# The Effect of Security Return Dispersion on Performance Measurement in a South African Context

**Bryce A. Gething**



A dissertation submitted to the Department of Actuarial Science, Faculty of Commerce, at the University of the Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

4th December 2014

*Master of Philosophy specializing in Mathematical Finance,*
*University of the Cape Town,*
*Cape Town.*

# Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy in the University of the Cape Town. It has not been submitted before for any degree or examination in any other University.

______________________________

4th December 2014

# Abstract

This work replicates a similar study performed by de Silva *et al.* (2001). Our study was performed on the South African market. de Silva *et al.* (2001) studied the effect of cross-sectional volatility (CSV) on fund managerial skill measurement. This lead to the conjecture that increased fund performance dispersion was primarily due to higher CSV, and not changes in informational efficiency or ranges in managerial talent.

In this dissertation we firstly critique the CSV-adjusted alpha as a measure of fund performance and show that it can only be used as a means of normalising fund performance, yet reveals very little with regard to managerial talent. Since fund performance is intrinsically linked to CSV, we find it difficult to disentangle the effects of CSV and managerial talent dispersion. Adjusting for CSV therefore also implies adjustment for managerial talent, and we conclude with ideas for how a CSV-adjusted alpha may be used to assess manager talent.

# Acknowledgments

I would like to express my very great appreciation to Dr Daniel Polakow for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been very much appreciated.
I wish to thank the National Research Foundation for funding me and providing me the opportunity to complete this work.
I would also like to thank Alexander Forbes for supplying me with real world fund return data.
I would like to express my deep gratitude to Obeid Mahomed and Ralph Rudd, for their patient guidance, enthusiastic encouragement and useful critiques of this research work.
Finally, I wish to thank my lecturers, family and friends for their support and encouragement throughout my study.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

In the United States, research on cross-sectional volatility (CSV) by de Silva *et al.* (2001) has found that fund return dispersion can be attributed to wider dispersion amongst the available stock choices in the market and therefore, by implication has little to do with the range of managerial skill. Researchers had to reconsider how fund performance was calculated as a result of the return dispersion increases. Instead of just considering performance measurement as a function of benchmark exceedances, researchers incorporated return dispersion information as well. They concluded their work by adjusting alpha estimates for the variance across the range of constituent securities using two distinct measures of security return dispersion: periodic and asset class. To compute alpha estimates researchers used ordinary least squares regression (OLS) given as

$$(r_{P,t} - r_{F,t}) = \alpha_P + \beta_P(r_{M,t} - r_{F,t}) + \varepsilon_{P,t}, \tag{1.1}$$

where $r_{P,t}, r_{M,t}$ and $r_{F,t}$ are time series returns for the active portfolio, market and risk free asset respectively.

**Remark 1.1** (Market Model Estimates)**.** *For a more intuitive understanding of Equation* (1.1)*, the market model, we can interpret $\alpha_P$, $\beta_P$ and $\epsilon_P$ as follows*

*($\alpha_P$)* ***Portfolio Alpha*** *Often referred to as abnormal return in the literature (see Strong (1992)), alpha is the intercept term of the market model. Alpha is a measure of risk compensation and can, when scaled appropriately, be used as a measure for fund performance comparisons. It is possible for alpha to take on any value along the real line however values larger (smaller) than 3% (-3%) are far less likely to occur in practice.*

*($\beta_P$)* ***Portfolio Beta*** *The beta of a portfolio is a measure of the portfolio's return sensitivity to movements in the benchmark's return to which it is*

*being compared. In practice, betas greater than one represent securities which are more sensitive to benchmark movements, betas between zero and one are less sensitive to benchmark movements and finally, securities with zero betas bear no benchmark sensitivity. Investors with higher risk aversions are attracted to betas less than one, and risk lovers, to betas greater than one. To understand why this is true, consider Equation* (1.2)*, the variance of the portfolio:*

$$\sigma^2_{r_{p,t}} = \beta^2_P \sigma^2_{r_{M,t}} + \sigma^2_{\varepsilon_{P,t}}. \tag{1.2}$$

*We define* $\sigma^2_{r_{p,t}}$*,* $\beta^2_P\sigma^2_{r_{M,t}}$ *and* $\sigma^2_{\varepsilon_{P,t}}$ *as the variance of the portfolio return, systematic risk and portfolio specific risk respectively. Note that as the beta increases ceteris paribus the variance of the portfolio also increases. Thus betas larger than one are associated with lower risk aversions. We extract portfolio betas from the market model, where beta is the slope of this model.*

*(*$\varepsilon_{P,t}$*)* ***Portfolio Epsilon*** *The error, commonly referred to in the literature as idiosyncratic risk has negligible correlation with market risk and is often diversified away by constructing funds which contain more than one asset. In the market model we assume epsilon is normally distributed with mean zero and variance* $\sigma^2_{\varepsilon_{P,t}}$*.*

Ankrim and Ding (2002) provide an insightful mathematical definition of CSV which we utilise in the next Section of this Chapter. The primary focus of their research was the relationship of active manager dispersion (95th versus 5th percentile) to CSV. They asked and answered three questions:

1. **Question:** Is the increasing CSV associated solely with U.S. large-caps?
   **Answer:** After investigating market wide CSV and equity markets outside of the U.S. they concluded that the association is not limited to U.S. large-caps.

2. **Question:** Has it happened before?
   **Answer:** Considering 20 years of historical CSV data from Japans equity market they found that this was not the first time markets experienced increasing CSV.

3. **Question:** Is it a result of the technology boom (bust)?
   **Answer:** By comparing CSV data computed inside the “technology boom” period (1999 to 2001) with data outside of the boom period they found no

statistically significant difference, concluding that the increasing CSV was not a result of the technology boom (bust).

Ankrim and Ding (2002) were unable to determine the cause of the increasing CSV observed from 1996 to 2001.

Raubenheimer (2012) posed the question: "How can fund sponsors fairly and accurately evaluate their managers performance in light of the changing CSV of realised returns?" The author established that the existence of dispersion in fund returns over time produced a heteroskedastic time series. Heteroskedasticity renders regular parametric tests meaningless namely, OLS regression. The alpha from OLS regression would normally be used to compare funds and establish which is superior. Raubenheimer advocated Sharpe ratio calculations, weighting returns and weighted least squares (WLS) regression as the appropiate adjustments in light of the heteroskedasticity evident in data. This approach mirrored de Silva *et al.* (2001).

This dissertation focuses on the concept of cross-sectional volatility (CSV), thus an appropriate definition follows as a prelude to the research. Chapter 2 details a simulation of a simplified market containing stocks, funds and a bond. In Chapter 3 we briefly discuss fund and CSV data observed from the South African market, choosing appropriate time windows for analysis. Next, applying the relevant methodologies defined in Chapter 4 we generate results found in Chapter 5. These methodologies include least squares regression (the market model), weighted least squares regression (CSV adjusted), and tests for heteroskedasticity. In Chapter 6 we conclude by answering the research question: "Does cross-sectional volatility effect performance measurement in South Africa?", summarising the relevant findings and suggesting further research.

**Important notice:** All the institutional data was graciously provided by Alexander Forbes. The institutional data used in this dissertation is a manipulated version of the Forbes data. Thus all alphas calculated and presented are not an exact reflection of the Forbes data (see Section 4.4). We would like to thank Alexander Forbes for providing this invaluable data for the purpose of this research.

## 1.1 What is Cross-sectional Volatility?

Cross-sectional volatility (CSV) is formally defined as a metric which measures the dispersion of a set of asset returns at a particular point in time. Note, that this is different from asset return volatility (also known as longitudinal volatility) which measures the return dispersion for a single asset over a time period. We have illustrated the difference between CSV and longitudinal volatility in Figure 1.1.

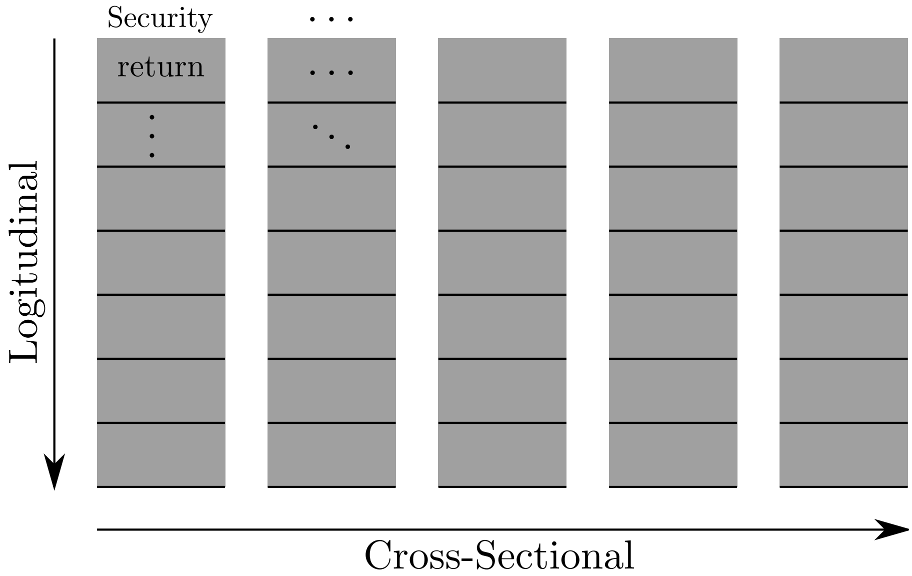


**Fig. 1.1:** Visual comparison of cross-sectional and longitudinal volatility.

Each column in Figure 1.1 represents an asset (security) and each grey block represents a return at a particular date. Longitudinal volatility is calculated by square-rooting the average of squared differences of the columns, whereas CSV is calculated by square-rooting the average of the squared differences of the rows.

As per Ankrim and Ding (2002), let cross-sectional variance be denoted as $X_t^2$ for a time period $t$. Note that we would have to square-root $X_t^2$ to attain CSV. Then for a market containing $N_t$ assets, cross-sectional variance is defined by

$$X_t^2 = \sum_{i=1}^{N_t} w_{it}(r_{it} - r_{mt})^2, \tag{1.3}$$

where $w_{it}$ is the market portfolio weight of stock $i$ at the beginning of time period $t$, $r_{it}$ is the return of stock $i$ over the time period associated with $t$, $r_{mt} = \sum_{i=1}^{N_t} w_{it} r_{it}$ is the weighted market return over time period $t$ and $N_t$ is the number of stocks in the market at the beginning of time period $t$. Substituting $\frac{1}{N_t}$ for $w_{it}$ in Equation (1.3) enforces equal weights for all stocks and calculates uniform CSV while the expression as given is used for market capitalisation-weighted CSV calculations.

# Chapter 2

# Simulating the Effect of Cross-sectional Volatility

In this chapter we study the effects of cross-sectional volatility (CSV) on fund performance in a controlled simulated market. We start off by making several assumptions and then close off with result interpretation.

## 2.1 Simulation Assumptions and Methodology

Firstly we list several assumptions about the market and the models' for stocks and funds:

1. Assume there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\mathcal{F}_t$. All assets in the market are adapted to $\mathcal{F}_t$.

2. A constant risk free rate of interest exists $(r)$ and we define $B_t$, a bond price process with the dynamics
$$dB_t = rB_t dt.$$

3. The stock price $S_t^{(i)}$ obeys the dynamics of a geometric Brownian motion.
$$dS_t^{(i)} = \mu_i S_t^{(i)} dt + \sigma_i S_t^{(i)} dX_t^{(i)},$$
where $X_t^{(i)}$ is a correlated $\mathbb{P}$-Brownian motion and $dX_t^{(i)} dX_t^{(j)} = \rho_{ij} dt$, $S_t^{(i)}$ the price of stock $i$ at time $t$ and $\mu_i$ and $\sigma_i$ are the drift and volatility parameters of stock $i$ respectively.

4. No stock pays dividends.

5. It is possible to borrow and lend any amount of cash at the risk free rate.

6. It is possible to go long any amount of stock (i.e. It is not possible to go short).

7. The market is frictionless, i.e. there are no costs associated with any of the above transactions.

8. If there are $N_S$ unique stocks in the market, a manager can only have half that number of stocks in their fund.

9. Fund managers re-balance at every time point.

10. Funds are constructed at random, i.e. Managers choose stocks at random and assign random weights to each stock.

11. Funds contain a maximum of 30% cash and 70% equity. It not possible for either of these to have zero weight.

12. CSV used in this model is uniformly weighed and computed using stock returns.

We made several further assumptions about the initial conditions of the market:

**Tab. 2.1:** Table of model parameters and their descriptions

| Parameter | Description |
|---:|---|
| Number of Stocks | A constant, $N_S = 50$ |
| Number of Funds | A constant, $N_F = 10$ |
| Path length | A constant, $n = 250$ |
| Initial Stock Price | A constant, $S_0^{(i)} = 100$ |
| Initial Bond Price | A constant, $B_0^{(i)} = S_0^{(i)}$ |
| Drift | $\mu_i = a_\mu + b_\mu u$ |
| Volatility | $\sigma_i = a_\sigma + b_\sigma u$ |

In Table 2.1, $a_\mu, a_\sigma$ and $b_\mu, b_\sigma$ are shifting and scaling parameters respectively and $u$ is a pseudo-random uniform number between zero and one.

Combining all the aforementioned assumptions we constructed a market by generating stock price paths and one bond price path (The market only contains two types of assets). We then computed returns, $\text{Return} = \frac{\text{Asset}_{t+1}}{\text{Asset}_t} - 1$, for each asset. Next, using a notional index value of 100 in conjunction with the growth of $S_t^{(i)}$ we generated a price path for each $S_t^{(i)}, i = 1, \ldots, N_S$. Averaging over all these price paths we computed the index (Also referred to as the benchmark in this simulation) price path[1], which was then used to compute the index return, $I_r$. Another set

[1] We have thus created a uniformly weighted index.

of data points we needed were fund returns, the following algorithm was used to construct the fund returns:

1. Generate $N_F$ uniform pseudo-random integers $k_j$ between two and $N_S/2$. $k_j$ is the number of stocks fund $j$ will pick.

