is the estimated alpha, \( \hat{\alpha}_p \), known as Jensen’s alpha. The parameter is an estimate of the fund’s average incremental rate of return per period that is due solely to the manager’s ability to forecast future security prices. Or equivalently, the average return above or below that which is predicted by the CAPM.

Jensen found that US equity mutual funds were not able to predict security prices well enough to outperform a passive (buy the market and hold) strategy from 1945–64. Using a sample of 115 open-ended funds, he found that the average estimated alpha was \(-0.011\) with a median value of \(-0.009\). The mean values of \( \hat{\beta} \) and \( R^2 \) were 0.84 and 0.87 respectively.

Using this method to evaluate mutual fund performance of course assumes that the CAPM is empirically valid, that is, it characterizes the true data generating process of the excess returns generated by actively managed funds. However since the development of the CAPM, several empirical contradictions of the model have been
identified leading to the development of additional empirical asset pricing models that consider further undiversifiable risk factors in expected stock returns.

2.2 Performance evaluation using multi-factor models

The first empirical tests of the CAPM were not particularly successful. Lit-\-ner (1965a) found that the relationship between the annual excess returns of 301 US stocks and the S&P index over 1954–63 were incredibly disperse—the market index explained less than 25% of the variance of a quarter of the stocks, and less than half the variance of over 60% of the stocks. Miller and Scholes (1972) argued that this was due to the instability in beta coefficients and the high degree of residual variance when running the CAPM on individual stocks, but predicted that these issues would be corrected using portfolios of stocks.

However when Reinganum (1981) tested the CAPM using portfolio returns, he found a weak relation between portfolio and market returns, and strong persistence in abnormal returns. Rather than concluding that the market was inefficient, Reinganum argued that the CAPM is misspecified and that alternative models of capital market equilibrium should be “seriously considered”. Tests performed using post-1963 data by Lakonishok and Shapiro (1986) and Fama and French (1992) did not even support the most basic prediction of the CAPM—that average stock returns are positively correlated with market returns. Ross (1976) argued that the CAPM’s empirical failures were due to (amongst other factors) the fact that the true market portfolio is unobservable since it consists of all individual assets, and that the use of a market proxy will give incorrect inferences. Hansen and Richard (1987) argued that even if the market portfolio was observable, the fact that the information sets of agents are unobservable makes empirical testing the CAPM impossible.

The empirical failures of the CAPM led to the identification of non-market strategies and characteristics that seemed to explain the behaviour of average returns. Whilst it was common for researchers to identify strategies/characteristics that gave high
average returns, later to be corrected by studies that showed that the strategy simply had a high beta (see Cochrane 2005), some anomalies were shown to exist even after accounting for the market beta. However the strong possibility that factors, state variables or sources of undiversifiable risk beyond movements in the market portfolio are required to explain why some groups have stocks have higher average returns than others has been recognised in asset pricing theory since Merton (1971, 1973).

One of the most prominent examples is the size effect first identified by Banz (1981), also referred to as the small-firm effect or the small-cap premium. Banz found that the average returns of stocks with a high (low) market capitalisation were lower (higher) than that predicted by the CAPM. Specifically, using all common stocks listed on the New York Stock Exchange between 1936 and 1975, he found that stocks in the bottom quintile portfolio by market capitalisation earned a risk-adjusted return that was 40 basis points a month higher than the other firms. Reinganum (1981) analysed US stocks from 1963–77 and found that the smallest size decile outperformed the largest decile by 1.77% per month. Lamoureux and Sanger (1989) found a size premium of 2.0% and 1.7% per month on NASDAQ and NYSE/Amex stocks respectively over the period 1973–85. Fama and French (1992) found that the smallest size decile outperforms the largest size decile by 0.63% per month over the period 1963–90 using US stocks. Dijk (2011) provides a survey of international studies on the size effect, finding that the size effect exists for every country in which it has been studied, subject to a dataset of at least ten years.

The most common explanation for the existence of the size effect is that it is caused by market microstructure issues (which are not considered in the CAPM), namely transaction costs and liquidity risk. Stoll and Whaley (1983), for example, find that it is not possible to earn abnormal returns on NYSE small stocks after considering transaction costs and Pastor and Stambaugh (2003) find that portfolios of small firms have the highest loadings a liquidity factor based on trading volume.  

This is the market value of the firm’s equity, i.e. the stock price times the number of shares outstanding.
Despite this, several studies contend that the size effect is a statistical fluke. Lo and MacKinlay (1990), and MacKinlay (1995) argue that publication bias is a major factor; and Keim (1983) argues that the US size premium has varied too much over time to give evidence of a size effect. Several recent papers have suggested that the size effect disappeared after the early-mid 1980’s, both in US and international markets (see inter alia, Dimson and Marsh (1999); Dimson, Marsh, and Staunton (2002); Eleswarapu and Reinganum (1993). Karolyi et al (2012) attributes this to the increased liquidity of capital markets resulting in lower bid-ask spreads and commissions on small-cap stocks. However Dijk (2011) notes that the size effect has made a “remarkable comeback” in the US during the 2000’s.

A further anomaly was found by Rosenberg, Rosenberg, Reid, and Lanstein (1985), who identified a positive relationship between the average returns on US stocks and the ratio of a firm’s book value of common equity (BE) to it’s market value of equity (ME). They found that between 1973 and 1984, the strategy of picking stocks with high BE/ME (book-to-market) ratios yielded an average excess return of 36 basis points per month. Fama and French (1992), in examining the cross section of expected stock returns between 1963 and 1990, established that the positive relationship between book-to-market ratios and average returns persists in both univariate and multivariate tests, and is even stronger than the size effect in explaining returns.

Capaul, Rowley, and Sharpe (1993) extend the analysis of book-to-market ratios across international markets, finding that value stocks (stocks with high book-to-market ratios) earned higher returns than growth stocks (stocks with low book-to-market ratios) in every market they analysed between 1981 and 1992. They found that the global added return of high book-to-market stocks (by subtracting the returns on a growth index from a value index) was 1.88%. The justification for characterising stocks with a low (high) BE/ME ratio as growth (value) stocks is given in Fama and French (1992), who find that low BE/ME stocks have persistently
high earnings on book equity that result in high stock prices relative to book equity, whereas a high BE/ME is associated with a persistently low earnings on book equity that result in low stock prices.

If stocks are priced rationally, a stock’s BE/ME ratio should be a direct indicator of the relative prospects of firms. Fama and French (1995) show that BE/ME is a proxy for relative distress, where firms with persistently low (high) earnings tend to have high (low) BE/ME. This is consistent with Chan and Chan (1991) who find covariation in returns related to relative distress that is not captured by the market return, but is compensated for in average returns.

Heaton and Lucas (1997) note that whilst the distress of a particular firm is idiosyncratic and therefore can be diversified away, if a credit/liquidity crunch or a flight to quality arises, stocks in financial distress (i.e. high BE/ME stocks) will perform particularly poorly. Since an investor’s income is likely to be sensitive to these financial events, this is precisely the time at which the investor would not want his stock to become worthless, so he will demand a substantial premium to hold value stocks. The value premium is also present in international returns (see Asness, Moskowitz, and Pedersen (2009); Fama and French (2012).

As with the size effect, there is some controversy about whether the value-premium is a result of mispricing. Lakonishok, Shleifer, and Vishny (1994) argue that value stocks are priced irrationally since the market fails to understand the temporary nature of earnings growth in the years prior to portfolio formation. De Bondt and Thaler (1985) notes that growth stocks are often chosen when ‘naïve’ strategies, such as assuming a trend in stock prices or overreacting to news are followed which drives down the prices of growth stocks relative to value stocks. Fama (2011) rejects this view, arguing that it presupposes that investors never learn about their behavioural biases, which is necessary to explain the persistence of the value premium. Barberis and Shleifer (2003) argue that many investors simply get utility from holding growth stocks (which they call ‘glamour’ stocks), which tend to be profitable and fast-
growing and can have persistent effects on asset prices. However the controversy over whether value strategies have produced higher returns due to behavioural biases or because they are fundamentally riskier is yet to be resolved.

A further relationship that is unexplained by the CAPM is the positive relation between average return and leverage, documented by Bhandari (1988). Whilst the CAPM implies that leverage should be captured by the market beta, Bhandari finds that leverage helps to explain the cross-section of average stock returns in tests that include both size and market beta.

Finally, Basu (1983) shows that earnings-price ratios ($E/P$) help to explain the cross-section of average returns, even when controlling for size and the market. Ball (1978) notes that since $E/P$ is likely to be higher for stocks with higher risks and expected returns, $E/P$ acts as a catch-all proxy for non-market factors in expected returns.

Following these findings, Keim (1988) and Fama and French (1992) argued that since size, BE/ME, leverage and $E/P$ all use the stocks price, which is inversely related to the expected stock return once expected dividends are accounted for, to explain expected returns, each of these variables are natural candidates to expose the failures of the CAPM. Using the cross-sectional regression approach of Fama and Macbeth (1973), Fama and French (1992) find that the combination of size and the ratio of a firm’s book value of equity to its market value of equity absorbs the role of $E/P$ and leverage in expected stock returns from 1963–1990. They conclude that these two empirically determined variables, size and the book-to-market ratio, have a high degree of explanatory power in explaining the cross-section of the average returns on US stocks.

To explain these patterns in average returns, Fama and French (1993) developed a three-factor model using the market return less the risk-free rate (RM-RF), the return of small stocks less big stocks (SMB), and the return of high book-to-market less low book-to-market stocks (HML), as the three factors to describe the behaviour
of the expected stock returns. The model is a regression of the excess returns (return less risk-free rate) of a stock or portfolio of stocks on the three factors. The coefficients estimate the exposure to the sources of undiversifiable risk that the factors proxy for, or simply the extent to which the stock returns are influenced by the three factor portfolios. The intercept (known as the excess return, abnormal return or alpha) is the portion of the return that exceeds what is predicted given the stock’s exposure to the undiversifiable risk factors (often referred to as common or systematic risk factors). This approach of explaining average returns with a time-series model provided an alternative to the then-dominant Fama-Macbeth (1973) two-stage procedure, which is not robust to correlated errors and often produces poor standard error estimates (Shanken and Zhou, 2007).

The three-factor model was extended by Carhart (1997), who augmented the Fama and French (1993) model with an additional factor that considered the effect of momentum in expected stock returns. The model was motivated by the three-factor model’s inability to explain the cross-sectional variation in momentum-sorted portfolio returns, where stocks stocks with above (below) average returns in recent months tend to outperform (underperform) other stocks in subsequent months. The model was motivated by the three-factor model’s inability to explain the cross-sectional variation in momentum-sorted portfolio returns. The momentum factor is based on Jegadeesh and Titman (1993), who find that buying stocks that performed well in the past year and selling stocks that performed poorly in the past year generates significant positive returns (around 1% per month) over three to twelve month holding periods. In Carhart’s (1997) extension, the returns from following a long prior year ‘winners’ and short prior year ‘losers’ strategy are constructed in a similar way the SMB and HML factors, and this is added as the fourth explanatory variable (MMC) in Carhart’s (1997) model. The four-factor model is applied to the returns of Australian equity funds in Chapter 3.

As with the size effect and the value premium, there exists some controversy as
to whether the returns from following a momentum strategy are rewards for risk
or are the result of mispricing. In the former category, Carhart (1997) attributes
momentum to the expense ratio and transaction costs. Moskowitz and Grinblatt
(1999) find that most apparent gains come from short positions in small, illiquid
stocks, which also have high transaction costs, and that a large proportion of these
gains come from taking short positions in anticipation of tax-loss selling. As with the
size effect, the efficient markets interpretation suggests that a profitable momentum
strategy is due to a microstructure glitch, rather than investor irrationality. In
the latter category, Barberis, Shleifer, and Vishny (1998) argue that momentum
effects are caused by behavioural biases due to investors underreacting to new
information.