2. Sample $k_j$ uniform pseudo-random integers between one and $N_S$ without replacement and store them in a vector $q_j$. For example, let $k_1 = 3$ and $N_S = 10$ then we choose 3 numbers at random between one and ten without replacement, $q_j = \{5, 2, 9\}$. Note that $q_j$ will never have repeated elements i.e. $q_s \neq q_t, \forall s \neq t$. The elements of $q_j$ represent the stocks picked by fund $j$.

3. Generate $w_j, j = 1, \ldots, N_F$ where $w_j$ is a uniform pseudo-random vector of integers. The first $k_j$ elements of $w_j$ are chosen between one and 70 (allocation to stock $i$) and the last element is chosen between ten and 30 (allocation to cash) therefore $w_j$ has $x = k_j + 1$ elements. Now calculate $w_j^N = \frac{w_j}{\sum_x w_x}$ for every vector $w_j$, the normalisation of $w_j$.

4. Multiply the first $k_j$ weights in $w_j$ by the $(k_j)$th stock returns selected, at the same time multiply the $(k_j + 1)$th weight by the bond return. Now sum all these products together to produce the fund return time series for fund $j$.

5. Repeat steps one to four for all $j = 1, \ldots, N_F$ funds.

The above algorithm generates $N_F$ funds each with a random holding in equity and a smaller random holding in bonds. The last time series we needed was the CSV of the stocks which we calculated using the formula:

$$X_t = \sqrt{\sum_{i=1}^{N_S} \frac{1}{N_S} \left( r_{it} - \frac{1}{N_S} \sum_{i=1}^{N_S} r_{it} \right)^2} \tag{2.1}$$

where $r_{it}$ is the return of stock $i$ for the time period associated with $t$, $N_S$ is the number of stocks in the market. $X_t$ is uniformly weighted CSV. We then computed normalised inverse CSV defined as follows

$$X_t^{\perp} = \frac{\frac{1}{X_t}}{\sum_t \frac{1}{X_t}}.$$

**Remark 2.1** (Least Squares Regressions)**.** *This remark details the first principle mathematics of OLS and WLS regressions.*

***Linear least squares*** *For this model we assume that the equation which ties the explanatory variable (x) to the response (y) is linear. We thus have an equation*

$$y = c_1 + c_2 x + \epsilon, \epsilon \sim N(0, \sigma^2) \tag{2.2}$$

*where* $c_1$ *and* $c_2$ *are coefficients and* $\epsilon$ *is residual error. We wish to solve for* $c_1$ *and* $c_2$ *thus we can set up an equation* $S$*, a system of* $n$ *simulataneous linear equations of two unknowns. Note that the system is overdetermined if* $n$ *is greater than the number of unknowns. We write*

$$S = \sum_{i=1}^{n} (y_i - (c_1 + c_2 x_i))^2 \tag{2.3}$$

*The least squares fitting process minimises the summed square of the residuals. The coefficients are determined by differentiating* $S$ *with respect to each parameter, and setting the result equal to zero. Mathematically,*

$$\frac{\partial S}{\partial c_1} = -2 \sum_{i=1}^{n} (y_i - (c_1 + c_2 x_i)) = 0$$
$$\frac{\partial S}{\partial c_2} = -2 \sum_{i=1}^{n} x_i (y_i - (c_1 + c_2 x_i)) = 0$$

*Often the estimates of the true parameters are represented by* $\alpha$ *and* $\beta$*. Substituting* $\alpha$ *and* $\beta$ *for* $c_1$ *and* $c_2$ *respectively the above equation become*

$$\sum_{i=1}^{n} (y_i - (\alpha + \beta x_i)) = 0$$
$$\sum_{i=1}^{n} x_i (y_i - (\alpha + \beta x_i)) = 0$$

*Suppressing the summation bounds while distributing the summation we have*

$$n\alpha + \beta \sum x_i = \sum y_i$$
$$\alpha \sum x_i + \beta \sum x_i^2 = \sum x_i y_i$$

*Solving for* $\beta$ *and* $\alpha$*, in terms of* $\beta$*, we have*

$$\beta = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sigma_{xy}}{\sigma_x^2} \tag{2.4}$$
$$\alpha = \frac{1}{n} \left( \sum y_i - \beta \sum x_i \right) = \bar{y} - \beta \bar{x} \tag{2.5}$$

*where $\bar{x}$ and $\bar{y}$ are means, $\sigma_{xy}$ is the covariance between $x$ and $y$ and $\sigma_x^2$ is the variance of $x$.*

$$\bar{x} = \frac{1}{n}\sum x_i \quad \bar{y} = \frac{1}{n}\sum y_i.$$

***Weighted least squares*** *Once again we use a linear model identical to Equation (2.2). The first change we note are the weights $w_i, i = 1, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$. Similarly to Equation (2.3) we seek to minimise an Equation now weighted by $w_i$.*

$$S_w = \sum_{i=1}^{n} w_i \left(y_i - (c_1 + c_2 x_i)\right)^2 \tag{2.6}$$

*Differentiating Equation (2.6) with respect to $c_1$ and $c_2$, and substituting $\alpha$ and $\beta$ for $c_1$ and $c_2$ respectively we obtain the equations*

$$\sum_{i=1}^{n} w_i(y_i - (\alpha + \beta x_i)) = 0$$
$$\sum_{i=1}^{n} w_i x_i(y_i - (\alpha + \beta x_i)) = 0$$

*Suppressing the summation bounds while distributing the summation we have*

$$\alpha \sum w_i + \beta \sum w_i x_i = \sum w_i y_i$$
$$\alpha \sum w_i x_i + \beta \sum w_i x_i^2 = \sum w_i x_i y_i$$

*Solving for $\beta$ and $\alpha$, in terms of $\beta$, we have*

$$\beta = \frac{\sum w_i(x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum w_i(x_i - \bar{x}_w)^2} = \frac{\sigma_{xy}^{(w)}}{\sigma_x^{2(w)}}, \tag{2.7}$$

$$\alpha = \bar{y}_w - \beta \bar{x}_w \tag{2.8}$$

*where $\bar{x}_w$ and $\bar{y}_w$ are weighted means, $\sigma_{xy}^{(w)}$ is the weighted covariance between $x$ and $y$ and $\sigma_x^{2(w)}$ is the weighted variance of $x$*

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} \quad \bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}.$$

We can now calculate alpha for both the OLS and WLS regressions. The regression equations are given as

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_{i,t}^2), \text{(OLS)} \tag{2.9}$$

and

$$\sqrt{w_t} r_{i,t} = \sqrt{w_t}(\alpha_i' + \beta_i' r_{m,t} + \epsilon_{i,t}), \quad \epsilon_{i,t} \sim N(0, \sigma_{i,t}^2), \text{(WLS)} \tag{2.10}$$

where

$$\begin{aligned} r_{i,t} &= \text{ return of fund } i \text{ for time period } t, \\ \alpha_i, \alpha_i' &= \text{ abnormal return of fund } i, \\ \beta_i, \beta_i' &= \text{ sensitivity of fund } i \text{ to changes in the market,} \\ r_{m,t} &= \text{ return of the market portfolio or benchmark,} \end{aligned}$$

and

$$\begin{aligned} \sqrt{w_t} = & \text{ normalised inverse CSV with respect to} \\ & \text{ the time period associated with } t. \end{aligned}$$

While considering Equations (2.5) and (2.4) in Remark 2.1, re-write beta as a function of the correlation between $x$ and $y$ as follows

$$\begin{aligned} \beta &= \frac{\sigma_{xy}}{\sigma_x^2} \quad \left(\text{note: } \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}\right) \\ &= \rho_{xy} \frac{\sigma_y}{\sigma_x}. \end{aligned} \tag{2.11}$$

Defining $\bar{x}$ and $\bar{y}$ as the means of $x$ and $y$ respectively, $\sigma_{xy}$ as the covariance between $x$ and $y$, $\sigma_x^2$ and $\sigma_y^2$ the variances of $x$ and $y$ respectively and $\rho_{xy}$ the correlation between $x$ and $y$. This equation helps us explain the relationship between beta and correlation. Fixing $\frac{\sigma_y}{\sigma_x}$ we see that there is a direct relationship between correlation and beta. Substituting Equation (2.11) for beta into Equation (2.5):

$$\alpha = \bar{y} - \rho_{xy} \frac{\sigma_y}{\sigma_x} \bar{x}. \tag{2.12}$$

Ceteris paribus we note the effect of a change in correlation on OLS alpha in Table 2.2

**Tab. 2.2:** Table demonstrating the effect of correlation on alpha

| Correlation | Effect on Alpha |
|---|---|
| $\rho_{xy} \to -1$ | $\alpha \to \bar{y} + \frac{\sigma_y}{\sigma_x} \bar{x}$ |
| $\rho_{xy} = 0$ | $\alpha = \bar{y}$ |
| $\rho_{xy} \to 1$ | $\alpha \to \bar{y} - \frac{\sigma_y}{\sigma_x} \bar{x}$ |

These effects of correlation on alpha are mimicked in the weighted least squares case.

We still haven't considered what happens to alpha when CSV is taken into account. We expected that weighting the regression by CSV would impact the value of alpha and for some funds, cause a change in their alpha ranking relative to other funds. We chose to analyse the effect of CSV on alpha numerically by computing alpha in three different cases. The next section of this chapter deals with the analysis and interpretation of various simulated market alphas.

## 2.2 Analysis of Simulated Data

Of all the simulated data, alpha is our primary focus in this section. Before we look at alpha we need to consider the effect of the choice of the drifts and volatilities in the stock price generator as these contribute to larger (or smaller) CSV. In these simulations we considered fixed and random volatilities and drifts. The method for choosing the random drifts and volatilities laid out in Table 2.1 implies that we have fixed the range from which the drifts and volatilities are chosen. The underlying assumption being that we can only observe stocks with drifts between $a_\mu$ and $a_\mu + b_\mu$ and volatilities between $a_\sigma$ and $a_\sigma + b_\sigma$. To further cement the idea that drift and volatility can generate CSV we create two hypothetical (extreme) stocks, $H$ and $L$. Stock $H$ has a drift of $a_\mu + b_\mu$ (the maximum possible) and stock $L$ a drift of $a_\mu$ (the minimum possible) if their volatilities are the same, their drifts will generate a return opportunity set which will create CSV. Alternatively we could generate $H$ and $L$ with the same drifts and different volatilities. This time round their volatilities would create a return opportunity set, producing CSV. As a closing point to this example, it is logical to assume that any mixture (as long as the stocks aren't the same) of drifts and volatilities to create different stocks would create multiple return opportunities and hence CSV.

We now consider three market scenarios: Perfectly correlated stocks with the same drifts and volatilities, Randomly correlated stocks with the same drifts and volatilities and Randomly correlated stocks with different drifts and volatilities.

#### Case 1: Perfectly correlated stocks with the same drifts and volatilities

In a market where stocks are perfectly correlated and have the same drift and volatilities we expect all stocks to react the same way to any new information. If a shock occurs in the market all the stocks will trend in the same direction. In this case all stocks are identical to one another (see Figure A.1 in Appendix A). An investor would be indifferent when stock picking. It is not possible for any stock to out-perform the benchmark. Essentially, the benchmark is identical to any of the stocks. Like the benchmark, any funds constructed from this universe of stocks will be identical to any of the individual stocks. In all the simulations CSV was calculated using the stock returns. Because there is no stock return opportunity set CSV should theoretically be zero. CSV is very close to zero because we cannot decompose a perfectly correlated martix using Cholesky decomposition, however the matrix we have decomposed is slightly offset from the perfect correlation matrix. When comparing the adjusted alpha to the unadjusted alpha in Figure 2.1 we saw that there was no difference between them. We noted that adjusting the regression by CSV did not change alpha and that none of the funds ranks changed. In Figure 2.2 we observe positively skewed alpha, due to the way in which the stocks and funds were constructed. It's also clear that the unadjusted alpha and adjusted alpha histogram are identical, a feature of the perfectly correlated market.

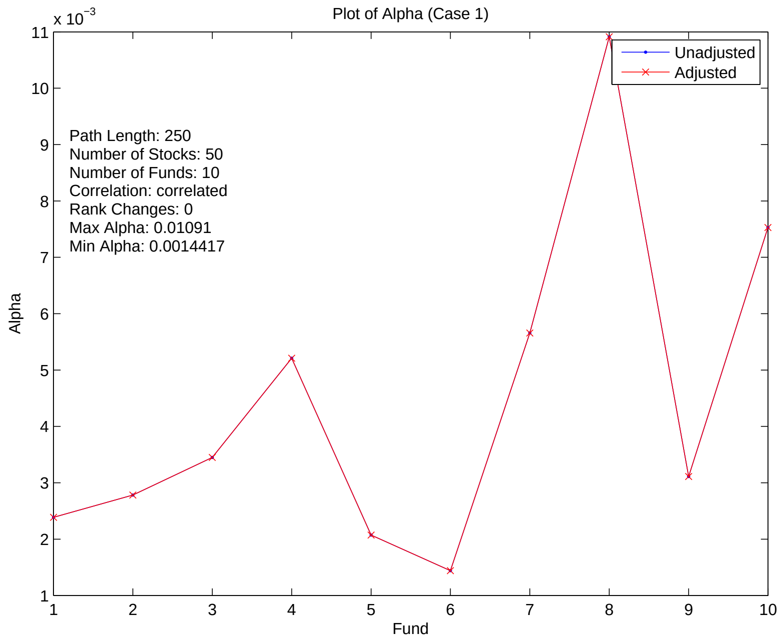


**Fig. 2.1:** Alpha computed from a market with perfectly positively correlated stocks with the same drifts and volatilities

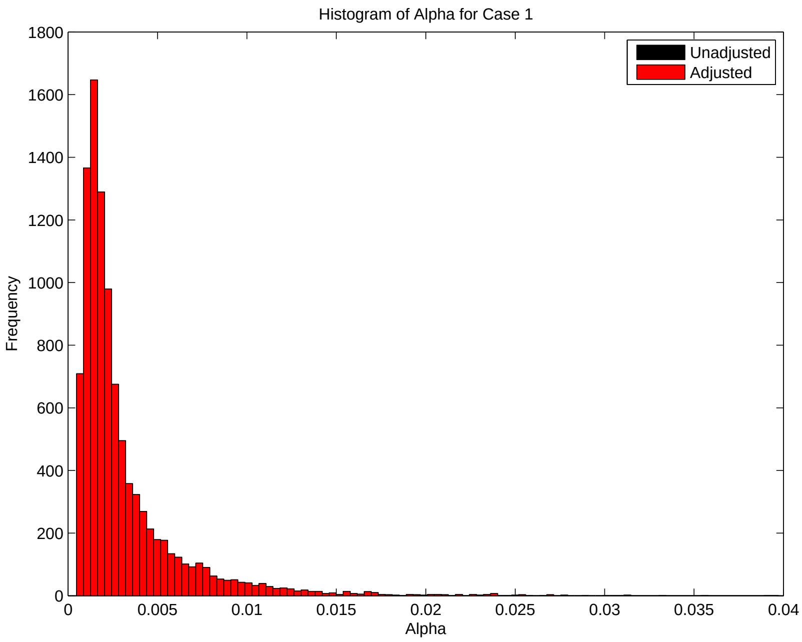


**Fig. 2.2:** Alpha histogram computed from a market with perfectly positively correlated stocks with the same drifts and volatilities

**Case 2: Randomly correlated stocks with the same drifts and volatilities**

The only difference between this case and the first was the choice to use randomly correlated stocks (see Figure A.2 in Appendix A), and hence we retain the drifts and volatilities from the first case. If this is the new scenario then the change to a randomly correlated stock universe should produce CSV and cause a difference between adjusted and unadjusted alpha. In this market, under- and out-performance are observable however CSV is generated strictly by the correlation of the stocks. Thus the funds constructed from this universe of stocks can out-perform the benchmark and achieve a larger positive alpha (unlike in the first case). It is also possible for funds to match the benchmarks performance or even under-perform. Changing the correlation introduced a wider return opportunity set into the market. The wider return set induced a larger CSV. When we studied the alphas in Figure 2.3 we noted that there were several value-wise changes in alpha as well as fund rank changes. An interpretation of rank changes, says that several funds actually have better (or worse) alpha than the OLS regression originally suggested. Figure 2.4 shows alpha is symmetrically dispersed around zero and that adjusted alpha is less dispersed than unadjusted alpha.

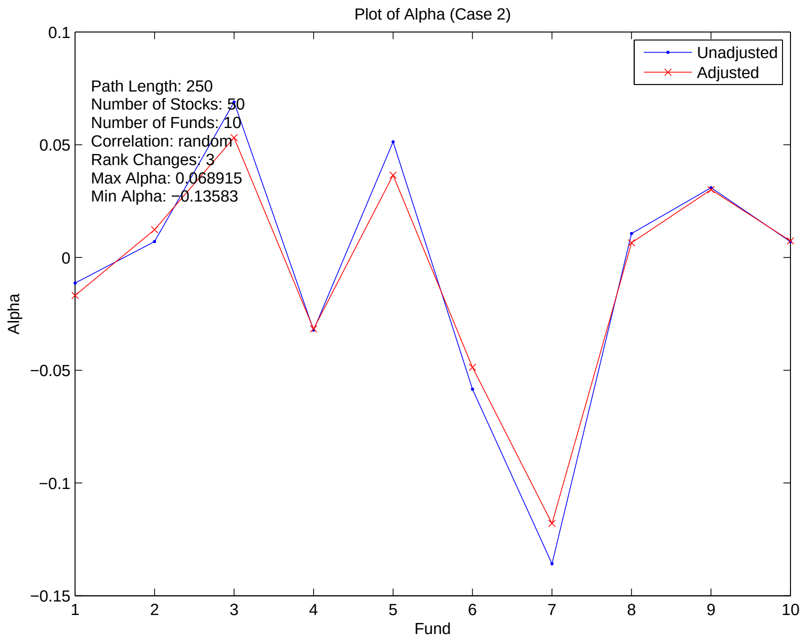


**Fig. 2.3:** Alpha computed from a market with randomly correlated stocks with the same drifts and volatilities

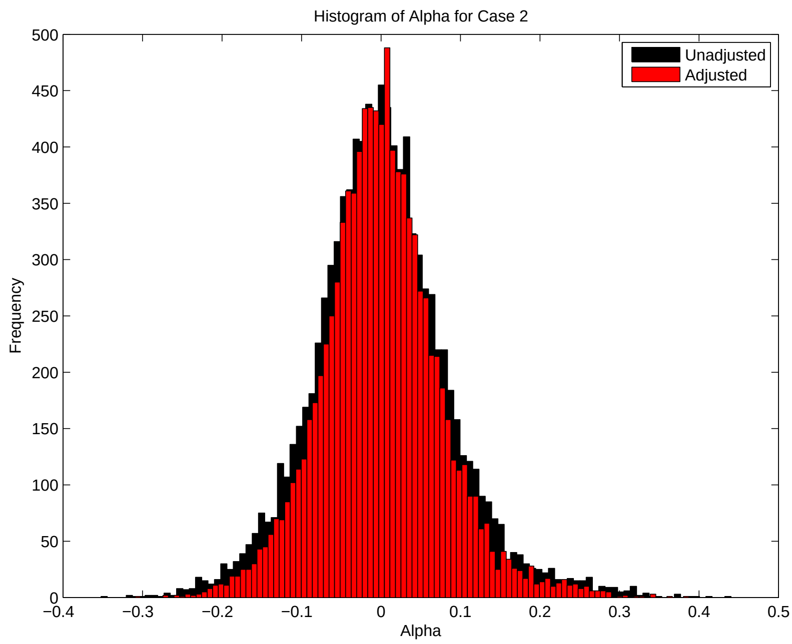


**Fig. 2.4:** Alpha histogram computed from a market with randomly correlated stocks with the same drifts and volatilities

Both case one and two are poor representations of reality, these examples helped us demonstrate the effect of correlation on CSV and alpha. In any real world market there exists correlations as well as varying drifts and volatilities. We now simulate a market scenario with randomly correlated stocks that have different drifts and volatilities (see Figure A.3 in Appendix A).

### Case 3: Randomly correlated stocks with different drifts and volatilities

This scenario is as close as our model gets to a real market however all the simplifying assumptions move us further away from reality. We have retained the correlation matrix from the previous case. Letting the drifts and volatilities vary per stock introduced additional CSV. Benchmark out performance, under performance and no performance were all possibilities in this case (like in case 2). We observed a larger opportunity set of stock returns because of the varied drifts and volatilities on top of the random correlation structure. Figure 2.6 demonstrates that alpha is symmetrically dispersed around zero and that adjusted alpha is less dispersed than unadjusted alpha. The tails of the unadjusted alpha are longer than the tails of adjusted alpha.

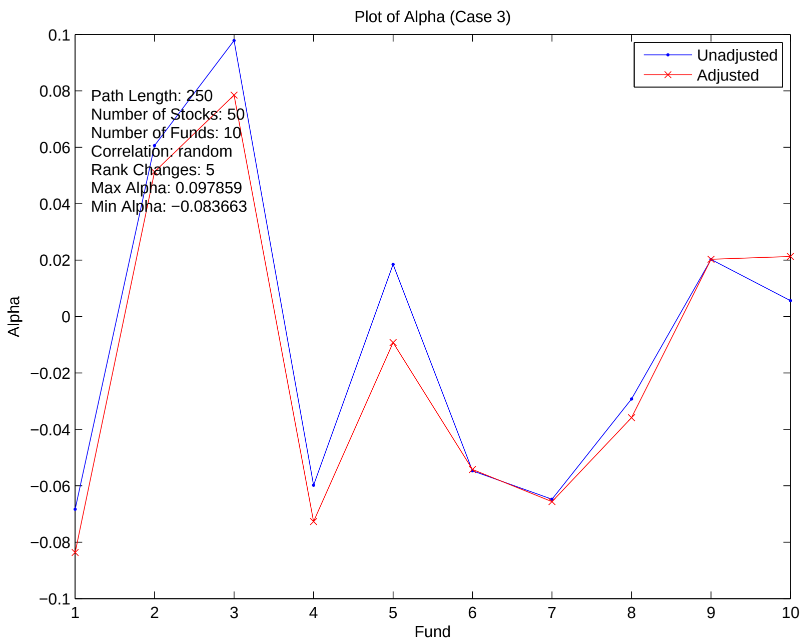


**Fig. 2.5:** Alpha computed from a market with randomly correlated stocks with different drifts and volatilities

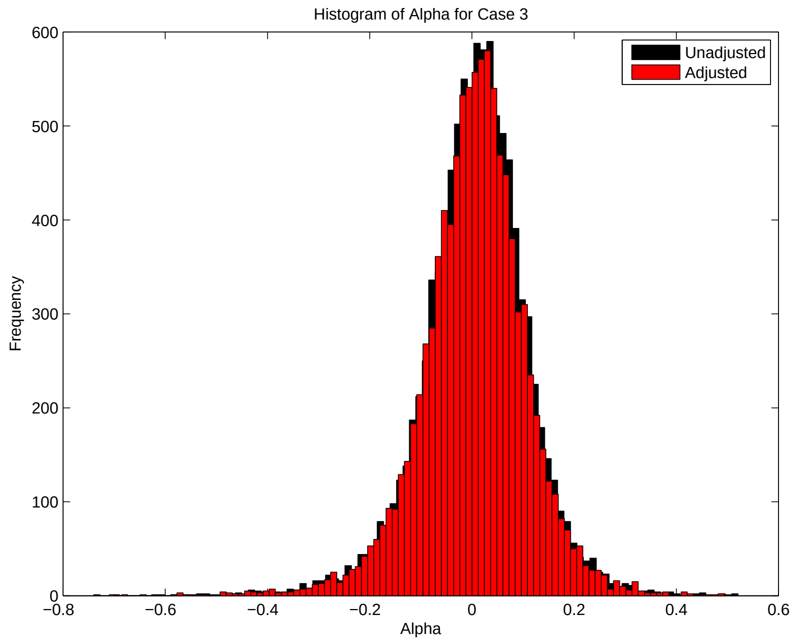


**Fig. 2.6:** Alpha histogram computed from a market with randomly correlated stocks with different drifts and volatilities

To explain why the adjusted and unadjusted alpha histograms are so similar in Figure 2.6 and less so in Figure 2.4 consider the following: For the unadjusted alpha, all of the weights are $1/250 = 0.004$. So the more concentrated the weights are around that number, the closer the adjusted alpha will be to the unadjusted one. Comparing Figure 2.7 to Figure 2.8 it is clear why case three is worse than case two in terms of dispersion  the weights are definitely centered at 0.004 in case three, while they're centered away from 0.004 in case two. This is totally an artifact of the random market we have simulated. Of course other simulated markets would have different weight distributions.

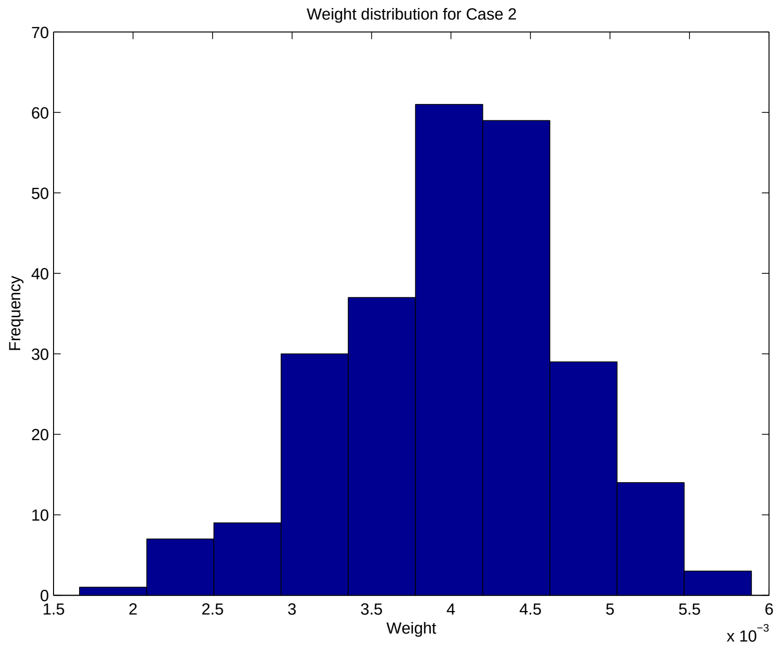


**Fig. 2.7:** Weight distribution computed from a market with randomly correlated stocks with the same drifts and volatilities

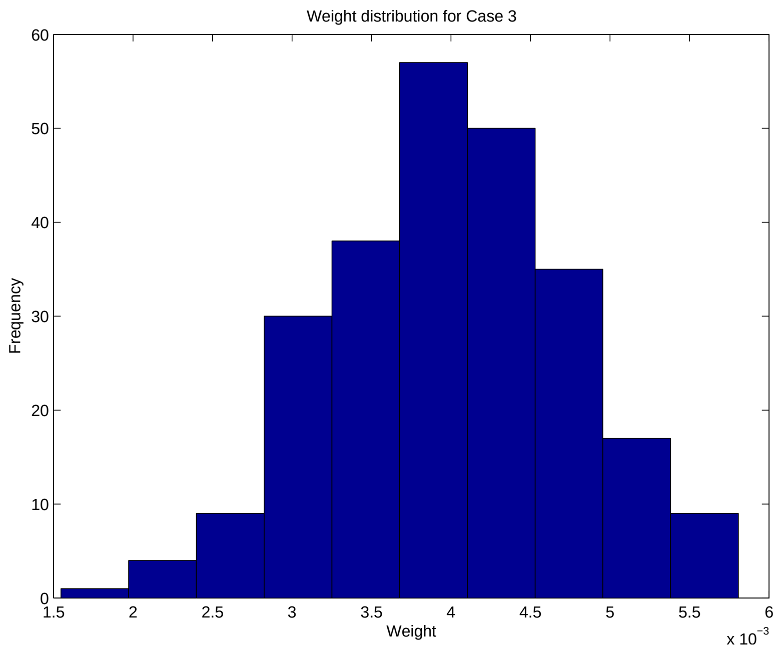


**Fig. 2.8:** Weight distribution computed from a market with randomly correlated stocks with different drifts and volatilities

### Perfect fund simulation using Case 3 data

In the last example of this section we demonstrate the effect of CSV on a perfect fund compared to the effect of CSV on non-perfect funds. Firstly, we defined a perfect fund as one which beats the benchmark at every point in time. In order for a fund to beat its benchmark consistently the manager must have the highest level of skill possible. The manager must also actively trade continuously. Conversely, non-perfect funds experience random return paths which can move above and below the benchmark. In this setting non-perfect funds are actively managed. We then generated 1000 non-perfect funds and chose those funds whose unadjusted alphas' were closest to the perfect fund's alpha (0.5% different from the perfect alpha). We then inspected the price path plots to check whether the alpha constraint detected similar funds correctly. It is clear from Figure A.4 that the price path of the perfect fund dominates the benchmark at every point in time. It is also evident that the similar funds have paths which move in tandem with the perfect fund. From the price plot we observed that one fund behaved very differently however its alpha was still close to the perfect alpha. Computing alphas for both the perfect fund and similar funds we obtained Figure 2.9. In the plot fund one is the perfect fund and the rest are non-perfect funds. The perfect fund's alpha did not change in this example however all the other non-perfect funds experienced an alpha change when moving from unadjusted to adjusted alpha.