The factor model approach of empirical asset pricing can also be viewed in the
context of arbitrage pricing theory (APT), developed by Ross (1976), who theorised
that since there are several sources of unsystematic risk, the expected return of a
stock can be modeled as a linear function of a number of factors, which represent
theoretical market indices or macro-economic variables. The rate of return derived
by an APT model is the equilibrium rate of return where prices are in alignment
to the extent that arbitrage opportunities have been eliminated. The model
dropped many assumptions which were implicit in the CAPM (for example that
only the mean and variance of returns are relevant for portfolio selection), but
nevertheless imposed a linear factor structure and assumed that idiosyncratic shocks
are uncorrelated both across assets and with each factor.

Numerous studies have used the Fama-French (1993) three-factor or the Carhart
(1997) four-factor model not only to describe the expected returns of an individual
stock, but as a performance attribution model. Results indicate that there is some
modest persistence in mutual fund returns, but only if excess fund returns are
adjusted to account for style biases (i.e. tilted towards or away from small-cap,
value or momentum stocks) using one of the aforementioned models.
The majority of empirical studies have found that actively managed mutual funds tend to underperform passive ones. Given Sharpe’s (1964) equilibrium accounting identity, this is perhaps not surprising—since index (passively managed) funds do not trade much, any transfers of wealth between active and passive funds will be small. So the gain of an actively managed fund will be at the loss of another active manager (broadly defined) since for every buyer of a positive-alpha stock there is of course a seller. So if a sample of active managers generate an aggregate positive abnormal return, this is at the expense of active investments elsewhere.

Kothari and Warner (2001) applies the Carhart (1997) model to estimate the abnormal returns of US equity mutual funds and find that the average fund has an alpha of 0.05, an alpha $t$-statistic of 0.16, and that the standard deviation of the cross-section of alphas is 0.31. They conclude that the question of whether significant excess performance of individual can be identified ex-ante “remains unanswered” due to the low power of the tests they perform. Bollen and Busse (2005) also use the Carhart model to evaluate the performance of US funds. They find that the cross-section of alphas is roughly symmetrical and centered around zero, with the abnormal returns of the top funds (0.07% p.m. in aggregate for the top decile of funds) largely being balanced by that of the bottom funds ($-0.08\%$ p.m. for the bottom funds).

Kosowski, Timmermann, Wermers, and White (2006) used Fama and French’s (1993) three-factor model and Carhart’s (1997) four-factor model to estimate the abnormal returns of US mutual funds, however they use a non-parametric bootstrap estimation approach, arguing that the parametric assumptions used in prior studies are not realistic. They argue that a parametric approach gives a (negatively) biased estimate of the number of funds that have the ability to generate excess returns, and found that using their simulation algorithm; approximately 10 per cent of US mutual funds have significant stock-selecting ability. The algorithm that I apply in Chapter 4 is similar to that of Kosowski et al. (2006); however the analysis of
results is quite different. Prior to this paper, most studies that used a factor-model approach to evaluate fund performance were done in the style of [Jensen (1968)], where the statistical significance of the intercept (or ‘alpha’) was used to determine whether a fund showed evidence of stock-selecting ability.

The simulation approach of [Kosowski et al. (2006)] uses non-parametric bootstrap simulations to generate a distribution of excess returns (alphas) for each mutual fund, assuming that the excess returns generated by the mutual fund in question were solely due to luck. They compare the estimated alpha from running Carhart’s (1997) four-factor model with this distribution to determine whether the excess return of a given fund was due to luck or skill. [Cuthbertson, Nitzsche, and O’Sullivan (2008)] use the same algorithm as [Kosowski et al. (2006)] to evaluate the performance of UK mutual funds. They find that between 3–8 per cent of UK funds show evidence of skill, compared with the roughly 10 per cent of US funds identified by [Kosowski et al. (2006)]. [Fama and French (2010)] use a different bootstrap algorithm to evaluate the performance of US funds. However their approach implicitly assumes that there is no autocorrelation in the returns of individual mutual funds, which is not required for the approach of [Kosowski et al. (2006)] (or the approach used in this study). However much of the finance literature (e.g. Fama, 1965) suggest that autocorrelation in stock returns will only be a minor problem.

### 2.3 Australian literature

Entire articles have been devoted to constructing the SMB, HML and MMC factors in Australia (not necessarily for fund performance evaluation), but the results have been contradictory. Past studies have shown that the SMB factor has a large degree of explanatory power for the returns of Australian stocks, but the average SMB value (that is, the average premium for small stocks) has found to be both positive and negative (in roughly equal proportions) even when studies use a similar time period. [Gharghori, Chan, and Faff (2006); Gharghori, Mudumba, and Veeraraghavan (2007)]
found that the HML factor (the value premium) is insignificant for explaining the returns on Australian stocks, but Brailsford et al. (2012) found that the HML factor tends to be significant. Humphrey and O’Brien (2010) identify a positive SMB, and a positive but statistically insignificant HML. Bowers and Heaton (2013) construct all four factors over the 2002-2010 and find that SMB, HML and MMC only marginally add to the explanatory power of a single-factor model that uses a value-weighted index.

It is very likely that these contradictory results are due to short time periods, limited data availability or running the model without a momentum factor. Humphrey and O’Brien (2010) and Brailsford et al. (2012) manually collected the book values of Australian firms from their annual reports and used the AGSM-CRIF database to for market capitalisation values. Halliwell, Heaney, and Sawicki (1999) who constructed the three factors for the 1980’s only used 26 per cent of listed companies. Gaunt (2004) and Gharghori et al. (2007), who constructed the factors over the 1990’s, had access to less than half of the total number of listed companies. Given these contradictory results, the four-factors have been constructed using Thomson Reuters’ Datastream database (see Chapter 3), which contains book values and market capitalisations for all relevant Australian companies, over the 1986–2013 period.\footnote{Many micro-cap stocks were missing from the database, but these will not have a significant impact on how the relevant portfolios are constructed (see Section 3.1). Specifically, the weights of these micro-cap stock in the ‘small-cap’ portfolio will not be large enough to substantially alter the returns of the portfolio of small-cap stocks. Moreover, the majority of funds used in this study only invest in ASX300 stocks, so the elements that influence the returns of stocks that are outside the ASX300 portfolio will be redundant in the evaluation of most of the funds that I will be evaluating.}

There has been little research on Australian mutual fund performance that has controlled for the common factors in stock returns as defined by Carhart (1997). The most prominent example is Humphrey and O’Brien (2010), who used the model to evaluate the performance of unit trusts and superannuation funds from 1992–2010.
finding no evidence of outperformance. Bilson et al. (2005) applied both the Carhart and Fama and French (1993) models to evaluate the performance of retail funds, finding that 8.2 per cent of funds have significant alphas (at the 5 per cent level), and that the alphas of most funds are marginally above zero. However there have been no Australian studies that have dropped the assumption of an ex-ante distribution from which excess returns are assumed to be drawn, or that have made an explicit adjustment for luck. This study attempts to fill this gap.
3.1 The four factors

The motivation for the use of a factor-model approach to evaluate the performance of equity mutual funds is twofold. Firstly, it controls for the common risk factors in the returns of stocks that cannot be diversified away. Whilst a fund manager can increase the funds expected return by increasing exposure to one or more risk factors, this will be accompanied by an increase in undiversifiable risk, which is the only relevant risk to a diversified portfolio. Hence this increase in expected return is not a reflection of the manager’s skill in identifying mispriced stocks, but merely a reward for taking on added undiversifiable (i.e. systematic) risk. (A central tenet of financial economics is that only the systematic portion of risk is relevant in a diversified portfolio). Secondly, the model may be used as a performance attribution model that estimates the portion of a mutual fund’s return that is not explained by exposure to undiversifiable risk factors, which in turn can be used to estimate the value-added of the fund in question. Carhart (1997) uses it in this way to evaluate the performance of US equity mutual funds.

The Carhart (1997) four-factor model builds on the three-factor model of Fama and French (1993), who were the first to identify non-market factors that proxy for the undiversifiable risk in the returns of stocks by extending the single-factor CAPM. Whilst the CAPM includes only a market factor (in empirical work, generally proxied by the value-weighted return of all stocks listed in a country’s major stock indices), Fama and French include a further two factors that further help to explain the
behaviour of average returns. These factors take into account the role that the size premium (the higher expected return of small stocks) and the value premium (the higher expected return of value stocks relative to growth stocks) have in expected stock returns, and Fama and French (1996) argues that these two factors proxy for non-market undiversifiable risk factors. Carhart (1997) augments the Fama and French model by including an additional factor that considers the higher expected return of momentum stocks (stocks which have performed well during the previous year) relative to contrarian stocks (those which performed poorly).

Whilst there is some minor controversy as to whether the size, value and momentum effects do in fact proxy for undiversifiable risk factors (discussed in Section 2.2), the view that these effects are caused by mispricing rather than reward for risk will not invalidate the results of this study. Under this view, the factors can simply be interpreted as diversified passive returns that help to explain average returns, irrespective of what the source of the average returns are.

3.1.1 Constructing the four factor portfolios for the Australian market

To explain how the size, value, and momentum-stock premiums on Australian stocks change over time, the monthly returns on the portfolios of small stocks less big stocks, value stocks less growth stocks, and momentum stocks less contrarian stocks are constructed mostly in the style of Fama and French (1993) and Carhart (1997).

To run the Carhart (1997) four-factor model on a specific mutual fund, the returns of the mutual fund are regressed against each of these three portfolios and a market portfolio. The coefficients on each of the four dependent variables estimate the fund’s exposure to each of the four factor portfolios, and known as the loadings on each factor.

Since I am evaluating the performance of Australian equity funds that invest in Australian stocks, I use the stocks listed on the Australian Securities Exchange
(ASX) to generate the monthly returns of the four portfolios, which covers more than 96 per cent of the Australian equity market. The portfolio returns are computed for the period from July 1986 to April 2013 (henceforth 1986–2013), and the data is provided by Thomson Reuters’ Datastream Professional database.

The variables which are used to construct these four portfolios are the monthly time series of returns, the book values, and the market values (i.e. size or market capitalisations) of every stock listed on the ASX at some point during the 1986–2013 sample period. For an individual stock, the return index that I used shows the theoretical growth in the value of a share holding a share assuming that dividends are re-invested and are used to purchase additional units of the stock at the closing price applicable on the ex-dividend date. Market value of equity (ME) is equal to the share price multiplied by the number of ordinary shares issued. For companies with multiple classes of equity capital, the value is expressed according to the individual issue, and the amount issued is updated following a capital change or whenever new tranches of stock are issued. And finally, a firm’s book value of equity (BE) is computed using the balance sheet value of the common (ordinary) equity in the company. It also includes (but is not limited to) retained earnings, the capital surplus and the capital stock premium.

Firms that don’t issue common equity (including foreign firms that are listed, real estate investment trusts or units of beneficial interest) were dropped from the sample. To adjust for survivorship bias, I did not include firms until they had appeared in the database for at least two years. If a firm had a negative BE in a certain month, the firm’s BE, ME and return in that month were excluded. An adjustment for dead equities was also required since Datastream continues to quote the most recent return index for a company after it gets delisted. I also dropped the smallest ‘micro-cap’ firms from the sample using the following rolling cut-off: Before 1995, if a firm’s

---

1The Datastream variables used to for the returns, book values and market values of each firm are the Total Return Index (RI), Market Value (MV) and Book Value (WC03501). [Schmidt, Schrimpf, Wagner, and Ziegler (2011)] provides a justification for using these specific variables when using Datastream to construct the Carhart factors. Further, the RI index is analogous to how the returns of the mutual funds are calculated in Section 3.2.
ME was less than $45 million in a given month, its return, ME and BE observations during that month were dropped, which rose to $65 million for the 1995–2003 period and $80 million from 2004. Results were robust to the cut-off points used. Finally, a firm’s book-to-market ratio was created by dividing its book value of equity for the fiscal year ending in calendar year \( t - 1 \) by its market value of equity at the end of calendar year \( t - 1 \).