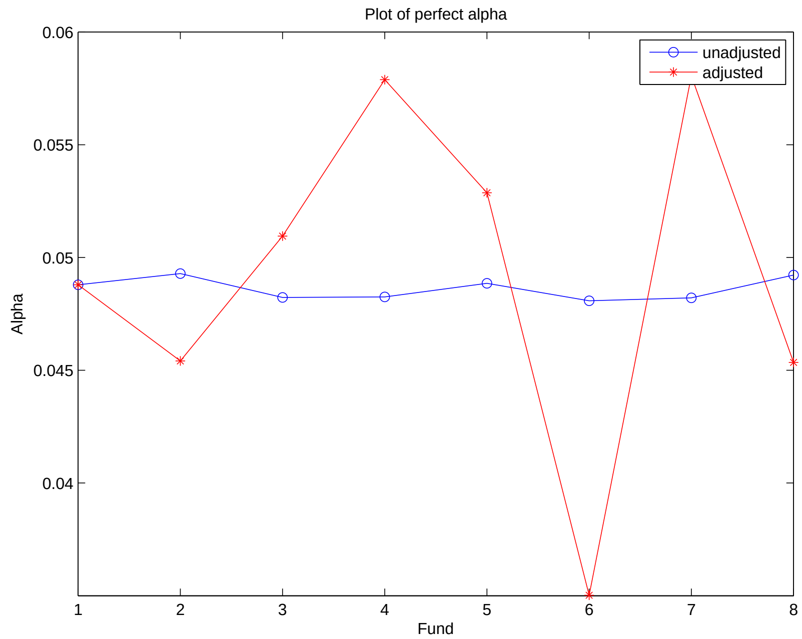


**Fig. 2.9:** Plot of perfect alpha (fund 1) in comparison to several similar non-perfect funds. Alphas are annualised.

In a real world situation it is highly unlikely to observe such a perfect fund. Thus we can only observe funds' whose alphas either increase or decrease after adjusting for CSV. We know that no change in alpha after adjusting for CSV indicates the perfect fund implying that alphas which change the least after adjusting for CSV should represent the best funds. Note that if fund $A$ experienced a change in alpha of 0.1% and fund $B$ experienced a change in alpha of $-0.1\%$ it is unclear which fund is better although choosing a fund which consistently obtains alpha across different CSV environments would be ideal.

**Tab. 2.3:** Table of detailed regression data from perfect fund example

| Table of regression variables | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| $\alpha$ | 0.0488 | 0.0493 | 0.0482 | 0.0482 | 0.0489 | 0.0481 | 0.0482 | 0.0492 |
| $\alpha_w$ | 0.0488 | 0.0454 | 0.0510 | 0.0579 | 0.0529 | 0.0350 | 0.0581 | 0.0453 |
| $\beta$ | 1.0000 | 0.6226 | 0.6688 | -0.3200 | 0.7797 | 0.5349 | 1.1666 | 0.8610 |
| $\beta_w$ | 1.0000 | 0.6279 | 0.6569 | -0.2618 | 0.7843 | 0.5711 | 1.1588 | 0.8737 |
| $\bar{x}$ | 0.1186 | 0.1186 | 0.1186 | 0.1186 | 0.1186 | 0.1186 | 0.1186 | 0.1186 |
| $\bar{x}_w$ | 0.1079 | 0.1079 | 0.1079 | 0.1079 | 0.1079 | 0.1079 | 0.1079 | 0.1079 |
| $\bar{y}$ | 0.1674 | 0.1232 | 0.1276 | 0.0103 | 0.1413 | 0.1115 | 0.1866 | 0.1514 |
| $\bar{y}_w$ | 0.1567 | 0.1132 | 0.1218 | 0.0296 | 0.1375 | 0.0966 | 0.1831 | 0.1396 |
| $\sigma_x$ | 0.0524 | 0.0524 | 0.0524 | 0.0524 | 0.0524 | 0.0524 | 0.0524 | 0.0524 |
| $\sigma_x^{(w)}$ | 0.0031 | 0.0031 | 0.0031 | 0.0031 | 0.0031 | 0.0031 | 0.0031 | 0.0031 |
| $\sigma_y$ | 0.0524 | 0.0560 | 0.0677 | 0.0677 | 0.0719 | 0.0559 | 0.0827 | 0.0750 |
| $\sigma_y^{(w)}$ | 0.0031 | 0.0035 | 0.0041 | 0.0041 | 0.0044 | 0.0035 | 0.0050 | 0.0046 |
| $\sigma_{xy}$ | 0.6854 | 0.4267 | 0.4584 | -0.2193 | 0.5343 | 0.3666 | 0.7995 | 0.5901 |
| $\sigma_{xy}^{(w)}$ | 0.0024 | 0.0015 | 0.0016 | -0.0006 | 0.0019 | 0.0014 | 0.0028 | 0.0021 |
| $\rho_{xy}$ | 1.0000 | 0.5816 | 0.5169 | -0.2475 | 0.5679 | 0.5011 | 0.7386 | 0.6009 |
| $\rho_{xy}^{(w)}$ | 1.0000 | 0.5545 | 0.4943 | -0.1970 | 0.5526 | 0.4994 | 0.7190 | 0.5900 |

Taking the perfect example further we calculated annualised values for all the regression variables used in the perfect alpha example. Column one of Table 2.3 details the regression of the perfect fund and columns two to eight represent regression results of the non-perfect funds. The correlation of the perfect fund to the benchmark is one because the perfect fund's path is calculated using the benchmark's path as a reference. We used Table 2.3 and Equation (2.12) to discern which variables may have caused alpha to change. Because all the variables responded to the WLS regression it was not clear that any particular variable contributed significantly to a change in alpha.

## 2.3 Simulation findings

The purpose of this Chapter was to expose any uncertainties surrounding alpha and CSV while working in a controlled system. What happens to alpha when all the stocks are perfectly correlated? What does CSV do when we vary the correlation between stocks? How does CSV affect alpha in a system where all the parameters are known? We answered these questions by designing an idealistic market with a simple model of stocks and stock picking. It was clear from the plots examined in Section 2.2 that alpha was changed by WLS regressions. It was also evident that correlation affected the magnitude of the change in alpha. We saw no noticeable change in alpha in the market with perfect correlation, when comparing OLS alpha to WLS alpha. Correlation affects the closeness of stock returns, which in turn shrinks the size of the stock return opportunity set. CSV used in this idealistic model was based off of stock returns and represents the opportunity set of these returns. We can therefore state that the higher the correlation, the lower the CSV. If all stocks are highly correlated then no combination of stocks will out-perform any other, including the benchmark (a weighted combination of all the stocks). If this is true then none of the funds constructed from this idealistic stock universe will achieve alpha when correlation is high. We also established that CSV leads to alpha. If CSV is very close to zero then alpha becomes harder to obtain as there is a smaller return opportunity set. Funds which obtain higher alpha in a low CSV environment show a higher level of managerial skill than those with lower alpha. The same can be said for a high CSV environment although it is easier to obtain alpha when CSV is high. WLS regressions will decrease or increase alpha. This depends on the CSV environment which the fund exists in. If there is no change in alpha the fund is perfect.

In the Chapters which follow we analyse and test real world data.

# Chapter 3

# Framework for Analysis and Data Collection

In this chapter we build a framework for fund data selection. We then select the fund data on the basis of this framework and end the chapter by describing the difficulties experienced with this data.

## 3.1 Cross-sectional Volatility in South Africa

The history of cross-sectional volatility (CSV) in South Africa is a short albeit an interesting one. CSV is a new concept thus many fund managers are still conducting research on the subject. The CSV data used in this research was captured between July 2002 and July 2012, for the ALSI, Top40, Resi20, Indi20, Fini25, MidCap, and SmallCap indices. In Figure 3.1 we consider setting $w_{it} = \frac{1}{N_t}$ in Equation (1.3), or equivalently uniformly weighting CSV[1]. We make the following observations: The Resi20 is the most volatile of the indices with spikes between July 2002 and July 2008, after which it settles post the 2008 crash, picking up again in the past year. The Resi20 also had the highest observed CSV at 25.96% in 2008. The Fini15 featured the lowest CSV for the period under observation, with a minimum of 1.47% in 2011. Looking at all the indices we find that CSV was highly volatile before 2008. From 2008 to early 2012, we see that CSV declined across all sectors and rose again in 2013.

[1] The values plotted are the square-root of Equation (1.3).

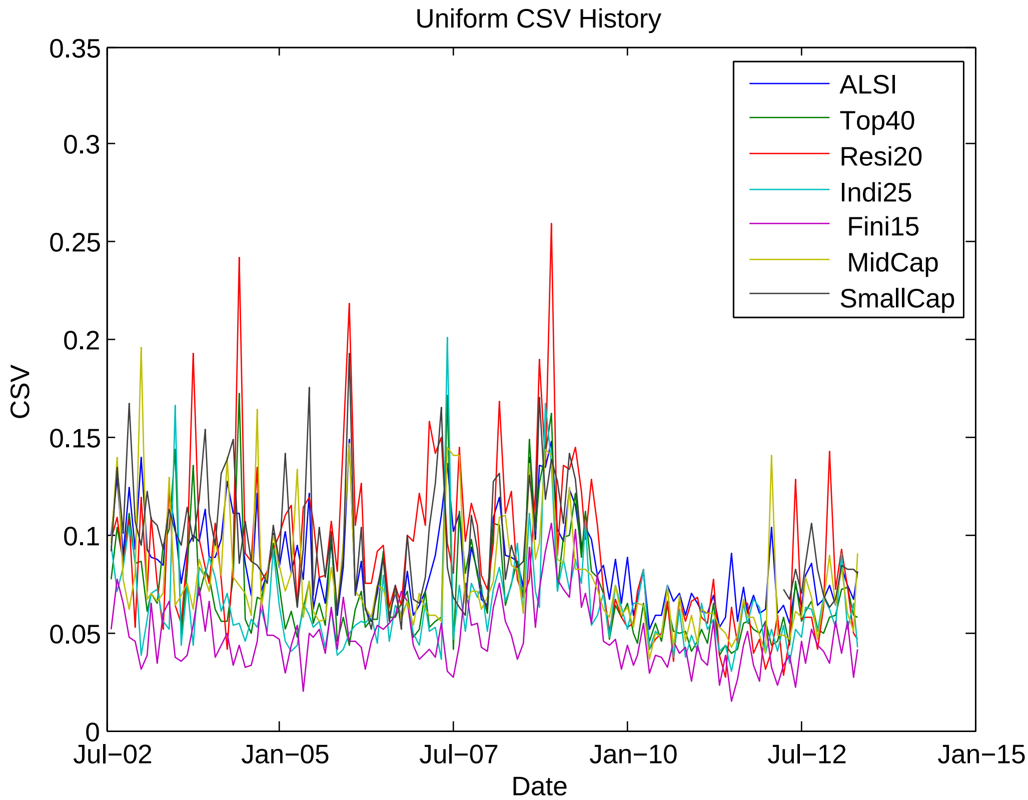


**Fig. 3.1:** Uniformly weighted CSV

Following the analysis of uniform CSV, we turned to market capitalisation weighted CSV visible in Figure 3.2. We see similar trends to the uniform weighting basis. Fini15 still produces the lowest CSV results on average throughout the period, with a minimum of 1.24% in 2011. Resi20 has smaller CSV values due to its proportionately larger market capitalisation weightings. SmallCap dominates with the largest average CSV, with a maximum of 21.74% (due to its proportionately smaller market capitalisation). The range of CSV values throughout the observation window is smaller when taking into account market capitalisation. We still see a trend of high variability in years prior 2008, and decreased variability post 2008, similar to the uniform case.