In order to classify each stock as big or small (based on market capitalisation) or value or growth (based on their book-to-market ratio) in a given month, I generated a series of median firm sizes (i.e. the market capitalisation of the median listed company in a given month), and the 30th and 70th percentile values of the book-to-market ratio for each month. Firms with a market cap above (below) the median in a given month were classified as big (small) for that month. Firms with a book-to-market ratio below the 30th percentile in a given month were classified as growth or low-BE/ME stocks, and firms above the 70th percentile were classified as value or high-BE/ME stocks. Firms in the middle 40 per cent were classified as core or medium-BE/ME stocks.

This results in six ways to classify a stock in a given month based on its size and BE/ME characteristics (two size categories times three BE/ME categories), and six portfolios are constructed containing the stocks in each of the six groups. I will refer to these portfolios as \( SL, SM, SH, BL, BM \) and \( BH \). The \( SL \) portfolio, for example, is the portfolio of monthly returns (weighted by size) that contains stocks that are both small \( (S) \) and have low book-to-market ratios \( (L) \). The weight of an individual component stock of the \( SL \) portfolio in month \( t \) is equal to its market capitalisation in month \( t \) divided by the total market capitalisation of the \( SL \) portfolio in month \( t \). The \( BH \) portfolio for example, is the monthly returns of stocks that are in both the big (market cap) group and the high (book-to-market ratio) group.

For each of the six portfolios, the series of size-weighted returns is generated from
July of year \( t \) to June of year \( t+1 \). At the end of year \( t+1 \) the portfolios are reformed, and the size-weighted returns for the subsequent year are computed. That is, stocks are placed into one of the six portfolios based on its size and BE/ME values in the previous June, rather than being reformed monthly.

Finally, the monthly returns of these six portfolios are used to generate the returns on the \( SMB \) (small minus big) and \( HML \) (high minus low) portfolios—these are two of the four factors in the Carhart four-factor model. The \( SMB \) and HML monthly returns are computed as follows:

\[
SMB_t = \frac{SL_t + SM_t + SH_t}{3} - \frac{BL_t + BM_t + BH_t}{3}
\]  \hspace{1cm} (3.1)

\[
HML_t = \frac{SH_t + BH_t}{2} - \frac{SL_t + BL_t}{2}
\]  \hspace{1cm} (3.2)

The \( SMB_t \) portfolio is equal to the simple average of the three small portfolios minus the simple average of the three large portfolios. The idea is to disentangle the book-to-market effects from size to give an indication of how the small-cap premium changes over time. The idea is the same for the \( HML_t \) portfolio, which is constructed by subtracting the average of the two high \( BE/ME \) portfolios by the two low portfolios. The movement of the \( SMB_t \) (\( HML_t \)) portfolio gives an indication of how the small-cap premium (value premium) changes each month.

The market factor in the Carhart model is equal to the monthly return of the size-weighted portfolio of all of the stocks in the six size-\( BE/ME \) portfolios, plus the stocks with negative book market values which were previously excluded. From this, I subtract the risk-free rate proxy, the 90-day bank accepted bill rate, to form the excess market return \( (RM - RF)_t \) which is proxies for the market factor in the Carhart model. The factor proxies for (undiversifiable) market risk, and is the premium on the expected return of Australian stocks relative to the risk-free rate, which is known as the equity risk premium.
The fourth factor is constructed in a similar way to the HML portfolio and considers the effect of momentum in stock returns, namely that stocks which have performed well during the past year to continue to do well over the next month (these are known as momentum stocks), and stocks that have performed poorly continue their poor performance during the following month (known as contrarian stocks). Whilst some Australian studies have argued that momentum is the strongest using a six month strategy rather than an annual strategy (Brailsford and O’Brien 2008; Demir, Muthuswamy, and Walter 2004; Durand, Limkriangkrai, and Smith 2006), I find in unreported tests that the annual strategy has far more explanatory power for the Australian market compared to a six-month strategy. This is consistent with the approach of Carhart (1997) and the vast majority of international literature, and Australian studies such as Hurn and Pavlov (2003).

In order to classify each stock as momentum or contrarian; I first generate a series of one-month lagged eleven month returns for each stock (henceforth eleven-month returns). That is, for each stock I generate a monthly series of \( R_t = \frac{(P_{t-1} - P_{t-12})}{P_{t-12}} \), where \( R_t \) is the eleven-month return and \( P_t \) is the stock price at the end of month \( t \). I then then generate the monthly series of the 30th and 70th percentiles of the eleven month returns of each stock for each month. Firms with an eleven month return above the 70th percentile in a given month are classified as momentum stocks, and firms below the 30th percentile are classified as contrarian stocks.

As with the HML factor, I categorise each stock into one of four categories to disentangle the effects of size from momentum: small momentum stocks (SMom), big momentum stocks (BMom), small contrarian stocks (SC) and large contrarian stocks (LC) each month. As with the market factor, these four portfolios are reformed monthly. These four portfolios are used to generate the fourth factor.
in the Carhart model, which I call $MMC_t$ (momentum minus contrarian). 

$$MMC_t = \frac{SMom_t + BMom_t}{2} - \frac{SC_t + BC_t}{2}$$  \hspace{1cm} (3.3)

The $MMC$ factor considers the premium on momentum stocks relative to contrarian stocks. Whilst the effect of size on momentum was not considered in the original Carhart (1997) study, the Australian literature indicates that size and momentum are related (Brailsford and O’Brien, 2008) so using the two independent sorts should neutralise the size effect within momentum.

### 3.1.2 The Australian factor portfolios

To show how the equity risk premium (RM-RF), and the premiums on small stocks (SMB), value stocks (HML) and momentum stocks (MMC) change over time, I graph the monthly series of returns for each of the four factors. I also smooth each series using a centred 11-month moving average filter (shown in the right panel) so any trends in the behaviour of the premiums can be seen more clearly:

Moreover, the returns on the factors are highly non-normal. The histograms of the returns of each factor are shown in Figure 3.2 and are overlaid with a normal distribution with the same mean and variance. The non-normality has implications on how the performance of the mutual funds are evaluated (discussed in Section 4.1):

Further, the correlations between each portfolio are quite small. I show this in Table 3.1 and the scatterplot matrix in Figure 3.3.

### Table 3.1: Cross-correlation matrix of the four factors

<table>
<thead>
<tr>
<th></th>
<th>RM-RF</th>
<th>SMB</th>
<th>HML</th>
<th>MMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM-RF</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMB</td>
<td>-0.075</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>-0.073</td>
<td>-0.217</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>MMC</td>
<td>-0.233</td>
<td>-0.455</td>
<td>0.147</td>
<td>1.000</td>
</tr>
</tbody>
</table>
I also compare results with the corresponding US factor portfolios. Table 3.2 shows the summary statistics of the previously generated factor portfolios during the 1986 to 2013 period, and the results are compared to the US portfolios over the same sample period. The average returns of the equity premium ($R_M - R_f$) and the small-cap premium (SMB) are similar in both countries, and are both insignificant from zero. The finding of an insignificant SMB premium has robust across all major international markets since the mid-1980’s (see for example Fama and French [2012]). However the premium on value stocks in Australia is twice as large, and the premium on momentum stocks is almost three times as large as the US. The large value and momentum effects in Australian stock returns are relatively well documented in the Australian literature. See, inter alia, Halliwell et al. [1999], Gaunt [2004] and Durand et al. [2006] for discussions on the value premium; and Hurn and Pavlov [2003] and

\footnote{US Data are taken from Kenneth French’s website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed 19 September 2013.}
Demir et al. (2004) who analyse the momentum premium.

Table 3.2: Australian vs. US factor portfolios, 1986–2013

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_M - R_f$</td>
<td>$SMB$</td>
<td>$HML$</td>
</tr>
<tr>
<td>Average Return</td>
<td>0.41</td>
<td>-0.17</td>
<td>0.56</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4.59</td>
<td>3.15</td>
<td>2.77</td>
</tr>
<tr>
<td>$t$-statistic</td>
<td>1.58</td>
<td>-0.95</td>
<td>3.63</td>
</tr>
</tbody>
</table>

In terms of other Australian studies, numerous papers have been primarily or exclusively focused on constructing the factor portfolios and analysing the equity, small-cap, value and/or momentum premiums in Australia. So-far, the findings have been contradictory. Two recent examples are Brailsford et al. (2012) and Gharghori et al. (2006), who find wildly different small-cap premiums. Another recent paper is Humphrey and O’Brien (2010) who construct the factors in order to analyse the
performance of superannuation funds and unit trusts.

Table 3.3 shows the summary statistics for these three papers. The average return, standard deviation of returns and the $t$-statistics for the portfolios are given in the first three rows over each estimation period, which is stated in row 5. Note that some of the statistics were not reported (NR) and the MMC portfolio was only generated in Humphrey and O’Brien (2010) since the other two papers used the Fama and French (1993) model. Also note the differences in sample periods—in order to compare my results with that of these three papers, I estimate the average returns of the factor portfolios (generated in the previous section) over the same estimation period of each of the three papers. The average returns estimated over each subset of dates are given in row 4.

Comparing Tables 3.2 and 3.3 it is clear that the results most closely accord with Brailsford et al. (2012), where the average monthly returns on the RM-RF, SMB and HML portfolios only differ by a couple of basis points, and the unreported
Table 3.3: Australian vs. US factor portfolios, 1986–2013

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave. Return</td>
<td>0.51</td>
<td>0.15</td>
<td>0.23</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>4.59</td>
<td>3.10</td>
<td>NR</td>
</tr>
<tr>
<td>t-statistic</td>
<td>1.92</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Ave. return over common sample period</td>
<td>0.54</td>
<td>0.54</td>
<td>NR</td>
</tr>
<tr>
<td>Data source</td>
<td>Hand collected from AGSM-Crif exc. book values (hans AGSM-Crif and Aspect annual reports)</td>
<td>AGSM-Crif and Aspect collected from annual reports</td>
<td>Financial databases</td>
</tr>
</tbody>
</table>

aSince Humphrey and O’Brien (2010) only reported the average market return, RM, I subtracted the average 90-day bank accepted bill rate over their estimation from $R_M$ to generate the $R_M - R_F$ factor for this study.

bThe returns were estimated over the July 1986 to Dec 2006 period for the Brailsford et al. comparison rather than the full Jan 1982 to Dec 2006 sample.

standard deviations and t-statistics are very similar. The results also suggest that Thomson Reuters Datastream provides a reasonable alternative to how Brailsford et al. obtained their data to create the factors, namely by manually searching through the annual reports of every company listed on the ASX (for post-1997 data) or by using the now defunct Company Reporting Service (CRS) for the 1981–97 period.

The results of Gharghori et al. (2006) are the most divergent, and it seems unlikely that the findings of this paper can be justified given past studies and economic fundamentals, most notably that the SMB portfolio generates an average annualised return of 54 per cent for Australian stocks. The result is particularly strange in light of the discoveries of disappearing size effects in all major international markets including Australia from the mid-1980’s.

3.2 Mutual Fund Data

For a mutual fund to be included in this study, it must be managed in Australia, invest only in equities (of which at least 80 per cent must be issued by Australian companies), be actively managed, be open-ended (i.e. shares can be invested
and redeemed at any time), have at least 36 months of continuous returns data between July 1986 and April 2013 and not be a duplicate fund. This leaves 329 funds with a total of 42,988 fund-month observations—one of the largest panels of fund returns every analysed for the Australian market—with around $250 Billion in assets under management. All mutual fund data is from Morningstar Direct, and with the exception of identifying passively managed and duplicate funds, this database was used to determine the sample of funds which satisfy the aforementioned requirements.

Morningstar does not include any data on whether its mutual funds are actively or passively managed. To determine which funds were passively managed (i.e. those who mimic the performance of an index rather than actively picking stocks), I first performed a separate internet search for all of the mutual funds in the sample to determine whether a given fund was actively managed. I then dropped all of the passively managed funds. As an additional measure, for each of the remaining funds I then regressed total returns on the S&P/ASX 20, 50, 100, 200, 300, Small Ordinaries, MidCap 50 and High Dividends indices, as well as the All Ordinaries index and the SMB, HML and MMC portfolios (both individually and in combinations). Funds that returned an R-Squared larger than 0.98 in any of these regressions were deemed to be passively managed (or not active enough to be included in the sample) and were dropped. Data for the indices were provided by McGraw Hill Financial, Yahoo Finance and the Center for Research in Security Prices (CRSP).

To determine which funds were duplicates, I first identified funds with very similar returns. For example, for firms that offer both a wholesale and retail product, the total returns of the two funds will differ by only a couple of basis points so I dropped the fund that had the smallest number of observations. Then, I sorted each fund by its firm name (examples include AMP Group which manages 12 mutual funds in the sample, or Macquarie Bank Group which manages 3). Then for each firm, I

\[ \text{3I often relied} \quad \text{http://www.au.investsmart.com.au or http://www.2020directinvest.com.au for this information. Otherwise, I would refer to the company website or other sources.}\]
generated the correlation matrix of total returns of all of the funds that managed by that firm. If any two funds had a correlation coefficient exceeding 0.99, I dropped the fund with the smallest number of observations.

(Report how many were lost at each step here e.g. dropping duplicates, no data funds, passive and survival bias fell from 746 to 355—all in Stata code).

This left a total of 329 funds that have been in operation at some point during the sample period. The growth in Australian mutual funds (which satisfy the requirements of this study) throughout the 1986-2013 sample period can be viewed in the graph below:

**Figure 3.4: Growth in Australian equity mutual funds**

---

3.3 **Standard estimation procedure**

To estimate the exposure that each fund has to each of the four risk factors I use the Carhart (1997) model, which is a regression of the excess monthly returns of a fund.
on the four factors. The excess return of a fund in month $t$ is equal to the fund’s total return during that month ($R_{it}$) minus the 90-day bank accepted bill rate in month $t$ ($R_{ft}$). Total return is computed in each month by taking the percentage change in monthly net asset value, reinvesting all income and capital gains distributions during that month. Reinvestments are made using the actual reinvestment NAV and daily payoffs are reinvested monthly. The returns are not adjusted for sales charges such as deferred or front-end loads or redemption fees to give a clearer indication of a fund’s performance. The returns do account for management and administrative fees and other costs taken out of fund assets. The annualised bank accepted bill rate is provided by the RBA.

The estimated equation for fund $i$ is as follows:

$$R_{it} - R_{ft} = \alpha_i + \beta_i (R_{Mt} - R_{ft}) + \gamma_i SMB_t + \delta_i HML_t + \varphi_i MMC_t + \epsilon_{it} \quad (3.4)$$

The constant in the regression equation estimates the portion of a fund’s return that is not due to exposure to its undiversifiable risk as proxied by the four factors, which provides a risk-adjusted measure of fund performance. This is known as a fund’s abnormal return or alpha.

The coefficients estimate the fund’s loading on (or exposure to) the four factors. If the beta coefficient (the coefficient on the market less risk-free rate proxy) is greater (less) than one, the fund in question is more (less) sensitive to market volatility than the market portfolio. A positive coefficient on SMB, HML or MMC suggests that the fund has more exposure to small-cap, value or momentum stocks than the market portfolio (RM). As an example, suppose a fund with neutral exposure to small-cap, value and momentum stocks (i.e. with 0 coefficients on the SMB, HML and MMC factors) decides to generate a higher average return by taking on extra unsystematic risk. If for example the fund manager overweights on value stocks, the fund will have a positive loading on the HML factor which will therefore generate a higher expected return in future periods as long as the average return of the HML
portfolio is positive in the future. One would expect this to be the case for the HML and MMC factors which are consistently positive, but perhaps not for the SMB factor due to the disappearing size effect (refer to Figure 3.1). The residuals for fund i estimate its exposure to idiosyncratic risk—the risk that is specific to the individual mutual fund (more on this later).

The average loadings for each of the mutual funds, as well as the proportion of funds whose loading is statistically significant from zero at the 5 and 10 per cent levels are shown below in Table 3.4. The coefficients are estimated using the ordinary least squares method.

Table 3.4: Loadings on the four factors

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}_i$</th>
<th>$\hat{\gamma}_i$</th>
<th>$\hat{\delta}_i$</th>
<th>$\hat{\phi}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.81</td>
<td>0.39</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>% significant at 5 per cent level</td>
<td>100</td>
<td>94</td>
<td>25</td>
<td>85</td>
</tr>
<tr>
<td>% significant at 10 per cent level</td>
<td>100</td>
<td>90</td>
<td>13</td>
<td>80</td>
</tr>
</tbody>
</table>

It is evident from the last two rows that for the majority of funds, the exposure to small-cap and momentum stocks differs from that of the market portfolio. However only 13 per cent of funds underweight or overweight on value stocks relative to the market portfolio when testing at the 10 per cent level. The finding that most funds have an insignificant loading on the HML factor is consistent with past UK studies (for example Cuthbertson et al., 2008 who also found the MMC factor tends to be insignificant for UK mutual funds), however past studies on US mutual funds (e.g. Carhart, 1997; Kosowski et al., 2006) find that most US funds have a significant loading on all of the factors.

A summary of the diagnostics, factor returns and the results of a regression on the equally weighted portfolio of all of the mutual funds included in this study are shown in Table 3.5.

From Table 3.5 it is clear that the average fund does not produce a significant abnormal return since the average estimated alpha is 0.01 per cent per month,
and the average $t$-statistic of the estimated alpha is 0.01. The abnormal return of the equally weighted portfolio is also insignificant (0.08% p.m., with a $t$-statistic of 0.89). The results also indicate that most funds overweight on small-cap and momentum stocks relative to the market. The diagnostics show that a substantial proportion of the funds have non-normal residuals, especially amongst the funds with the highest estimated alphas, which suggests that the approximations made by an OLS (or otherwise parametric) approach may lead to erroneous inferences. These are precisely the funds that are of most interest to investors, and whose performance is most attributed to skill. This point is expanded on in Chapter 4.

However, to evaluate fund performance, I primarily use the $t$-statistic on the estimated alpha. Whilst the estimated alpha measures the economic size of abnormal performance, it is likely to suffer from a lack of precision. The $t$-statistic however has better sampling properties and is a pivotal statistic (that is, its probability distribution is not a function of unknown parameters, such as the error variance). Moreover, if a fund has a relatively short life or engages in a relatively high level in risk taking, the standard error on the alpha is likely to be high and the point estimate of its alpha will be a spurious outlier in the cross-section distribution of estimated alphas. Using the $t$-statistic corrects for these outliers by normalising the estimated alphas by its estimated standard deviation. The $t$-statistic is related to the Treynor and Black (1973) appraisal ratio, which Brown et al (1992) suggests using for mitigating survival bias problems. The cross-section of $t$-statistics also has better properties than the cross-section of estimated alphas in the presence of heterogeneous fund volatilities due to differing risk levels or differing life spans.

In Figure 3.4 I show the cross-section of the OLS estimated alphas (left panel) and the OLS t-stats on the estimated alphas (right panel). Each histogram is overlaid with a scaled normal density function with the same mean and standard deviation as the data.

With the exception of a couple of funds in the far right tails, funds with positive
abnormal returns seem to be balanced by funds with negative abnormal returns, and
the abnormal returns are clustered around zero, supporting the results in Table 3.5.
Specifically, the top three funds have estimated alphas of 1.74, 1.39 and 1.10 per
cent per month; or equivalently, annualised abnormal returns of 23.0, 18.0 and 14.2
per cent. Further, both of the cross-section distributions are very non-normal. The
tails are far fatter than a normal distribution with the same mean and standard
deviation, and both distributions are somewhat positively skewed.

A statistically significant and positive alpha suggests that the returns of the fund
in question are significantly higher than what can be explained by exposure to the
four risk factors alone, and provides evidence of skill in selecting mispriced stocks.
Conversely, funds with a significant negative alpha provide evidence of ‘negative’
skill, and funds with an insignificant alpha do not have any evidence that they can
provide returns higher than the four passive benchmarks. The proportions of funds
that satisfy these characteristics are given in the Table 3.6 below:

The proportion of funds that have evidence of skill (both positive and negative)
is very close to the type I error rate, suggesting that there is little skill to
be found in the Australian equity mutual fund market. Testing for statistical
significance of estimated alphas is the technique that all past Australian papers (e.g. Humphrey and O’Brian, 2010) and almost all international studies which employed a factor-model approach have used to evaluate the performance of actively managed funds (not necessarily mutual funds). However this parametric approach relies on highly simplified and often unrealistic assumptions about the nature of the ex ante distribution from which fund returns are assumed to be drawn. I drop these assumptions in the next section by applying bootstrap simulations to the monthly returns of each mutual fund, and compare the results to the parametric approach.
Table 3.5: Diagnostics of the four-factor models

Diagnostics and output from equally weighted portfolio using Carhart’s four-factor model.

Average coefficient (average t-statistic in parentheses)

<table>
<thead>
<tr>
<th>Component</th>
<th>Average Coefficient (%)</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha (constant)</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>RM-RF</td>
<td>0.81</td>
<td>14.98</td>
</tr>
<tr>
<td>SMB</td>
<td>0.39</td>
<td>4.36</td>
</tr>
<tr>
<td>HML</td>
<td>-0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>MMC</td>
<td>0.12</td>
<td>2.92</td>
</tr>
</tbody>
</table>

Diagnostics (all funds)

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Adjusted R-squared</td>
<td>0.72</td>
</tr>
<tr>
<td>Average SIC</td>
<td>601.7</td>
</tr>
<tr>
<td>% funds with non-normal residuals</td>
<td>32</td>
</tr>
<tr>
<td>% of funds with first order autocorrelation</td>
<td>54</td>
</tr>
<tr>
<td>% of funds with first-sixth order autocorrelation</td>
<td>48</td>
</tr>
</tbody>
</table>

Diagnostics (top 20 funds)

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Adjusted R-squared</td>
<td>0.62</td>
</tr>
<tr>
<td>Average SIC</td>
<td>520.1</td>
</tr>
<tr>
<td>% funds with non-normal residuals</td>
<td>50</td>
</tr>
<tr>
<td>% of funds with first order autocorrelation</td>
<td>44</td>
</tr>
<tr>
<td>% of funds with first-sixth order autocorrelation</td>
<td>42</td>
</tr>
</tbody>
</table>

Coefficients on EW portfolio (t-stats in parenthesis)

<table>
<thead>
<tr>
<th>Component</th>
<th>Average Coefficient (%)</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha (constant)</td>
<td>0.08</td>
<td>0.89</td>
</tr>
<tr>
<td>RM-RF</td>
<td>0.84</td>
<td>39.59</td>
</tr>
<tr>
<td>SMB</td>
<td>0.36</td>
<td>10.96</td>
</tr>
<tr>
<td>HML</td>
<td>-0.14</td>
<td>0.42</td>
</tr>
<tr>
<td>MMC</td>
<td>0.06</td>
<td>4.39</td>
</tr>
</tbody>
</table>

Table 3.5 shows results from the estimation of the Carhart four-factor model. Newey and West (1987) heteroskedasticity and autocorrelation adjusted standard errors are used to generate the t-statistics. The t-statistics used for the average coefficients are the cross-sectional average of the absolute value of the t-statistics of each regression. I use the D’Agostino, Belanger, and D’Agostino (1990) skewness-kurtosis test to test for normality in residuals with the adjustment suggested by Royston (1991). The Breusch-Godfrey serial correlation LM test is used for the testing for first order autocorrelation and the joint test of the first six lags. The diagnostics for the top 20 funds include the funds with the 20 highest estimated OLS alphas.