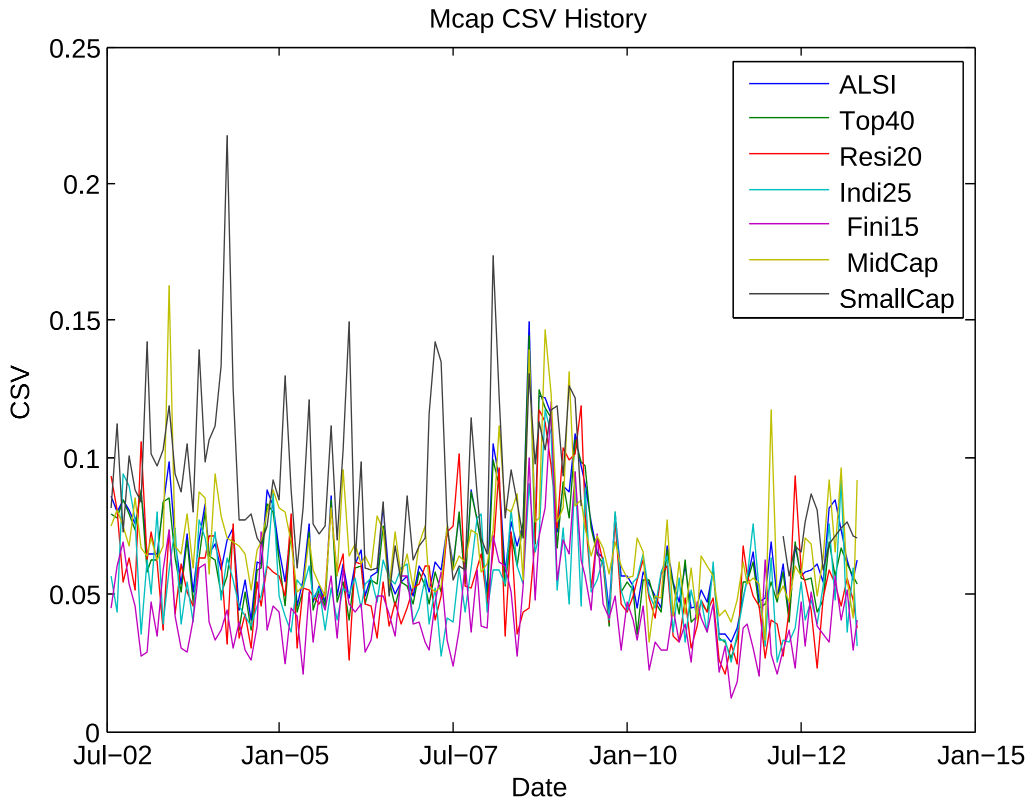


**Fig. 3.2:** Market capitalisation weighted CSV

Further interest revolves around the global decline of CSV experienced in 2008. Every sector was effected by the crash, all featuring a decline in share value simultaneously. When share prices dropped, all of them dropped over the same time window thus decreasing the CSV (the dispersion between stocks). This behaviour continued till late 2011. Between January 2012 and April 2013, CSV has increased because some sectors have recovered, while others are still recovering.

Finally, having visually assessed figures 3.1 and 3.2 we were able to make a tentative decision on the appropriate time windows on which to perform the analysis. The primary factor behind our choice of analysis periods was the behaviour of CSV between 2002 and 2013. Both periods chosen exhibited stable CSV over time. Hence, the inability to use the period between January 2008 and December 2010 where CSV was highly negatively correlated over time. The time windows we chose were 1 January 2006 to 31 December 2007 (2 years), referred to as Time Window 1, and 1 January 2011 to 31 December 2012 (2 years), referred to as Time Window 2.

## 3.2 Data description

For this research we acquired monthly fund price data for 16 retail funds, of which only six were usable. The funds were required to have a low cash and a majority equity holding. The availability of data and the choice of Time Window 1 and Time Window 2 forced us to work with only six of the 16 retail funds (see Table 3.1 for the chosen list of retail funds). The low cash, high equity holding requirement ensured the presence of dispersion between funds. As an example, consider two funds with large cash holdings (40%) and calculate their dispersion. This dispersion is far smaller compared to funds with majority equity holdings. One further driver for the decision to work with low cash funds presents itself in the paper by de Silva *et al.* (2001) which states that a link exists between fund return dispersion and security returns, not cash.

**Tab. 3.1:** Usable retail funds

| Retail Funds | |
|---|---|
| Fund Name | Bloomberg Code |
| Allan Gray Equity Fund | ALEQTYF |
| Coronation Top 20 Fund | CORTP20 |
| Investec Active Quants | INVINDX |
| Kagiso Top 40 Tracker Fund | CORALSI |
| Momentum Equity Fund | RMBEQTY |
| Prudential Equity Fund | PRUOPTM |

In addition to the retail funds, Alexander Forbes provided us with a data set of historical institutional fund returns. From the data set we chose 12 funds (see Table 3.2) which matched the criteria previously stated. For each fund, retail and institutional we collected a nominated benchmark[2] returns.

[2] A nominated benchmark is the benchmark contractually assigned to the fund when it is created.

**Tab. 3.2:** Usable institutional funds

| Institutional funds |
|---|
| Cadiz Equity SWIX |
| Coronation Core Equity |
| Investec Active Quants |
| Investment Solutions Pure Equity |
| Prudential Core Equity ALSI |
| Prudential Core Equity SWIX |
| Stanlib Core Equity |
| Investec Equity |
| ABSA Asset Management Core Equity |
| Fraters Equity |
| Coronation Houseview Equity |
| Prescient Equity Quant Fund |

For the purpose of further analysis we gathered monthly price data for sector indices i.e. the TOP40, RESI20, INDI25, FINI15, STEFI and ALBIP defined in Table B.1 (used in the calculation of a proxy benchmark, see Section 4.5). The final data set attained was monthly ALSI CSV data, both uniform and market capitalisation weighted (see Chapter 3).

## 3.3 Data difficulties

The first problem we encountered was that of missing data. All funds with missing data were excluded from the analysis. The exclusion of funds with missing data was a necessary step to ensure the regressions were calculable.

The institutional data provided by Alexander Forbes presented us with an anomaly of positive alpha across all the funds. Through the regressions in Section 4.3 we found that no matter the time window, or fund, we always received a positive alpha figure. In Chapter 4 we consider two ideas to correct for this anomaly: Firstly, the proxy benchmark, i.e. solving a tracking error minimisation problem (see Section 4.5) and the zero-sum method (see Section 4.4). Thus, the alpha estimates in this research are not a pure representation of the Alexander Forbes institutional data.

The final issue worth noting, is selection bias. As mentioned earlier in this chapter, we chose 16 retail funds, however 10 of them did not have data in the chosen time windows. Consequently, we were left with six retail funds. Similarly for the institutional funds, we were left with 12 funds. We created a selection bias by excluding funds which did not have data in the time period over which we performed the analysis.

# Chapter 4

# Methodology

In this chapter, we explain the method we followed for alpha estimation, along with a test for heteroskedasticity. We also explain how alpha was re-estimated using a cross-sectional volatility (CSV) adjustment. We close off this chapter by discussing two new concepts, namely the zero-sum alpha and the proxy benchmark.

## 4.1 Why test for heteroskedasticity?

We test for the presence of heteroskedasticity because it biases the variance of estimates, therefore regressions between predictors (the benchmark return) and outcomes (the fund return) will yield biased relationships. As an example, the p-value for testing whether the slope term (beta) is significantly different from zero can be misinterpreted when the data are heteroskedastic.

In the exploratory analysis of the return data we investigated scatter plots of benchmark return versus fund return. Informally we determined whether heteroskedasticity was present in the data by studying these scatter plots for wedge-like shapes. These wedge formations indicate increasing variance between predictors and outcomes. Formally we tested for heteroskedasticity using the Breusch-Pagan hypothesis test clearly defined by Breusch and Pagan (1979). In linear regression, using the model $Y_i = \theta_1 + \theta_2 X_i + \epsilon_i$, assuming $\epsilon_i \sim N(0, \sigma_i^2)$. Here $Y_i$ is the return vector of fund $i$, $X_i$ is the nominated benchmark return vector of fund $i$, $\epsilon_i$ is the residual error vector for fund $i$ and $\theta_1$ and $\theta_2$ are the intercept and slope of linear regression respectively. Perform the following steps:

1. Estimate the model by ordinary least squares regression (OLS) and obtain the residuals $\{\hat{\epsilon}_1, \hat{\epsilon}_2, \dots\}$.

2. Estimate the variance of these residuals i.e. $\hat{\sigma}_i^2 = \frac{\sum \epsilon_i^2}{(n-2)}$. The vector $\epsilon_i$ has $n$ elements, and $n-2$ is a degrees of freedom adjustment to $n$. Note that $\hat{\sigma}_i^2$ is a scalar quantity.

3. Run the regression $\frac{\epsilon_i^2}{\hat{\sigma}_i^2} = \kappa_1 + \kappa_2 X_i + \varepsilon_i$ and compute the explained sums of squares (ESS) from this regression.

4. Reject the hypothesis of homoskedasticity in favour of heteroskedasticity if $\frac{ESS}{2} > \chi_{\nu=1}^{2(\alpha=0.01)}$. The null and alternative hypothesis are defined as follows:

$$H_0 : \text{Homoskedasticity present}$$
$$H_1 : \text{Heteroskedasticity present}$$

The Breusch-Pagan hypothesis test is a form of statistical inference. We use it to draw conclusions about the population on data sampled from the population. The test is designed so that the null hypothesis is the least favoured outcome, however it is not necessarily rejected. We choose to reject or accept a hypothesis on the basis of evidence found in the data by comparing the test statistic to the theoretical value of that test statistic (with some level of significants). In our case we used an inequality which compares the Breusch-Pagan test statistic to a theoretical Chi-Squared statistic (see above).

## 4.2 Testing for Heteroskedasticity

Firstly, we inspected the scatter plots (Figures 4.1 and 4.2) of fund return versus benchmark return. We observed a linear relationship between fund returns and their nominated benchmarks (see Figure 4.1, CORALSI). For further insight into why this was the case, consider Remark 4.1 regarding trading strategies. Comparing Figure 4.1 to Figure 4.2 we saw decreased linearity between dependent and independent variables moving from Time Window 1 to Time Window 2. Using only the plots we could not conclude with certainty that heteroskedasticity was present in the return data. Thus we proceeded with the Breusch-Pagan Test for heteroskedasticity.

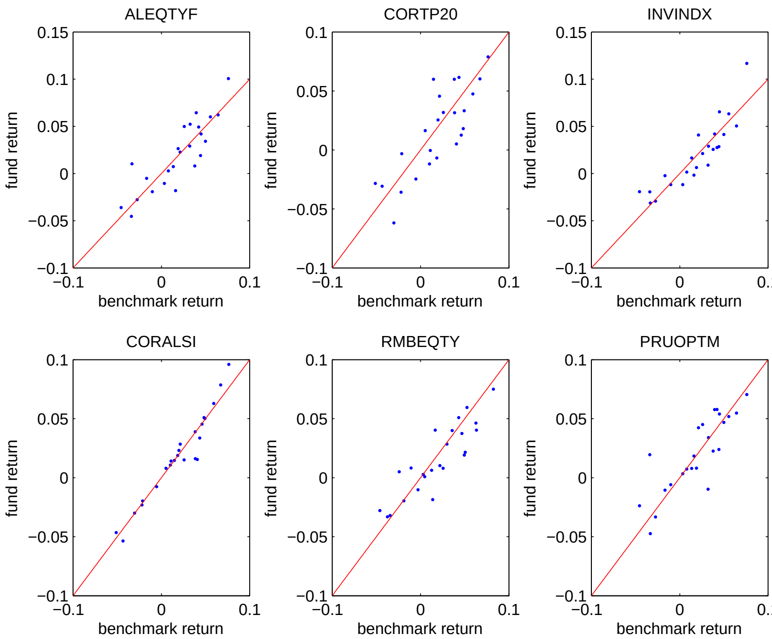


**Fig. 4.1:** Scatter plot of fund return vs. benchmark return for window 1 for retail funds, unadjusted

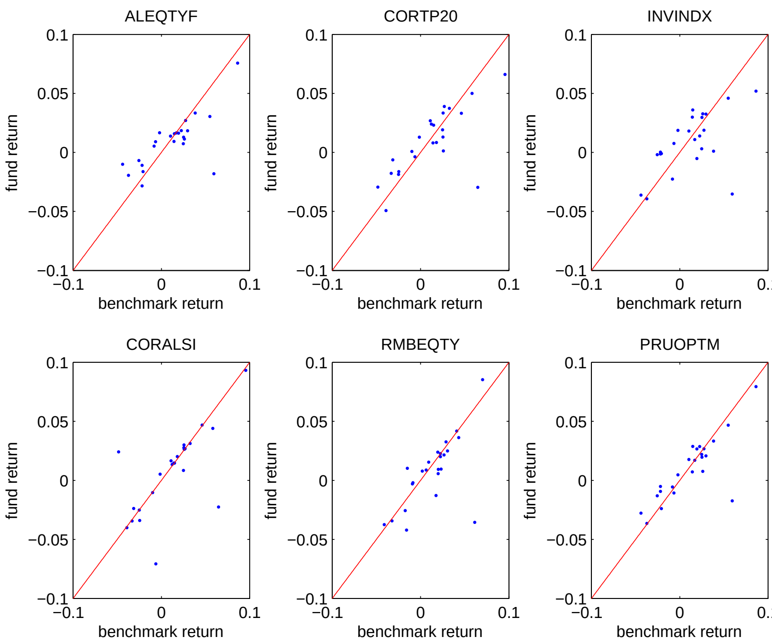


**Fig. 4.2:** Scatter plot of fund return vs. benchmark return for window 2 for retail funds, unadjusted

**Remark 4.1** (Trading Strategies)**.** *To understand why there is a strong relation between retail fund returns and their benchmarks we first need to define two types of investment strategies:*

***Active Trading*** *The primary goal of active management is benchmark out performance (positive alpha). Skill levels, active fund performance, and new information are closely monitored by all market participants making it difficult to exploit market inefficiencies (if they ever existed in the first place). In active funds we see more human involvement than in the passive style. As a result of more human oversight active fund management is a costly and more risky style of management than passive management. As a consequence of higher risks active funds are more likely to yield higher abnormal returns than passive strategies.*

***Passive Trading*** *In passive trading strategies, managers are concerned with benchmark tracking. Passive strategies are cheaper than active strategies due to a lower human participation factor. Passive funds bear a larger proportion of benchmark risk as the majority of idiosyncratic risk is diversified away. Unfortunately, the lower risk through diversification will ultimately produce lower absolute returns, however, lower management costs results in a better (relative) net effect than in the actively managed alternative. This makes sense because active fund managers experience much larger costs which they have to overcome before any profits can be realised.*

*The majority of retail funds follow a passive trading strategy and thus have a mandate to track their nominated benchmark as closely as possible creating the linear relationship we have already alluded to.*

Performing the Breusch-Pagan test on the retail funds we attained the results seen in Table 4.1.

**Tab. 4.1:** Results from the Breusch-Pagan hypothesis test for heteroskedasticity

| **Test for Heteroskedasticity Unadjusted** | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **ALEQTYF** | Reject $H_0$ | Reject $H_0$ |
| **CORTP20** | Reject $H_0$ | Reject $H_0$ |
| **INVINDX** | Reject $H_0$ | Reject $H_0$ |
| **CORALSI** | Reject $H_0$ | Fail to Reject $H_0$ |
| **RMBEQTY** | Reject $H_0$ | Reject $H_0$ |
| **PRUOPTM** | Reject $H_0$ | Reject $H_0$ |

These results provided a clearer picture regarding heteroskedasticity. Looking at Time Window 1 we noted that all funds in the window rejected $H_0$. In Time Window 2 we noted that only CORALSI failed to reject $H_0$, while all other funds in the window rejected $H_0$. A reminder again, that in this test we wanted to reject $H_0$ as this was the hypothesis of homoskedasticity. From the evidence in Table 4.1 we concluded that weighted least squares (WLS) was necessary because heteroskedasticity was present in almost all the funds tested.

## 4.3 Performance measurements and estimation

There are several ways which fund managers can evaluate their fund's performance (see Remark 4.2). In OLS regression we obtain alpha (see Remark 1.1) by regressing the return of a fund's benchmark on the fund's return.

**Remark 4.2** (Performance Measures)**.** *We have already defined alpha, thus we shall define a further four performance measurements commonly used in practice.*

***Orthogonalised Return*** *A fund return which is independent of its benchmark, i.e. moves in the benchmark are not reflected in the orthogonalised return. We observe orthogonalised returns as the error components of the OLS regression.*

***Relative Return*** *A measure of fund performance, determined by subtracting a funds benchmark from a funds return. Mathematically,*

$$RR_{Fund} = R_{Fund} - R_{Benchmark}.$$

*Active fund managers aim to maximise relative return, while passive portfolio strategies require the relative return to be as close to zero as possible. Positive relative returns imply benchmark out performance, and negative relative returns imply that a fund has under performed the benchmark. Relative return is often used to describe the skill of a fund manager, however it does not provide an indication of the risk which the manager has taken on. In this case one would consult the information ratio, a risk adjusted performance measure taking volatility into account.*

***Sharpe Ratio*** *A performance measurement meaningfully implemented when funds contain more than two securities. To compute the Sharpe ratio we use the equation*

$$S = \frac{E[R_P - R_F]}{\sigma_P}, \tag{4.1}$$

*where $R_P$ is the return of a portfolio, $R_F$ is the risk free rate of return, $\sigma_P$ is the volatility of the portfolio and $E[R_P - R_F]$ is the expected excess return of the portfolio relative to the risk free rate. Holding $\sigma_P$ constant, portfolios with higher Sharpe ratios are more attractive to investors because they provide higher excess returns for the same level of risk.*

***Information Ratio*** *Different from the Sharpe ratio, the information ratio uses the benchmark return to calculate expected excess returns. The information ratio is frequently used by analysts to gauge the skill of fund managers. Unlike relative return , the information ratio encompasses the amount of risk the manager has taken on. It is important to correctly choose the benchmark as this can misrepresent the information ratio by either over or under estimating excess returns. We calculate the information ratio using the equation*

$$IR = \frac{E[R_p - R_b]}{\sigma_p}, \tag{4.2}$$

*where $R_p$ is the return of a portfolio, $R_b$ is the benchmark return, $\sigma_p$ is the standard deviation of excess returns and $E[R_p - R_f]$ is the expected excess return of the portfolio relative to the benchmark.*

We estimate these parameters using the model,

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_{i,t}^2), \tag{4.3}$$

where

$$\begin{aligned}r_{i,t} &= \text{return of security } i \text{ for time period } t,\\ \alpha_i &= \text{abnormal return of security } i,\\ \beta_i &= \text{sensitivity of security } i \text{ to changes in the market}\end{aligned}$$

and

$$r_{m,t} = \text{return of the market portfolio.}$$

This alternative to the market model, introduced in Chapter 1, simplifies calculation and references pure returns instead of premiums above the risk free rate. Note that the returns $r_{i,t}$ are calculated on a monthly basis in a relative fashion.

Application of WLS will either completely remove the presence of heteroskedasticity or reduce its effect. The weighting we chose was the annualised normalised CSV of the Johannesburg Securities Exchange All Share Index (ALSI). We explain the choice of this particular weighting after the mathematical formulation of WLS,

$$\sqrt{w_t} r_{i,t} = \sqrt{w_t}(\bar{\alpha}_i + \bar{\beta}_i r_{m,t} + \epsilon_{i,t}), \quad \epsilon_{i,t} \sim N(0, \sigma^2_{i,t}). \tag{4.4}$$

where

$$\begin{aligned}\bar{\alpha}_i &= \text{abnormal return of security } i,\\ \bar{\beta}_i &= \text{sensitivity of security } i \text{ to changes in the market}\end{aligned}$$

and

$$\begin{aligned}\sqrt{w_t} = \ &\text{annualised normalised inverse CSV applying}\\ &\text{over the time period associated with } t.\end{aligned}$$

Note that we are now weighting the input variables $r_{i,t}$ and $r_{m,t}$ by $\sqrt{w_t}$ from Equation (4.3) thus performing WLS regression.

A noteworthy problem with WLS is that the choice of weighting has a big impact on the coefficients of regression. We have in this procedure chosen the ALSI CSV as we deem this most appropriate for the types of funds we were considering. An argument could be made for fund CSV rather than ALSI CSV, however the unavailability of fund CSV forced us towards the ALSI. Because we have chosen the ALSI we are also considering a market average CSV, one which is not style specific and thus the adjustment we make (through weightings) is not as accurate as using the sector or fund specific CSV.

All the techniques discussed above were applied to the available retail funds to produce interpretable performance measures which are analysed in Chapter 5. In the next two sections we define the zero-sum alpha and proxy benchmarks.

## 4.4 Zero-sum alphas

During the exploratory analysis phase of this research, while calculating the institutional alphas using standard regression procedures (presented in Section 4.3), we discovered an interesting anomaly. We found an overwhelming amount of positive alpha and too few negative alpha over the periods of investigation. In reality, when studying fund alphas, we should observe an even occurrence of positive and negative alpha because not every fund can experience abnormal performance simultaneously (This is a highly improbable notion). We could not proceed with the (globally positive alpha) return data thus we transformed it to create zero-sum alphas[1]. Leaving the data untouched would produce results that poorly imitate the market as we would observe too many positive alphas.

The zero-sum alpha helped us generate more realistic alphas for the institutional data. In this research, zero-sum alpha acts as a substitute for the most basic alpha produced by OLS the regression given in Equation (4.3). To create a zero-sum alpha we implement the following procedure,

1. Using the OLS defined in Equation (4.3) we can estimate the idiosyncratic terms, i.e. the orthogonal returns[2]. Restating Equation (4.3) we can perform the following re-arrangement to calculate $\epsilon_{i,t}$ as

$$\begin{aligned} r_{i,t} &= \alpha_i + \beta_i r_{m,t} + \epsilon_{i,t}, \\ \epsilon_{i,t} &= r_{i,t} - \alpha_i - \beta_i r_{m,t}. \quad \text{(Orthogonalisation)} \end{aligned}$$

2. We then use these orthogonal returns to perform a new regression

$$\epsilon_{i,t} = \widetilde{\alpha}_i + \widetilde{\beta}_i r_{m,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, \widetilde{\sigma}^2_{i,t}), \tag{4.5}$$

where

$$\begin{aligned} \epsilon_{i,t} &= \text{ orthogonal return of security } i \text{ for time period } t, \\ \widetilde{\alpha}_i &= \text{ abnormal return of security } i \end{aligned}$$

and

$$\widetilde{\beta}_i = \text{ sensitivity of security } i \text{ to changes in the market.}$$

[1] This transformation was suggested by Daniel Polakow a derivatives specialist at Old Mutual Investment Group South Africa.

[2] See Remark 4.2

We extract the $\widetilde{\alpha}_i$'s from Equation (4.5). Note that the regression in Equation (4.5) is a regression of the nominated benchmark on the orthogonalised returns of the fund.

3. Using the $\widetilde{\alpha}_i$'s we calculate their arithmetic average, $\bar{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\widetilde{\alpha}_i$.

4. Lastly we compute $\dot{\alpha}_i = \widetilde{\alpha}_i - \bar{\alpha}$ for every $i$. These new alphas are zero-sum alphas. That is $\sum_{i=1}^{N}(\widetilde{\alpha}_i - \bar{\alpha}) = \sum_{i=1}^{N}\dot{\alpha}_i = 0$. Note that they are centered around zero.

These assimilated alphas ($\dot{\alpha}_i$) can now be used to create a new return series for each respective fund through reverse engineering of the regression equation as follows

$$\dot{r}_{i,t} = \dot{\alpha}_i + \beta_i r_{m,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_{i,t}^2), \tag{4.6}$$

where $\dot{r}_{i,t}$ is the new return series, $\dot{\alpha}_i$ is the zero-sum alpha for fund $i$ and the remaining parameters as defined in Section 4.3. We can now apply the methodology in this section to $\dot{r}_{i,t}$ to generate the necessary results for the institutional data.

## 4.5 Proxy Benchmark

Similar to an investor observing fund returns we have no knowledge of the holdings the fund managers take in the various sectors, or the method they choose to allocate capital to various stocks. The proxy benchmark is a benchmark constructed using mathematical optimisation of tracking error [3] to estimate these proportional holdings. We then apply these proportions to their associated index, summing the weighted indices together to form the proxy benchmark. Using the sector indices listed in Chapter 3.2, let $F_i$ represent index $i$'s returns and $w_i$ be the weighting of $F_i$ in the proxy benchmark. Then $R_{\text{Proxy}} = \sum_i F_i w_i$ is used in the following optimisation procedure:

$$\min_{w_i} TE_{\text{Fund}}^2 = \frac{\left[\sum_{j=1}^{N}(R_{\text{Fund}} - R_{\text{Proxy}})^2\right]}{N-1},$$

with constraints

$$\sum_{i=1}^{6} w_i = 1$$

and

$$0 \le w_i \le 1 \quad \forall i,$$

[3] See Remark 4.3 for a full definition of tracking error.

where $F_i, R_{\text{Fund}}, R_{\text{Proxy}}$ are vectors of returns, $w_i$ is a scalar and $N$ is the number of elements in $R_{\text{Fund}}$. We have essentially minimised the tracking error by choosing optimal weightings in each of the sector indices to construct our proxy benchmark return. This benchmark allows us to create alphas similar to those found in passive strategies [4], where the fund return follows the benchmark as close as possible with the aim of minimising tracking error.

**Remark 4.3** (Tracking Error). *A metric which measures how closely a fund tracks its benchmark (closeness of a fund and its benchmark), usually an index in the same market as the fund. Passive funds tend to have lower tracking errors, conversely, active funds have higher tracking errors. Tracking error is mathematically defined as:*

$$TE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_{p,t} - r_{b,t})}, \tag{4.7}$$

*where $TE^2$ is tracking error variance, $N$ the number of observations, $r_{p,t}$ and $r_{b,t}$ are the portfolio and benchmark return over time period $t$, respectively. A more tractable definition of tracking error variance follows,*

$$TE^2 = (\beta_p - 1)^2\sigma_b^2 + \sigma_\epsilon^2, \tag{4.8}$$

*where $\beta_p$ is the beta of the fund, $\sigma_b$ is the benchmarks volatility and $\sigma_\epsilon^2$ is the idiosyncratic fund risk. From equation (4.8) we note that tracking error can be incurred in one of three ways: (i) through beta alignment ($\beta_p = 1$), leaving only idiosyncratic risk ($\sigma_\epsilon^2 > 0$); (ii) through beta misalignment ($\beta_p \neq 1$) and the absence of idiosyncratic risk ($\sigma_\epsilon^2 = 0$); or (iii) a combination of beta risk ($\beta_p \neq 1$) and idiosyncratic risk ($\sigma_\epsilon^2 > 0$).*

[4] See Remark 4.1 for detail on trading strategies.

# Chapter 5

# Interpretation of results

In this chapter we interpret all available results from the regressions. This includes model fit statistic R-square[1] and the various return measures discussed in Chapter 4. We also consider methods for measuring whether or not there were any changes in alpha when taking cross-sectional volatility (CSV) into account, i.e. ranking-wise and value-wise changes of alpha.

## 5.1 Model quality

How well does our choice of model fit the data? This question is answered using the coefficient of determination. Firstly, we compute R-squared for the unadjusted regression and observe the values in Table 5.1. For Time Window 1, it is clear that the R-squared value is at an acceptable level. It is also evident that Time Window 1 is fitted better than Time Window 2, for every fund. As an example, a fund with an R-squared of 0.7, means that 70% of the funds movements can be explained by movements in the benchmark.

**Tab. 5.1:** R-squared for OLS regression (Retail)

| **R-Squared Unadjusted** | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **ALEQTYF** | 0.70616 | 0.57256 |
| **CORTP20** | 0.63837 | 0.55078 |
| **INVINDX** | 0.78778 | 0.35538 |
| **CORALSI** | 0.94023 | 0.50588 |
| **RMBEQTY** | 0.79446 | 0.48594 |
| **PRUOPTM** | 0.73227 | 0.67633 |

[1] See Remark 5.1 for a definition.

Note that R-squared does not tell us about the performance of the fund, however it does tell us about the correlation between the fund and its associated benchmark. The higher the value of R-squared the more the fund's performance patterns can be explained by the performance of its benchmark. The weakest relationship presents itself in Time Window 2 for INVINDX indicating that this fund did not track its benchmark too closely during that time period. The same can be said for all R-square values below 70% as this is a suitable cut-off, see Remark 5.1.