Table 3.6: Evidence of skill at the 5 and 10 per cent levels

<table>
<thead>
<tr>
<th>Skill Type</th>
<th>5 per cent level</th>
<th>10 per cent level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>4.24</td>
<td>7.56</td>
</tr>
<tr>
<td>Negative</td>
<td>0.30</td>
<td>1.82</td>
</tr>
<tr>
<td>No skill</td>
<td>95.4</td>
<td>90.6</td>
</tr>
</tbody>
</table>

39
CHAPTER 4

Accounting for Luck

4.1 Motivation for the bootstrap

The result in the previous section that the cross sectional distribution of estimated alphas is highly non-normal motivates the following question: What proportion of funds are expected to generate large alpha (say, above 5 per cent a year) by luck alone, and what proportion of funds actually generate such an alpha above this level in the Australian market?

To answer this question (amongst others), I apply a semi-parametric cross-sectional bootstrap algorithm to each of the individual funds to explicitly separate skill from luck in abnormal returns in the presence of the highly non-normal distributions of idiosyncratic risk associated with many of the funds. The procedure is used to estimate the cross-sectional abnormal returns distribution that would be generated by luck alone for each of the funds, including those in the tails of the estimated alpha distribution. Typically, these are the funds which are of most interest to investors and the financial media, and the inferences about performance persistence implied by this process can substantially differ from that of a parametric approach.

In Section 4.2, the bootstrap simulations involve resampling the residuals of each fund. Finally, the results are generalised to estimate of the value added for the entire equity mutual fund market in Australia.

There are numerous reasons why a bootstrap resampling approach is superior to the standard parametric approach used in all previous Australian studies. As shown
Table 3.5, many funds have non-normal residuals, particularly those in the tails of the estimated alpha cross-section, and the rejection is often very large for these funds. This strong finding of non-normal residuals challenges the validity of the past Australian literature—all of which relied on this normality assumption—and the challenge to the standard $t$- and $F$-tests for significance strongly suggests a bootstrap procedure is necessary to determine if significant alphas are due to manager skills (at least in part) or if they are due to luck alone. These non-normalities in individual funds translate to non-normalities in the cross-section distribution of estimated alphas, plotted in Figure 3.5.

In addition, the bootstrap algorithms formally model the cross-sectional nature of the ex post sorts of funds by forming the empirical joint distribution of the cross-section of alphas. This is necessary to capture unknown forms of heteroskedasticity and cross-fund correlations in returns, as well as possible clustering in the levels of the idiosyncratic risk taken amongst the funds. This clustering of high and low risk funds may be explained by the principal-agent problems in mutual funds as discussed by Chevalier and Ellison (1997). Specifically, they find that since the fund company desires to maximise its inflow of investments as opposed to risk-adjusted returns, which the investor desires to maximise, fund managers have an incentive to increase or decrease the risk of a fund based on their year-to-date return due to the shape of the estimated fund-performance relationship. So even if the individual fund alphas were normally distributed, heterogeneity and clustering in risk taking amongst the mutual funds will lead to excess kurtosis in the cross-section of alphas, or even tails that are thicker at some percentile points and thinner in others. Looking at Figure 3.5, this seems to be the case for the Australian data.

In terms of the theoretical justification of the bootstrap, Bickel and Freedman (1984) and Hall (1986) show that the bootstrap can substantially improve on the approximation that the constant term or error is normally distributed. Davison and Hinkley (1997) show that a major advantage of residual resampling is improved
quantile estimation when normal-theory distributions of the estimated parameters are not accurate. Efron and Tibshirani (1994, p. 115) shows that taking bootstrap samples of the residuals give reasonable results in the presence of endogeneity or if the probability structure of the regression model is not specified correctly. Further, Horowitz (1994; 1997) shows that the bootstrap can greatly reduce the difference between the true and nominal probabilities of correctly rejecting a null hypothesis. Kosowski et al. (2006) give two further reasons why the cross-section of mutual fund alphas may be non-normal. Firstly, given the non-normality of the factor returns (Figure 3.2), there may be co-skewness in the returns of the factor portfolios and individual stocks, and the returns of individual stocks may exhibit varying levels of autocorrelation. Secondly, funds may implement dynamic strategies that involve changing their levels of risk-taking in response to changing levels of market or other undiversifiable risk, or in response to the performance of their competitors. They argue that even for a very large mutual fund portfolio, normality is likely to be a poor approximation when conducting statistical inference.

Since ranking funds by their alpha $t$-statistics helps to control for risk-taking across funds (as discussed in Section 3.3), the $t$-statistics rather than the estimated alphas will be the primary measures of mutual fund performance. However this does not solve all problems—the cross-section of OLS-estimated alpha $t$-statistics (Figure 3.5) remains distinctly non-normal due to the influence of higher moments of the alphas of the individual funds. The bottom line is, even when using the alpha $t$-statistics, the bootstrap remains crucial for identifying funds with significant skill primarily due to the clustering in risk levels and non-normalities in individual fund alphas. The result is a cross-sectional distribution that represents a complex mixture of the individual distributions of the sample of mutual funds.
4.2 Residual Resampling

The technique of applying Efron's (1979) bootstrap to resample the residuals of a regression model was introduced by Freedman (1981). Variants of this original technique have been frequently applied in the economics and finance literature for statistical inference that is more robust to the aforementioned problems discussed in Section 4.1. In preparation for the bootstrap procedure, the OLS-estimated alphas, factor loadings, residuals and the alpha \( t \)-statistic of each fund are saved, as estimated in Section 3.3. For each fund \( i \) \((i = 1, \ldots, 329)\), I denote the OLS-estimated coefficients as \( \{\hat{\alpha}_i, \hat{\beta}_i, \hat{\gamma}_i, \hat{\delta}_i, \hat{\varphi}_i\} \), the time series of estimated residuals as \( \{\hat{\varepsilon}_{i,t} : t = i_1, \ldots, i_T\} \) where \( i_1 \) is the date of the first monthly return for fund \( i \) and \( i_T \) is the last, and \( \hat{t}_{\alpha,i} \) is the \( t \)-statistic of the estimated alpha of fund \( i \).

To apply the bootstrap residual resampling algorithm to fund \( i \), I draw 999 random samples with replacement (i.e. bootstrap samples) of resampled residuals which I denote as \( \{\hat{\varepsilon}^b_{i,t_b} : t_b = e^b_{i_1}, \ldots, e^b_{i_T}\} \) where \( b \) is an index denoting each of the bootstrap resamples \((b = 1, \ldots, 999)\), and where the time indices for bootstrap sample \( b \) (i.e. \( e^b_{i_1}, \ldots, e^b_{i_T} \)) are drawn randomly from \([i_1, \ldots, i_T]\) such that the original series of residuals for fund \( i \) are re-ordered. I will refer to these resampled residuals as pseudo-residuals.

I then simulate 999 total excess monthly return series’ for each fund \( i \), imposing the null hypothesis of no true performance \((\alpha = 0)\) on all of the funds. The simulated excess return series’, which I call pseudo-returns, are generated as follows:

\[
\hat{r}^b_{it} = \hat{\beta}_i (R_{Mt} - R_{ft}) + \hat{\gamma}_i SMB_t + \hat{\delta}_i HML_t + \hat{\varphi}_i MMC_t + \hat{\varepsilon}^b_{it} \quad (4.1)
\]

Whilst Equation (4.1) uses the same factor coefficients that were estimated by OLS, the true alpha (and alpha \( t \)-statistic) of the pseudo-returns is set to zero by construction. Note that the pseudo-return \( \hat{r}^b_{it} \) could have been equivalently written
as \((R_{it} - R_{ft})^b\), that is the bth simulated total monthly return less risk-free rate series of fund i. Hence this process will generate \(999 \times 329\) pseudo-return series’ (one pseudo-return series for each of the 999 pseudo-residual series, for each of the 329 funds), with the length of the pseudo-return series for fund i equal to \(i_T - i_1 + 1\) (the average length is 131 observations of returns).

I then run the Carhart four-factor model on each fund 999 times using the pseudo-return series’ rather than the actual excess return series. Note that a positive (negative) estimated alpha will result if a high number of positive (negative) residuals are drawn. Running the regressions using the pseudo-return series’ of each fund will yield a cross-section of 999 bootstrapped alphas that results entirely from sampling variation, or luck alone. If the bootstrapped alphas of fund \(i\) (denote as \(\hat{\alpha}_{i,b}\)) consistently fall below the fund’s OLS-estimated alpha, we can conclude that stock-picking skills had a role to play in generating the fund’s abnormal returns, and its alpha was not entirely the result of randomisation. And in terms of the entire mutual fund market, if the bootstrapped alphas generate fewer extreme positive alphas compared to those actually observed (Figure 3.5 left panel), this provides evidence that skills in stock-selection actually exist in the Australian mutual fund market, and high abnormal returns are not purely the result of luck. The alpha \(t\)-statistics rather than the alpha point-estimates are also used to address these questions, which avoids some of the problems associated with heterogeneous risk-taking discussed in Sections 3.3 and 4.1.

---

1This addresses the question raised in the first sentence of the introduction—are the funds with the highest alphas actually skilled, or merely the luckiest of the 329 funds? In the latter case, the generation of abnormal returns would be akin to that of a coin-flipping contest, where (in a sample of 329), one fund is expected to return eight heads in a row, 10 are expected to return seven from the first eight attempts; but such strong performance of flipping heads of course doesn’t provide any evidence of coin-flipping ability. Generating the ‘luck-alpha’ distribution is akin to forming the distribution of heads flipped—if the distributions are not significantly different (which would be expected in the coin-flipping example), we can conclude that the top performers were merely the luckiest in a large sample, and the lowest performers were the unluckiest, and there is no evidence of skill.
4.3 Analysis of individual funds

I begin with reporting the results of individual funds. Table 4.1 gives the results from both the parametric and residual resampling approaches. In the top panel, for various points along the cross-sectional estimated-alpha distribution, I test the null hypothesis of zero true alpha (that any abnormal returns are solely attributed to luck, i.e. randomisation) against an alternative of a positive true alpha (evidence of skill in selecting stocks) for positive estimated-alpha funds, and a negative true alpha (‘negative’ skill) for negative estimated-alpha funds. I test these hypotheses using both the parametric and bootstrap simulation approaches, and report the p-values of these tests at various percentiles of the estimated alpha cross-section.

In the bottom panel, I perform the same tests, but rank the funds according to the cross-section of estimated t-statistics, and the bootstrapped p-values are computed according to the simulated t-statistics of each fund, rather than the simulated alphas.