**Remark 5.1** (Coefficient of Determination)**.** *The coefficient of determination is often referred to as R-squared or $R^2$. Mathematically we define R-Squared as*

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

*where $SS_{res}$ is the sum of squares of residuals and $SS_{tot}$ is the total sum of squares. R-squared tells us how well the model fits the data, the higher the value of R-squared the better the model fits the data and visa versa. When is the R-squared value high enough? The choice of the cut-off depends on the consequences it may have in the research. If the consequences of choosing the cut-off are severe we choose a large R-squared value enforcing a strong model fit. If the consequences are minor we choose a smaller cut-off, showing lenience towards a weaker model fit. Hence the choice of 70% as a suitable cut-off.*

**Tab. 5.2:** R-squared for WOLS regression (Retail)

| **R-Squared Adjusted** | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **ALEQTYF** | 0.73265 | 0.56726 |
| **CORTP20** | 0.62030 | 0.53036 |
| **INVINDX** | 0.76982 | 0.33753 |
| **CORALSI** | 0.93415 | 0.49133 |
| **RMBEQTY** | 0.78349 | 0.47203 |
| **PRUOPTM** | 0.69915 | 0.66243 |

After performing the weighted least squares (WLS) we found that there were improvements (seen in Table 5.2), although Time Window 1 still showed stronger correlations than Time Window 2. Because the WLS removed performance noise we saw that some of the R-squared values changed, although none changed significantly ($\pm 1\%$ throughout). For the most part retail ordinary least squares (OLS) and WLS models were well suited for the data, with some disappointing results in Time Window 2, a feature of market instability in the period (and recovery). Expanding on this point we note that when a regression model is poorly fit by data, R-square values will typically be low ($\leq 0.5$). Thus when the market is unstable (recovering or crashing) we observe data extrema which cannot be predicted by the regression model, producing a poor data fit and hence the low R-squared values.

**Tab. 5.3:** R-squared for OLS regression (Institutional)

| **R-Squared Unadjusted** | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Fund 1** | 0.94569 | 0.89482 |
| **Fund 2** | 0.95533 | 0.94417 |
| **Fund 3** | 0.89876 | 0.88741 |
| **Fund 4** | 0.98176 | 0.94980 |
| **Fund 5** | 0.96203 | 0.98104 |
| **Fund 6** | 0.96562 | 0.96613 |
| **Fund 7** | 0.93389 | 0.95171 |
| **Fund 8** | 0.86766 | 0.93422 |
| **Fund 9** | 0.96077 | 0.93899 |
| **Fund 10** | 0.73090 | 0.90613 |
| **Fund 11** | 0.92563 | 0.93457 |
| **Fund 12** | 0.99718 | 0.91894 |

**Tab. 5.4:** R-squared for WOLS regression (Institutional)

| **R-Squared Adjusted** | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Fund 1** | 0.93433 | 0.41381 |
| **Fund 2** | 0.95183 | 0.44568 |
| **Fund 3** | 0.90536 | 0.45484 |
| **Fund 4** | 0.98146 | 0.44725 |
| **Fund 5** | 0.96022 | 0.44538 |
| **Fund 6** | 0.96252 | 0.45390 |
| **Fund 7** | 0.92978 | 0.46416 |
| **Fund 8** | 0.85039 | 0.41265 |
| **Fund 9** | 0.95760 | 0.45191 |
| **Fund 10** | 0.68730 | 0.43752 |
| **Fund 11** | 0.92320 | 0.43935 |
| **Fund 12** | 0.99675 | 0.46798 |

Having performed the exact same R-squared calculations on institutional data (Tables 5.3 and 5.4), we observed R-squared values in excess of 90% for the OLS regression in Time Windows 1 and 2, and under WLS regression we noted disappointing results (R-squared values less than 50%) in Time Window 2. For any given time window or regression no fund had a dominating R-squared. This is owing to the similarity between the various funds (primarily equity based, low cash) and their market performance during the time periods. Throughout all the institutional R-squared values, the largest change lay between OLS Time Window 2 and WLS Time Window 2 where R-squared moved from values in excess of 90% to values less than 50%. This large change in R-squared indicated that the model quality of OLS was better than WLS for Time Window 2.

## 5.2 Alpha interpretation

In this section we analyse all alphas produced by the regressions defined in Chapter 4. Note that for the institutional data we have calculated zero-sum alphas (see Section 4.4) as our raw alpha. We then apply the WLS regression to calculate CSV adjusted alphas, the core purpose of this research. Finally, we consider the implications of a proxy benchmark (defined in Section 4.5) and its impact on both raw alpha and CSV adjusted alpha.

Alpha provides us with a means to compare performance between fund managers by way of a simple ranking system, the highest alpha receiving a rank of one. Fund ranking was the main method for comparing raw alphas to CSV adjusted alphas. For the purpose of this analysis we have produced several tables of alpha values: retail and institutional, raw and CSV adjusted (Tables C.1, C.2, C.3 and C.4 in Appendix C), using graphs to simplify interpretation.

**Remark 5.2** (Wilcoxon Rank Sum Test)**.** *The Wilcoxon rank sum test as suggested by Wild and Seber (1999) is performed using the following methodology. In each case (for each p-value computed) we have a two samples of alphas from two populations. Let these samples be from population $A$ and $B$, containing $n_A$ and $n_B$ observations respectively. We test the null hypothesis $H_0 : A = B$. Wilcoxons rank sum test tries to detect location shifts, and hence differences in the median of $A$ and $B$. Thus the alternative hypothesis can be written as $H_1 : A > B$, $H_1 : A < B$ or $H_1 : A \neq B$. The test ranks the sum of $n_A + n_B$ observations of the combined sample. Each observation has a rank, the smallest having rank 1[2], the largest having rank $n_A + n_B$. The test statistic for the Wilcoxon rank sum test is the sum of the ranks for the observations from one of the samples. For our purposes we have set the significance level at 5%.*

**Tab. 5.5:** p-values from the Wilcoxon rank sum test for retail alpha, rejection of $H_0$ at 5%

| **Wilcoxon ranksum p-value** | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Nominated benchmark** | 0.58874 | 0.58874 |
| **Proxy benchmark** | 0.58874 | 0.93723 |

**Tab. 5.6:** Comparison of the standard deviations of retail alpha

| **Standard deviation comparison** | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Nominated benchmark** | 1.00831 | 0.98481 |
| **Proxy benchmark** | 2.04647 | 1.11711 |

Firstly, we studied the retail alpha, Figures 5.1 and 5.2. Visually, the biggest alpha change appeared in Time Window 1 using the proxy benchmark. Ranking changes were sparse and the Wilcoxon rank sum test in all cases suggested that there were no median differences (p-values in Table 5.5). We noted that the ratio of alpha standard deviations[3] were more pronounced in Time Window 1 for both the nominated and proxy benchmarks (See Table 5.6). These ratios, both larger than one, indicated that the standard deviation of alphas increased when adjusting for CSV. The proxy benchmark produced higher alpha deviation changes in Time Window 1 and Time Window 2. Retail alpha rankings changed, but these changes were not statistically significant (referring to the size of the p-value of the Wilcoxon rank sum test).

**Remark 5.3** (Standard Deviation Ratio)**.** *The ratio of CSV adjusted vs. unadjusted alpha standard deviations mathematically,*

$$\textit{Alpha Standard Deviation Ratio} = \frac{\sigma_{\alpha \ CSV \ adjusted}}{\sigma_{\alpha \ unadjusted}}.$$

*We have essentially constructed ratios between the standard deviations of the various alpha groups. Note that we only calculate the standard deviation ratios between CSV adjusted alpha and raw alpha. We therefore have four ratios per fund type. Higher ratio values (greater than one) imply that the standard deviation of a CSV adjusted alpha is larger than the standard deviation of its raw alpha counter part. Lower ratio values (less than one) tell us that the standard deviation of the CSV adjusted alpha is smaller than the standard deviation of the raw alpha. This ratio tells us how sensitive the standard deviation of alpha is to the WLS regression.*

[3] Refer to Remark 5.3 for more detail regarding this ratio.

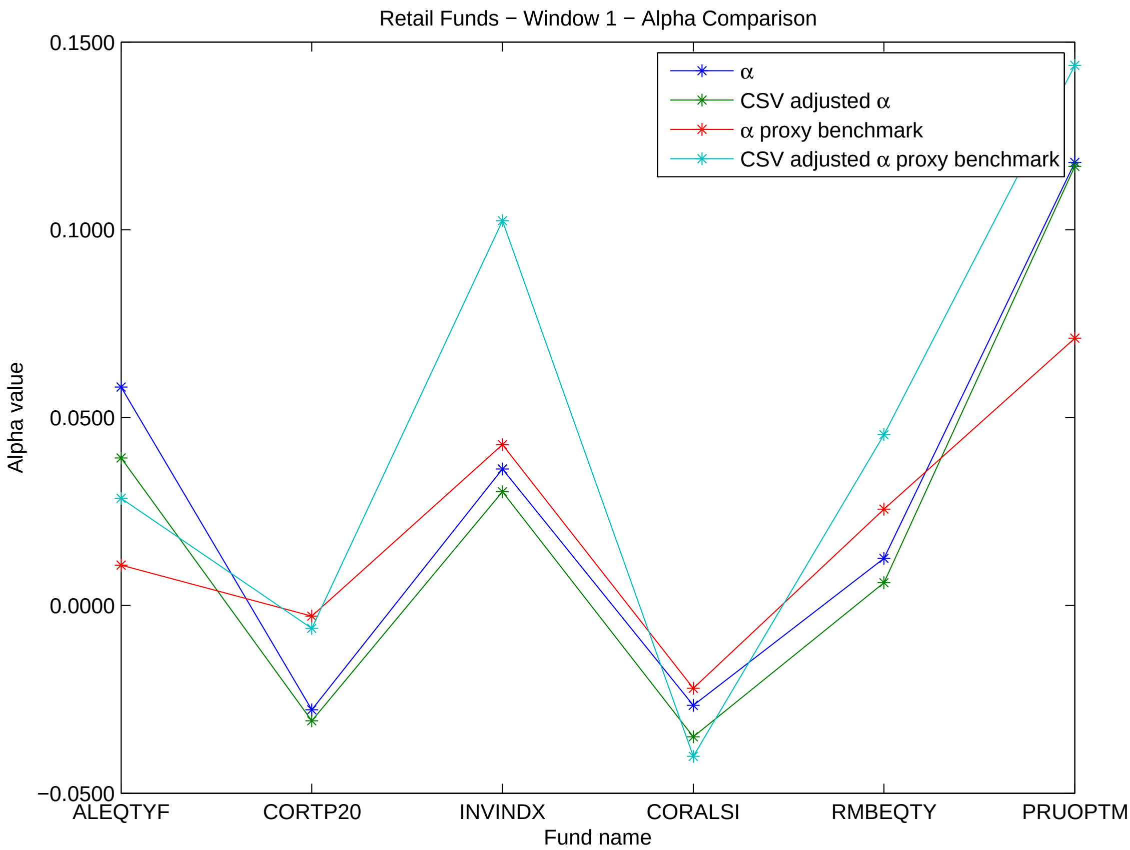


**Fig. 5.1:** Line plot comparing various retail alpha values in time window 1

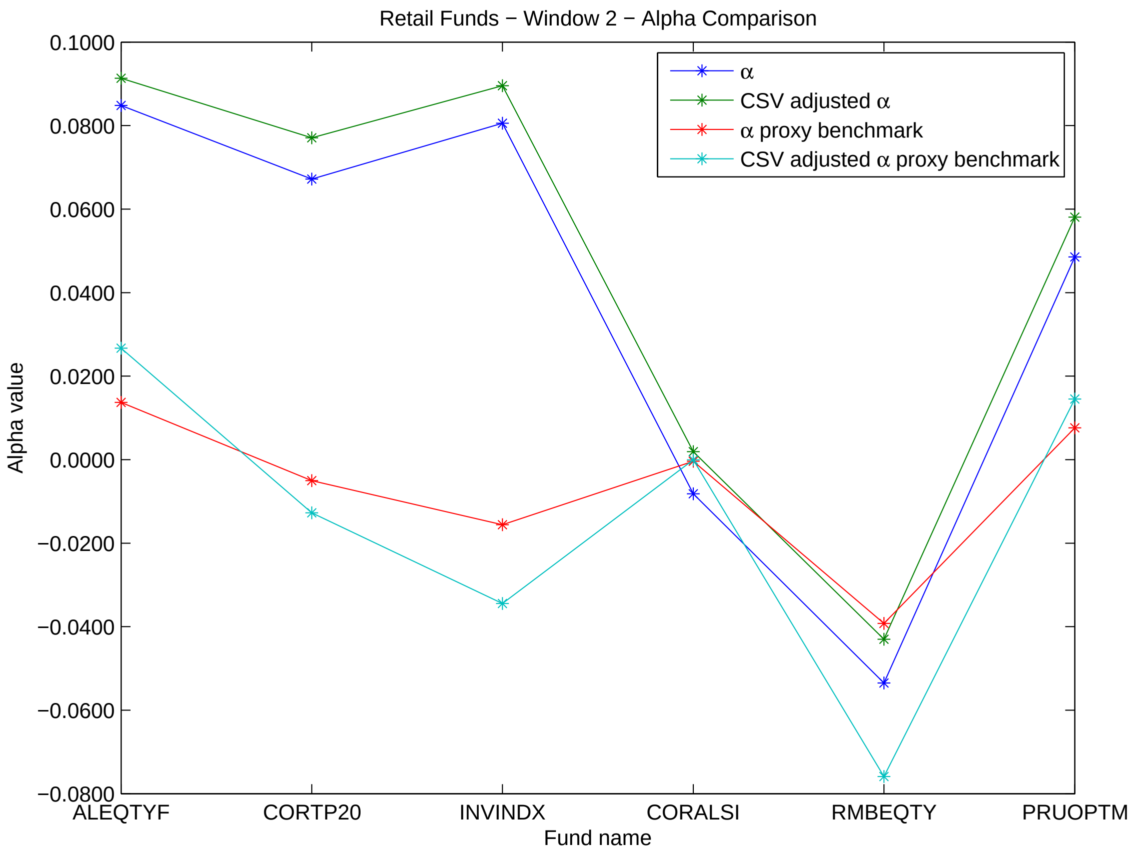


**Fig. 5.2:** Line plot comparing various retail alpha values in time window 2

**Tab. 5.7:** p-values from the Wilcoxon rank sum test for institutional alpha, rejection of $H_0$ at 5%

| Wilcoxon ranksum p-value | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Nominated benchmark** | 0.37084 | 0.50672 |
| **Proxy benchmark** | 0.17485 | 0.40250 |

**Tab. 5.8:** Comparison of the standard deviations of institutional alpha

| Standard deviation comparison | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Nominated benchmark** | 2.16379 | 1.97356 |
| **Proxy benchmark** | 2.22392 | 1.24370 |

Looking at the institutional alphas, Figures 5.3 and 5.4 we found large alpha changes in Time Window 1 for the proxy benchmark case, across multiple funds. In Time Window 2, large alpha changes were exhibited when using the nominated benchmark, however only for Fund 7. Visually, all the institutional alphas remained close together, with few shifts or changes when applying the CSV adjustment. Turning to the institutional alpha standard deviation ratios in Table 5.8 we noted high ratio values (larger than one). This tells us that alpha deviations increased when adjusting for CSV. For rank changes, once again we used the Wilcoxon rank sum test, which indicated that none of the alphas (comparing nominated with nominated and proxy with proxy) had significantly different medians, respectively. For Time Window 1 we found funds 3, 7 and 10 achieving the largest alpha values, similarly for Time Window 2. The institutional fund rankings also changed, however this change was not statistically significant either.