Table 4.1: Testing for skill vs. luck at various percentiles in the cross-section

<table>
<thead>
<tr>
<th></th>
<th>max</th>
<th>99%</th>
<th>98%</th>
<th>95%</th>
<th>90%</th>
<th>80%</th>
<th>70%</th>
<th>50%</th>
<th>30%</th>
<th>20%</th>
<th>10%</th>
<th>5%</th>
<th>2%</th>
<th>1%</th>
<th>min</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top Panel: Ranking based on estimated alpha</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated alpha</td>
<td>1.74</td>
<td>1.10</td>
<td>0.79</td>
<td>0.51</td>
<td>0.33</td>
<td>0.16</td>
<td>0.08</td>
<td>−0.03</td>
<td>−0.13</td>
<td>−0.18</td>
<td>−0.26</td>
<td>−0.34</td>
<td>−0.43</td>
<td>−0.50</td>
<td>−0.87</td>
</tr>
<tr>
<td>Annualised est. alpha</td>
<td>23.0</td>
<td>14.1</td>
<td>9.86</td>
<td>6.35</td>
<td>4.03</td>
<td>1.98</td>
<td>0.96</td>
<td>−0.30</td>
<td>−1.50</td>
<td>−2.19</td>
<td>−3.11</td>
<td>−4.04</td>
<td>−5.08</td>
<td>−5.82</td>
<td>−9.91</td>
</tr>
<tr>
<td>Parametric p-value</td>
<td>&lt; 0.01</td>
<td>0.01</td>
<td>&lt; 0.01</td>
<td>0.03</td>
<td>0.15</td>
<td>0.20</td>
<td>0.34</td>
<td>0.40</td>
<td>0.27</td>
<td>0.13</td>
<td>0.03</td>
<td>0.14</td>
<td>&lt; 0.01</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>Bootstrap p-value</td>
<td>0.01</td>
<td>0.01</td>
<td>&lt; 0.01</td>
<td>0.01</td>
<td>0.13</td>
<td>0.16</td>
<td>0.33</td>
<td>0.43</td>
<td>0.35</td>
<td>0.17</td>
<td>0.06</td>
<td>0.21</td>
<td>&lt; 0.01</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>Rank (delete later)</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>17</td>
<td>33</td>
<td>66</td>
<td>99</td>
<td>165</td>
<td>231</td>
<td>264</td>
<td>298</td>
<td>314</td>
<td>324</td>
<td>327</td>
<td>330</td>
</tr>
</tbody>
</table>

|                |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| **Bottom Panel: Ranking based on alpha t-statistic** |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Alpha t-statistic | 3.93 | 2.77 | 2.64 | 1.92 | 1.40 | 0.84 | 0.46 | −0.13 | −0.68 | −1.00 | −1.41 | −1.73 | −1.99 | −2.13 | −2.67 |
| Parametric p-value | < 0.01 | < 0.01 | 0.01 | 0.03 | 0.08 | 0.20 | 0.32 | 0.45 | 0.25 | 0.16 | 0.08 | 0.04 | 0.02 | 0.02 | < 0.01 |
| Bootstrap p-value | < 0.01 | 0.01 | 0.01 | 0.03 | 0.07 | 0.21 | 0.32 | 0.46 | 0.25 | 0.18 | 0.06 | 0.05 | 0.03 | 0.05 | < 0.01 |
The results in the top panel show that funds with estimated alphas at the 95\textsuperscript{th} percentile and above almost always show strong evidence of skill under the bootstrap. That is, the hypothesis that the realised alphas of these funds are simply random draws from their empirical distribution of pseudo-alphas (i.e. the distribution of abnormal returns under luck alone) is typically rejected even at small significance levels. The results are somewhat more mixed under the parametric approach. Whilst approximately half of the funds between the 90\textsuperscript{th} and 95\textsuperscript{th} percentile show evidence of skill using the bootstrap, and a quarter of the 80–90\textsuperscript{th} percentile-funds show evidence of skill, any positive abnormal returns generated by funds below the 80\textsuperscript{th} percentile seem to be solely the result of luck.

At the other end of the distribution, very few funds show evidence of ‘negative’ skill, that is, persistent underperformance relative to their benchmarks, particularly when using the bootstrapped alpha $p$-values. Amongst funds with low estimated alphas, the bootstrapped $p$-values are almost always larger that the parametric $p$-values, but the results are more mixed for the highest performing funds.

In the bottom panel, I rank the funds by their estimated alpha $t$-statistics. The main difference is that the null of no outperformance is rejected more often (see Horowitz (2003) for a discussion on the better power properties of the bootstrap), leading to a more equal proportion of funds with positive and negative skill. This is because in the tails of the distribution, the bootstrap ‘luck distributions’ have far fatter tails (more probability mass) than what is implied under the parametric assumptions due to the complex interaction between the (non-normal) estimated alphas of the individual funds, and the complexity of how these distributions are combined by the resampling process. And as mentioned previously, since the $t$-statistic scales the alpha point estimate by its standard error, the method of resampling from the cross-section of estimated $t$-statistics is superior to resampling based on the alpha cross-section.

These results are also evident from Table 4.2, where I compare these results with the
I compare the results of the parametric approach with the semi-parametric bootstrap approaches that use the simulated estimated alphas or simulated alpha $t$-statistics for permutation-based inference.

**Table 4.2: Inferring skill using parametric and bootstrap approaches**

<table>
<thead>
<tr>
<th></th>
<th>Parametric 5% level</th>
<th>Parametric 10% level</th>
<th>Resampling alphas 5% level</th>
<th>Resampling alphas 10% level</th>
<th>Resampling alpha $t$-statistics 5% level</th>
<th>Resampling alpha $t$-statistics 10% level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Skill (%)</td>
<td>3.65</td>
<td>8.21</td>
<td>4.56</td>
<td>6.99</td>
<td>7.60</td>
<td>13.07</td>
</tr>
<tr>
<td>Negative Skill (%)</td>
<td>2.13</td>
<td>6.38</td>
<td>0.30</td>
<td>1.51</td>
<td>6.08</td>
<td>12.76</td>
</tr>
<tr>
<td>Luck or no skill (%)</td>
<td>94.2</td>
<td>85.4</td>
<td>95.1</td>
<td>91.5</td>
<td>86.3</td>
<td>74.2</td>
</tr>
</tbody>
</table>

The main result is that the $t$-statistic simulations reject the null hypothesis more often than under the parametric approach, and the funds that show evidence of skill are largely balanced by those with negative skill. However, the vast majority of funds show no evidence of superior or inferior performance relative to passive investment strategies.

To concretise the effect that non-normally distributed abnormal returns have on hypothesis testing procedures, I present the simulated distributions of pseudo-alphas for funds at various points in the alpha cross-section in Figure 4.1. The vertical line represents the OLS-estimated alpha (hereafter, actual alpha) of each fund, which are compared to the simulated distributions which are smoothed with a kernel density estimator, using a Gaussian kernel. The optimal bandwidth (which considers the trade-off between a smoother density and increased bias) is chosen using the rule suggested by Silverman (1986), which is derived by minimising the integrated mean squared error. The top fund, for example, is the fund with the highest actual alpha, and the distribution approximates the distribution of alphas that would be generated under the null of no skill.

For positive-alpha funds, when testing the hypothesis $H_0 : \alpha_i \leq 0$, $H_1 : \alpha_i > 0$, I compute the $p$-value as the proportion of simulated alphas which exceed the actual alpha, and multiply this $p$-value by two to perform two-sided tests (the latter approach was used in Table 4.2 so the results could be compared with the parametric
results). For a fund with negative alphas, I test the hypothesis $H_0 : \alpha_i \geq 0$, $H_1 : \alpha_i < 0$, using the same method to compute $p$-values. The alpha of the top fund, for example, is clearly significant since its actual alpha of 1.74% per month is well within the right-tail rejection region of the simulated distribution.

I apply the same method to the simulated Newey-West adjusted $t$-statistics in Figure 4.2. Again, the vertical lines show the actual $t$-statistic of the funds relative to their ‘luck distributions’, which assume a true alpha (or equivalently, a true $t$-statistic) of zero. However the funds are now chosen based on their position in the $t$-statistic cross-section, so funds which are included differ from those included in Figure 4.1. The hypothesis tests are performed in the same way, except the $t$-statistics are used rather than the alphas. Once again the top fund, with $t_\alpha = 3.93$, which is well to the right of almost all observations in the $t$-statistic luck distribution (the $p$-value is 0.001), has produced true superior performance. Whilst this result is consistent with parametric testing procedures, there is much more divergence amongst the funds with poor performance, for example at the 2nd and 5th percentiles.

4.4 Analysis of the entire industry

To analyse the aggregate performance of all funds included in this study, I compare the cross-sectional distributions of simulated alphas and alpha $t$-statistics with the actual distributions. The simulated distributions are the $999 \times 329$ bootstrapped alphas or $t$-statistics that were generated for each fund in Section 4.2, and as before the simulated distributions assume that the returns of all of the funds are solely due to luck by imposing $H_0 : \alpha_i = 0$ for all funds. If the simulated and actual distributions are statistically different, we can conclude that at least some fund managers are skilled in selecting stocks, and the abnormal return generating process is influenced by factors other than exposure to undiversifiable risk. As before, the skill may be positive (where the fund consistently generates a return higher than its
common risk exposures) or negative.

Figure 4.3 (left panel) compares the cross-sectional density of actual alphas (navy) with the simulated alphas generated by the bootstrap (maroon). The corresponding empirical cumulative distributions are shown in the right panel, which allows us to compare the proportion of funds that are expected to achieve a given level of alpha performance purely by random chance (luck) with the proportion that actually achieved this level of performance. A Kolmogorov-Smirnov test gives evidence that they are statistically different ($p < 0.01$).

By observing Figure 4.3, it is clear that the top funds perform slightly better than what is expected under luck alone. For example, letting $F_A$ be the e.c.d.f of actual alphas and $F_S$ be the e.c.d.f of simulated alphas, examining the values of these functions at $\alpha = 0.5$ for example yields $F_A(0.5) = 0.944$ and $F_S(0.5) = 0.968$. So
the proportion of funds that achieve a level of alpha performance in excess of 0.5% p.m. (5.6% of funds) is somewhat larger than the proportion predicted to achieve this level of alpha performance solely by chance (3.2% of funds). This is also clear by examining the p.d.f’s, where the actual density has greater probability mass in the right tail.

At the other tail of the distribution, it is also evident that the worst performing funds perform slightly better than what is predicted by chance. That is, they generate abnormal returns that are less negative than what is predicted when the true alphas of all funds are set to zero. It is only in the centre of the distribution where \( F_A > F_S \), that funds in the centre of the distribution (ranked by their estimated alphas) perform worse than predicted by chance alone. A two-sample Kolmogorov-Smirnov test confirms that the two distributions are statistically different \((p < 0.01)\).
The equivalent graphs for the annualised abnormal return distributions are shown in Figure 4.4, where the annualised alphas are defined as $(1 + \alpha_m)^{12}$. In Table 4.3, I report the proportion of funds that are predicted to achieve a certain level of alpha performance purely by chance, and compare this with the proportion that actually achieve this performance. Note that since the absolute alpha is given in the left column, for negative alphas I report the proportion of funds with alphas that are more negative than each of the levels provided in the left column.

As discussed earlier, the large deviations from normality that are observed amongst the funds with the largest absolute alphas are typically a result of the high-risk strategies employed by these funds. This motivates the use of the alpha $t$-statistic as the primary measure used in performance evaluation. In Figure 4.5, I provide the density and distribution functions of the cross-section of alpha $t$-statistics, rather than the point estimates of the abnormal returns.

Once again, it is clear that actual and simulated distributions have quite different
shapes. For example, the density of actual alpha $t$-statistics has more probability mass in both tails, and less mass in the centre relative to the simulated distribution, and the actual distribution also has many non-standard features such as a fat right tail and positive skewness. The tails especially do not seem to be explained by sampling error in abnormal returns (represented by the simulated distribution). This again confirms that inference will differ from the parametric case since the bootstrap is able to capture the complex shape of the cross-section of $t$-statistics under the null hypothesis.

In terms of the e.c.d.f, the implied proportion of funds that exceed a level of performance by chance compared to what is actually observed are qualitatively consistent with the findings when analysing results by estimated abnormal returns. The results confirm that many funds exhibit significant outperformance that is not explained by luck alone, and there does seem to be some skill involved in generating abnormal returns amongst the top funds. The bottom line is that the top funds perform better, the bottom funds perform less worse, and the middle funds perform
Table 4.3: Comparison of observed annualised alphas with distribution under luck alone

<table>
<thead>
<tr>
<th>Absolute Alpha (p.a.)</th>
<th>Positive alpha values</th>
<th>Negative alpha values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Observed</td>
<td>% Under luck alone</td>
</tr>
<tr>
<td>&gt; 20%</td>
<td>0.31</td>
<td>0.04</td>
</tr>
<tr>
<td>&gt; 18%</td>
<td>0.61</td>
<td>0.08</td>
</tr>
<tr>
<td>&gt; 16%</td>
<td>0.61</td>
<td>0.13</td>
</tr>
<tr>
<td>&gt; 14%</td>
<td>0.61</td>
<td>0.23</td>
</tr>
<tr>
<td>&gt; 12%</td>
<td>1.22</td>
<td>0.42</td>
</tr>
<tr>
<td>&gt; 10%</td>
<td>1.83</td>
<td>0.79</td>
</tr>
<tr>
<td>&gt; 9%</td>
<td>2.45</td>
<td>1.10</td>
</tr>
<tr>
<td>&gt; 8%</td>
<td>2.75</td>
<td>1.57</td>
</tr>
<tr>
<td>&gt; 7%</td>
<td>4.58</td>
<td>2.32</td>
</tr>
<tr>
<td>&gt; 6%</td>
<td>5.81</td>
<td>3.47</td>
</tr>
<tr>
<td>&gt; 5%</td>
<td>7.65</td>
<td>5.38</td>
</tr>
<tr>
<td>&gt; 4%</td>
<td>10.09</td>
<td>8.58</td>
</tr>
<tr>
<td>&gt; 3%</td>
<td>11.93</td>
<td>13.87</td>
</tr>
<tr>
<td>&gt; 2%</td>
<td>20.18</td>
<td>22.41</td>
</tr>
<tr>
<td>&gt; 1%</td>
<td>29.66</td>
<td>34.80</td>
</tr>
<tr>
<td>&gt; 0.5%</td>
<td>36.70</td>
<td>42.16</td>
</tr>
</tbody>
</table>

more poorly relative to the case where all abnormal returns are generated by chance, or equivalently where no fund managers have any skill in selecting stocks. This holds when the funds are ranked based on their estimated alpha, or their alpha $t$-statistic.

The finding that the performance of the top funds cannot be solely attributed to luck allows us to reject the hypothesis that the Australian equity market is perfectly efficient in the strong-form sense of Fama [1970], or in the semi-strong-form if we assume that no mutual funds trade based on insider information. That is, the level of skill reflected in the returns of Australian funds implies that there are at least some cases where prices do not fully reflect all available information and there are some inefficiencies to exploit. But given that there are so few funds that significantly outperform (or underperform) relative to the ‘luck only’ case; the model of market efficiency seems to broadly hold rather well in the Australian equities market.

According to the equilibrium derived in Grossman and Stiglitz [1980], the finding that markets are “mostly, but not completely efficient” is required for fund managers
to make the effort to determine whether stocks are mispriced. So although the contest between active managers might be a zero-sum game, active management as a whole is not, since by making markets more efficient, active management improves capital allocation, thereby contributing to economic efficiency and growth. Thus, the abnormal returns generated by the mutual funds can be viewed as an economic rent for gathering and processing information, with the process of identifying mispriced stocks making the Australian equity market more efficient. I use this definition to approximate the value added by Australian equity mutual funds in Section 5.1.

However one caveat is that market efficiency cannot be tested without imposing a model of market equilibrium (in this study, the Carhart model) which specifies equilibrium expected returns and stipulates a measure of outperformance. This is known as the joint hypothesis problem (Fama, 1970). Whilst the Carhart (1997) model is by far the most widely used returns-based model used to evaluate equity managers, it is possible that there are other sources of common variation that explain the expected returns of stocks, or that the variation is not controlled for fully. For example, Fama and French (2013) suggest an alternative four-factor model that uses a profitability proxy based on Novy-Marx (2012) in place of momentum as the fourth factor. I leave the implications of the inclusion of non-Carhart factors for further research.²

²The fact that the Fama and French (2013) paper was only made available on SSRN in July, and has already been downloaded more than 2000 times (as of October 2013), suggests that there will be a substantial amount of work in this area.
Figure 4.5: Actual vs. simulated alpha $t$-statistics
CHAPTER 5

Further Analysis

5.1 Industry Value Added

Using the results from Section 4.4, I develop a new method that can be used to approximate the value added created by any subset of actively managed funds, and apply this method to the sample of funds used in this study. This allows the overall economic impact of the findings in the previous section to be evaluated, and also controls for the size of the funds: If the assets under management of the top performing funds are relatively low, for example, this is unlikely to have any significant implications for the average mutual fund investor, and the economic rent generated by these funds will be relatively small.

To approximate the value added created by the entire Australian equity mutual fund industry (subject to the conditions discussed in Section 3.2), I compare the income generated by each fund that is not explained by exposure to undiversifiable risk ($\alpha \times$ Fund size) to the income that would be generated by luck alone (where the true alpha is zero for all funds), estimated by the simulated alpha distribution. The difference gives an estimate of value added. The justification of this approach is that under luck alone, active management is providing no true value to investors, so there is zero value added.

Specifically, I compare the histogram of the actual alphas with the histogram implied by the bootstrap simulation. If the actual alpha is larger than the simulated alpha at a given percentile, actively managed funds are creating more value that what
would be expected under luck alone at that percentile, resulting in positive value added. The intuition of this approach is provided in Figure 5.1, where I show the values of the actual and simulated alpha distributions at each percentile.

Figure 5.1: Actual and simulated alphas at each percentile

To compute estimated value added, I first order the ‘actual’ monthly alphas, and denote these as \( \hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N \), where \( N = 329 \) in this study. With this, determine at what percentiles these alphas lie on the actual alpha distribution, and denote these percentiles as \( p_1, p_2, \ldots, p_N \).\(^1\) Next, I compute the values of

\[ p_1 = 0.302, p_2 = 0.604, \ldots \]
the simulated alphas at the same percentiles of the simulated alpha distribution, and denote these simulated alphas as \( \alpha_s^{(1)}, \alpha_s^{(2)}, \ldots, \alpha_s^{(N)} \). I then retrieve the sizes of each fund, and define \( Size_{(i)} \) to be the size of the fund with the \( i \)th smallest actual alpha. The net asset values of each fund are used as a proxy for fund size. Recall that the total returns of each fund are computed by taking the percentage change in net asset values each month, reinvesting all income and capital gains distributions. Finally, I estimate the monthly value added of the entire Australian equity mutual fund market (VA) as follows:

\[
VA = \sum_{i=1}^{N} \left[ (\hat{\alpha}_{(i)} - \alpha_s^{(i)}) \times Size_{(i)} \right]
\]  

Applying this formula to each of the \( N=329 \) funds included in this study yields a value-added estimate of A$95 million per month. This compares to the A$150 billion that the funds have under management as of April 2013. I conclude that the industry is generating a model amount of wealth for their investors (\( \sim 0.06\% \) of the aggregate NAV each month), however the wealth creation is concentrated in a few select top performing funds. An analysis of fund performance by net asset value is provided in Section 5.3.

5.2 Performance by investment style

The proliferation of mutual funds that specialise in either value or growth stocks, or small-cap or large-cap stocks, was sparked in large part due to the findings of Fama and French (1992), who showed that small-cap stocks and value (high book-to-market) stocks have higher expected returns than the market, and that these anomalies are not explained by the CAPM market factor. The influence that this finding has had on the Australian market is clear from the high proportion (91 per cent) of funds which that tilt toward or away from at least one of these two

---

2It is currently the second highest cited article in the Journal of Finance since 1970, behind Fama (1970).
dimensions types of stocks. However the issue of whether there are any systematic differences in fund manager ability across the size and book-to-market dimensions is yet to be explored in an Australian context.

To classify each fund by investment style, I use data provided by Morningstar which identifies whether a fund tilts across any of these two axes. In the case of missing data (around a quarter of funds), I used company websites or internet searches to retrieve this information.

I will begin by analysing performance based on exposure to small-cap vs. large-cap stocks, which yields the most significant results when analysing performance by observable characteristics. Using the definitions given in Morningstar; there are 64 small-cap funds; 203 large-cap funds, and 62 funds which focus on mid-cap stocks or don’t tilt with respect to size, that is, they don’t select stocks based on their market-capitalisations (I include these funds in the mid-cap category). The distribution of abnormal returns of small-cap, mid-cap and large-cap funds are statistically different from each-other, and this also holds for the alpha \( t \)-statistics. The densities are shown in Figure 5.2 and the corresponding cumulative distribution functions are shown in Figure 5.3. As in previous sections, I smooth the distributions using a Gaussian kernel.

The small-cap funds are by far the best performers—for example the average small-cap fund has an alpha of 0.2% per month and a \( t \)-statistic of 0.53. Perhaps even more impressively, 26.6 per cent of small-cap funds show evidence of positive skill (using the \( t \)-statistic bootstrap) compared to just 3.1 per cent that have evidence of negative skill. This compares to 13.2 and 12.8 per cent of the total population of funds that have positive and negative skill respectively. A full analysis (including the results by value/growth) is given in Table 5.1A in the ‘Size Dimension’ row. In contrast, there are more negatively skilled large-cap and mid-cap funds than positively skilled funds. The abnormal return of the average mid-cap fund is only 0.05% p.m. \((t = 0.04)\), and only \(-0.06\)% p.m. \((t = -0.28)\) for the average large-cap
fund. This compares to an average alpha of 0.01% p.m. ($t = -0.06$) for all of the funds included in this study.

**Figure 5.2: Probability density functions of alphas and $t$-statistics by size tilt**

There is less of a divergence when analysing the results based on the book-to-market dimension. Morningstar identifies 34 value funds, 73 growth funds, and 222 core funds (which includes funds that don’t tilt along the book-to-market axis), and I again find a significant difference in the distributions. First, the average alphas and $t$-statistics are marginally positive for value and growth funds, but negative for core funds. Whilst the proportion of ‘positive’ skill is roughly the same across the three groups, a very high proportion of core funds are negatively skilled (15.5 per cent), but this proportion is quite low for value and core funds (of which 5.9 and 2.2 per cent respectively are negatively skilled). Again, I plot the PDFs and CDFs of the alphas and $t$-statistics for these three groups (Figures 5.4 and 5.5), and the results are reported in Table 5.1A. However it is also interesting to analyse the results by considering both of these axes together, as I report in Table 5.1B. An interesting result is whilst the superior performance of small-cap funds seems
to be unaffected by their exposure to value or growth stocks, the performance of value and growth stocks seem to be far more affected by their exposure to small or large-cap stocks. The implication is that the size dimension has far more power in explaining fund performance compared to the book-to-market dimension. I conclude that, after disentangling the ‘size’ dimension from the ‘book-to-market’ dimension, there doesn’t appear to be a substantial difference in the performance of value funds relative to growth funds.

In Figures 5.4 and 5.5, I show the p.d.f’s and c.d.f’s of the alphas and alpha $t$-statistics, and a summary of this section’s results is given in Table 5.1.
Figure 5.4: PDF’s of alphas and t-statistics by book-to-market tilt

Figure 5.5: CDF’s of alphas and t-statistics by book-to-market tilt
Table 5.1: Analysis of results by investment style

A:

<table>
<thead>
<tr>
<th>Investment Style</th>
<th>Freq.</th>
<th>Per cent</th>
<th>Ave. alpha p.m. (%)</th>
<th>Ave. alpha p.a. (%)</th>
<th>Ave. t-stat</th>
<th>Positive Ave mgmt.</th>
<th>Negative Ave mgmt.</th>
<th>Ave fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>34</td>
<td>10.33</td>
<td>0.09</td>
<td>1.1</td>
<td>0.10</td>
<td>11.8</td>
<td>5.9</td>
<td>1.32</td>
</tr>
<tr>
<td>Growth</td>
<td>73</td>
<td>22.19</td>
<td>0.05</td>
<td>0.6</td>
<td>0.13</td>
<td>16.4</td>
<td>2.2</td>
<td>1.36</td>
</tr>
<tr>
<td>Core</td>
<td>222</td>
<td>67.48</td>
<td>−0.01</td>
<td>−0.2</td>
<td>−0.16</td>
<td>12.3</td>
<td>15.5</td>
<td>1.38</td>
</tr>
<tr>
<td><strong>BE/ME dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-cap</td>
<td>64</td>
<td>19.45</td>
<td>0.20</td>
<td>2.5</td>
<td>0.53</td>
<td>26.6</td>
<td>3.1</td>
<td>1.4</td>
</tr>
<tr>
<td>Mid-cap</td>
<td>62</td>
<td>18.84</td>
<td>0.05</td>
<td>0.6</td>
<td>0.04</td>
<td>16.4</td>
<td>19.7</td>
<td>1.37</td>
</tr>
<tr>
<td>Large-cap</td>
<td>203</td>
<td>61.70</td>
<td>−0.06</td>
<td>−0.8</td>
<td>−0.28</td>
<td>7.9</td>
<td>13.9</td>
<td>1.36</td>
</tr>
<tr>
<td><strong>Size dimension</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>329</td>
<td>100</td>
<td>0.01</td>
<td>0.1</td>
<td>−0.06</td>
<td>13.2</td>
<td>12.8</td>
<td>1.37</td>
</tr>
</tbody>
</table>

B:

<table>
<thead>
<tr>
<th>Investment Style</th>
<th>Freq.</th>
<th>Per cent</th>
<th>Ave. alpha p.m. (%)</th>
<th>Ave. alpha p.a. (%)</th>
<th>Ave. t-stat</th>
<th>Positive Ave mgmt.</th>
<th>Negative Ave mgmt.</th>
<th>Ave fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>17</td>
<td>5.17</td>
<td>−0.06</td>
<td>−0.7</td>
<td>−0.33</td>
<td>0.0</td>
<td>5.9</td>
<td>1.32</td>
</tr>
<tr>
<td>Growth</td>
<td>39</td>
<td>11.85</td>
<td>−0.03</td>
<td>−0.4</td>
<td>−0.13</td>
<td>10.3</td>
<td>7.7</td>
<td>1.37</td>
</tr>
<tr>
<td>No BE/ME tilt</td>
<td>147</td>
<td>44.68</td>
<td>−0.07</td>
<td>−0.9</td>
<td>−0.32</td>
<td>8.2</td>
<td>16.4</td>
<td>1.36</td>
</tr>
<tr>
<td><strong>Large-cap funds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>7</td>
<td>0.91</td>
<td>0.37</td>
<td>4.5</td>
<td>1.21</td>
<td>28.6</td>
<td>0.0</td>
<td>1.39</td>
</tr>
<tr>
<td>Growth</td>
<td>20</td>
<td>5.47</td>
<td>0.11</td>
<td>1.3</td>
<td>0.35</td>
<td>20</td>
<td>10</td>
<td>1.38</td>
</tr>
<tr>
<td>No BE/ME tilt</td>
<td>6</td>
<td>1.82</td>
<td>0.24</td>
<td>2.9</td>
<td>1.19</td>
<td>33.3</td>
<td>0.0</td>
<td>1.72</td>
</tr>
<tr>
<td><strong>Mid-cap funds/No size tilt</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>10</td>
<td>3.04</td>
<td>0.20</td>
<td>2.4</td>
<td>0.38</td>
<td>20.0</td>
<td>0.0</td>
<td>1.34</td>
</tr>
<tr>
<td>Growth</td>
<td>14</td>
<td>4.26</td>
<td>0.21</td>
<td>2.6</td>
<td>0.71</td>
<td>28.6</td>
<td>0.0</td>
<td>1.4</td>
</tr>
<tr>
<td>No BE/ME tilt</td>
<td>40</td>
<td>12.16</td>
<td>0.20</td>
<td>2.4</td>
<td>0.50</td>
<td>27.5</td>
<td>5.0</td>
<td>1.42</td>
</tr>
<tr>
<td><strong>Small-cap funds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No tilt</td>
<td>29</td>
<td>8.81</td>
<td>−0.08</td>
<td>−0.9</td>
<td>−0.53</td>
<td>7.1</td>
<td>25.9</td>
<td>1.36</td>
</tr>
<tr>
<td>All</td>
<td>329</td>
<td>100</td>
<td>0.01</td>
<td>0.1</td>
<td>−0.06</td>
<td>13.2</td>
<td>12.8</td>
<td>1.37</td>
</tr>
</tbody>
</table>

63
5.3 Performance by other observable characteristics

I also analyse the performance of the funds based on their management fee and size (net asset value). The management fee encompasses all direct expenses involved in managing the investments, which includes the hiring of the portfolio manager and investment team. However, the costs involved in buying and selling securities are not included in the management fee. In Figure 5.6, I plot the histogram of the management fees in the left panel. Given the very wide dispersion of net asset values, I plot the histogram of NAVs of funds below $100 million in the middle panel, and the funds with NAV’s above $100 million are plotted in the right panel.

The results are reported in Table 5.2. With respect to the management fee, the high-fee funds perform the poorest. The average alpha of the high-fee funds is $-0.05\%$ p.m., and the proportion of negative skilled funds is almost double the proportion of positively skilled funds. Whilst the low-fee funds perform marginally better than the medium-fee funds, there is not much of a difference.

Figure 5.6: Histograms of management fees and net asset values

The relationship between performance and net asset value is even stronger, and there is a clear negative relationship between the size of the fund and its net asset value.
A huge 39.1 per cent of high NAV funds show evidence of positive skill under the $t$-statistic bootstrap, compared to only 3.1 per cent of funds that show evidence of negative skill, suggesting that there may be a ‘smart money effect’ in the Australian market, where investors are able to predict mutual fund performance and invest accordingly. These full results are reported in Table 5.2.

**Table 5.2: Performance by management fee and net asset value**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Split</th>
<th>Ave. alpha p.m. (%)</th>
<th>Ave. alpha p.a. (%)</th>
<th>Ave. $t$-stat</th>
<th>Positive skill (%)</th>
<th>Negative skill (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management fee</td>
<td>High</td>
<td>-0.05</td>
<td>-0.6</td>
<td>-0.37</td>
<td>7.9</td>
<td>15.9</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0.04</td>
<td>0.5</td>
<td>0.04</td>
<td>17.0</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0.00</td>
<td>-0.1</td>
<td>-0.03</td>
<td>6.3</td>
<td>7.9</td>
</tr>
<tr>
<td>Net asset value</td>
<td>High</td>
<td>0.18</td>
<td>2.2</td>
<td>0.87</td>
<td>39.1</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>-0.01</td>
<td>-0.2</td>
<td>-0.21</td>
<td>7.3</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>-0.08</td>
<td>-1.0</td>
<td>-0.52</td>
<td>6.3</td>
<td>19.1</td>
</tr>
</tbody>
</table>

In the appendices (Figures A.1 to A.4), I plot the management fee and log net asset value against the two measures of fund performance (the alpha point estimate and the alpha $t$-statistic) by the fund’s investment style. In Figures A.5 to A.8, plot the PDFs and CDFs of each management fee and NAV group for both the estimated alphas and the alpha $t$-statistics. It seems that the NAV-performance relationship is strongest when using the alpha $t$-statistic as the performance measure, and that the relationship is strongest for growth funds.

In Figure 5.7, I compare these Morningstar ratings with the estimated alpha and $t$-statistics. Whilst the relationship is positive and statistically significant for both measures of performance, the relationship is stronger for the estimated alpha. I do not find evidence of a statistically significant bias in the Morningstar ratings with respect to the investment styles of the funds. However the relatively weaker relationship with the alpha $t$-statistics suggest that the heterogeneity in risk taking amongst the funds has not been fully accounted for by the Morningstar analysts.
Figure 5.7: Morningstar analyst ratings vs. alpha and alpha $t$-statistics
Using a data set of Australian equity mutual funds that primarily invest in Australian stocks and have at least three years of data between 1986–2013, I apply Carhart’s four-factor model to each fund to isolate the portion of monthly returns generated by stock-selection ability, rather than exposure to common (and therefore undiversifiable) equity risk factors. These factors were constructed using the monthly time series of return indices, market capitalisations, book values and market values of every stock that has been listed on the Australian Securities Exchange between 1985 and 2013.

I subsequently apply a semi-parametric bootstrap residual resampling algorithm to each of the funds to simulate the ‘luck distribution’ of abnormal returns and alpha t-statistics of each fund. I find that the proportion of funds with significant outperformance or underperformance (conditional on their exposure to the common risk factors) is a couple of percentage points higher than case where all non-zero abnormal returns are solely the result of luck.

The use of a parametric approach makes it much more difficult to disentangle the effects of luck from true stock-selection ability—given the large number of funds used in this study (329); it is likely that a true null hypothesis of zero alpha will be rejected for some of the funds. Since this will be combined with a (small) number of funds that exhibit genuine outperformance, I evaluate the performance of the funds in aggregate to approximate the proportion of funds that achieve high (and low) alphas or alpha t-statistics purely by chance, and compare this to the proportion
that is actually realised. I find that the best and worst performing funds generate higher abnormal returns that what is predicted purely by chance, but the funds in the centre of the distribution perform worse relative to the case where all returns are generated by chance. Since the sample of funds (in aggregate) hold a portfolio of stocks that exhibits similar behaviour to the market portfolio, for the most part I expect a fund which beats the market in month $t$ to be balanced to some degree by a fund who loses in month $t$.

When analysing the results by observable characteristics, I find that the top funds tend to be made up of either small-cap funds, or funds with high net asset values. I also generalise the results to develop a new way of estimating the value created by some population of active fund managers. Applying this measure to the funds included in this study, I find that the funds create an average of $95$ million in value each month, which compares to the $150$ billion in aggregate net assets under management (as of April 2013).

The bottom line result is that the study has provided strong evidence in favour of using the bootstrap to evaluate the performance of Australian mutual funds. However even if a parametric approach is used, the alpha $t$-statistic rather than the point estimate should be used to evaluate the performance of mutual funds in order to mitigate the problems associated with heterogeneity and clustering in risk-taking amongst equity mutual funds.
**Appendix A**

Additional Figures

**Figure A.1:** Management fee vs. estimated alpha by investment style

![Figure A.1](image1)

**Figure A.2:** Management fee vs. alpha $t$-statistic by investment style

![Figure A.2](image2)
Figure A.3: Log net asset value vs. estimated alpha by investment style

Figure A.4: Log net asset value vs. alpha $t$-statistic by investment style
Figure A.5: PDF’s of alphas and $t$-statistics by management fee

Figure A.6: CDF’s of alphas and $t$-statistics by management fee
Figure A.7: PDF’s of alphas and $t$-statistics by net asset value

![PDF’s of alphas and $t$-statistics by net asset value](image)

Figure A.8: CDF’s of alphas and $t$-statistics by net asset value

![CDF’s of alphas and $t$-statistics by net asset value](image)


Bowers, C. and C. Heaton (2013). What does high-dimensional factor analysis tell


MacKinlay, A. C. (1995). Multifactor models do not explain deviations from the


Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*
(2nd ed.). Basil Blackwell.

Merton, R. (1971). Optimum consumption and portfolio rules in a continuous-time
model,. *Journal of Economic Theory* 3, 373–413.

Miller, M. and M. Scholes (1972). Rate of return in relation to risk: A reexamination
of some recent findings. In M. C. Jensen (Ed.), *Studies in the Theory of Capital

Moskowitz, T. J. and M. Grinblatt (1999). Do industries explain momentum? *The

768–783.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedastic-
ity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–
708.


based on earnings’ yields and market values. *Journal of Financial Economics* 9,
19–46.