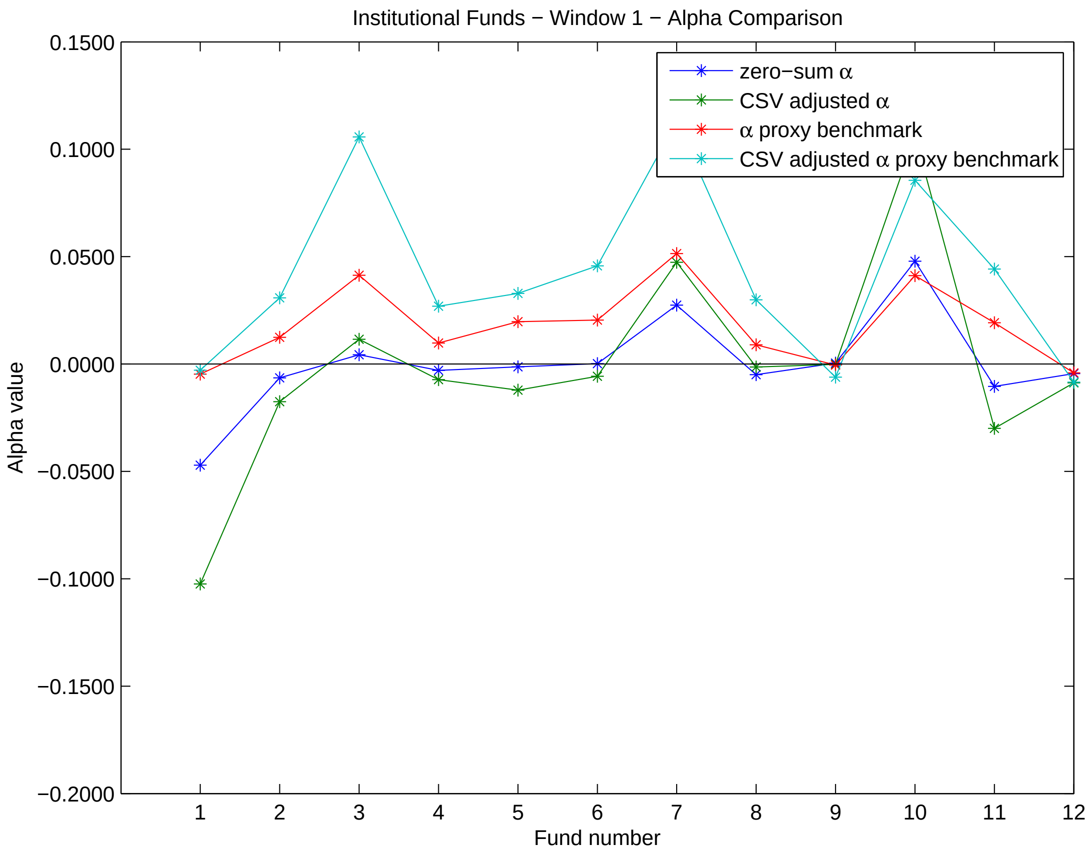


**Fig. 5.3:** Line plot comparing various institutional alpha values in time window 1

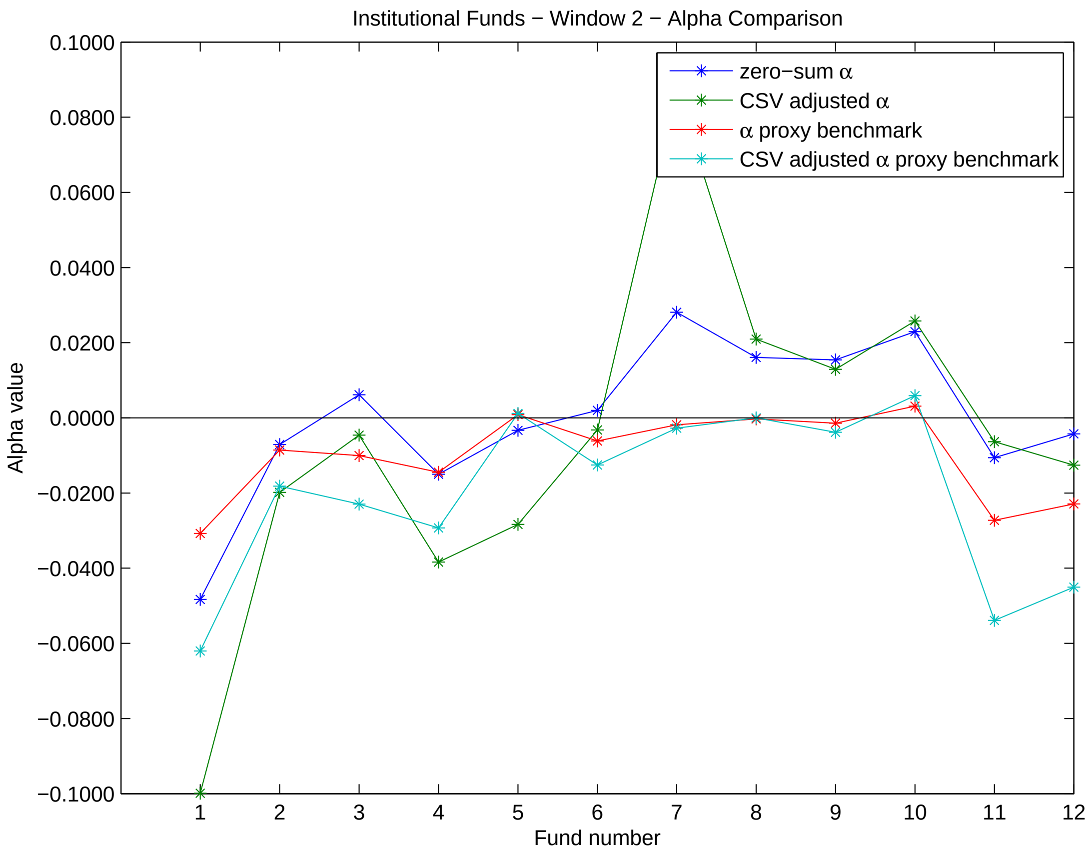


**Fig. 5.4:** Line plot comparing various institutional alpha values in time window 2

With both institutional and retail alpha in mind, the proxy benchmark created a more sensitive alpha compared to the nominated benchmark alpha. Using the nominated benchmark, we found that even when CSV was taken into account the alpha values were far more similar to one another, even more so in the retail environment. Without the consideration of the Wilcoxon rank sum test, turning to the raw ranking tables (see appendix D) we can conclude that although the Wilcoxon test did not pick up any median difference there were still rank changes present. The impact of these rank changes should not be ignored in practice and they are discussed briefly in the conclusion of this work.

# Chapter 6

# Conclusion

In this chapter we conclude on our analysis, answer the research question and comment on any other interesting pieces of information we have uncovered during this investigative process. We finish with suggestions for further research.

## 6.1 Digressions

The choice of time window gave us exposure to the movements experienced by both retail and institutional funds in South Africa. We ignored the spike in 2008 caused by the crash as we believed this would create erroneous regression coefficients (namely, alpha) by skewing the data. During the preliminary data exploration we noticed several problems which lead to the creation of the proxy benchmark and the zero-sum alpha[1] approaches. We then had the opportunity to investigate another dimension of the research problem. Zero-sum alphas (only for the institutional funds), centered around zero, were more sensible alpha estimates of fund performance[2]. We found that when adjusting for cross-sectional volatility (CSV), alphas derived using the proxy benchmark were more sensitive, experiencing larger changes in comparison to the nominated benchmark alternative. When establishing the rankings for each fund we discovered that there were only slight differences which lead us to the Wilcoxon rank sum test. The Wilcoxon rank sum did not detect any significant differences between the ranks. Before consulting the rankings directly we created one final measure to inform us, the ratio of alpha standard deviations. It was now possible to determine whether or not cross-sectional volatility had any effect on performance measurement.

[1] See Chapter 4 for more detail on the proxy benchmark and zero-sum alpha.

[2] As mentioned in Chapter 4, the positive alpha anomaly is a poor representation of the actual market

## 6.2 Does cross-sectional volatility have an effect on performance measurement in South Africa?

Yes, CSV does have an effect on performance measurement in South Africa. Since we were interested in measuring cross-sectional volatilities effect on fund performance (or managerial skill) we have centered our focus on alpha. After careful consideration of all the results we found it necessary to approach the discussion using three specific concepts: Rank-wise changes to alpha, value-wise changes to alpha and remuneration effects.

Firstly, considering the ranked alphas having used the Wilcoxon test we established that there was no significant difference when moving from ordinary least squares (OLS) to the weighted leasts squares (WLS) framework. Thus we can state that taking CSV into account has no effect rank-wise.

As indicated in Chapter 5 the value changes of alpha cannot be ignored (even though they are small). These value changes can still impact the remuneration of (un)skilled fund managers. Thus when taking CSV into account the value changes in alpha can have effect on the remuneration. A supporting example: Funds whose alphas increased when moving from OLS to WLS regression indicate that these managers are under compensated, and should, in-light of the larger alpha, receive an alpha adjusted compensation. This compensation should be linked to the alpha produced by the WLS regression. Therefore, taking CSV into account should also effect the remuneration of fund managers. What we cannot say, is how large or small this compensation should be. Given the results from Chapter 2 we cannot state whether an increase or decrease in alpha is better although consistency across different CSV environments would be ideal.

Although WLS using normalised inverse CSV did not affect the rankings of fund alphas the value change can be attributed to the regression coefficients change. When applying $\sqrt{w_i t}$ to the standard OLS regression (See Equation (4.3)) we are effectively discounting imprecision (See Equation (4.4)) hence the change in alpha arises when correcting for the heteroskedastic noise. This is best explained by quoting Wasserman (2006): "Trying to give equal attention to all parts of the input space (benchmark return) is a waste of time; we should be more concerned about fitting well where the noise is small, and expect to fit poorly where the noise is big". In Chapter 2 we noted that alpha dispersion decreased when adjusting for CSV however this change was not evident in the real world data. This can be attributed to the scale of the study (lack of data) and the data problems.

Finally, CSV leads to generation of alpha. Managers cannot expect to outperform, or generate alpha, without the existence of CSV.

## 6.3 Further research

We believe this research can be extended by considering the following list of ideas for further research

- Measurement: Consideration of alternative skill measurement metric. In this research we identified the characteristics of alpha for the perfect fund. We found that CSV adjustments either increased or decreased the alpha of non-perfect funds. We were unsuccessful in detecting rank changes of funds. As mentioned above, CSV allows managers to obtain alpha. How would alpha behave across different CSV regimes? If we chose a particular fund and observed its CSV adjusted alpha across these regimes we could then measure the managers consistency and rank the funds accordingly. i.e. Consistency refers to the managers ability to maintain a level alpha across different CSV regimes. The manager with the highest consistency measure could be the highest skilled manager, however it is also possible that he is the luckiest manager. It is necessary to construct this consistency measure so that it removes the possibility of luck. Using the perfect fund as an example, we would expect to see no change in alpha adjusted for every CSV regime.

- Remuneration: How should one go about adjusting fund managers remuneration when comparing alpha to adjusted alpha?

- Scale: Increase the size of the study, include more funds while keeping it within the South African market.

# Appendix A

# Simulated price paths

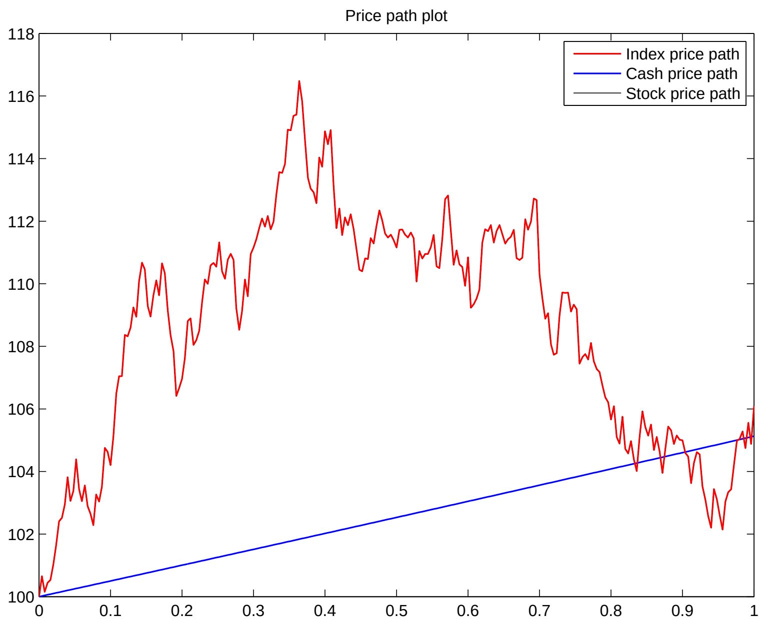


**Fig. A.1:** Price paths for Case 1

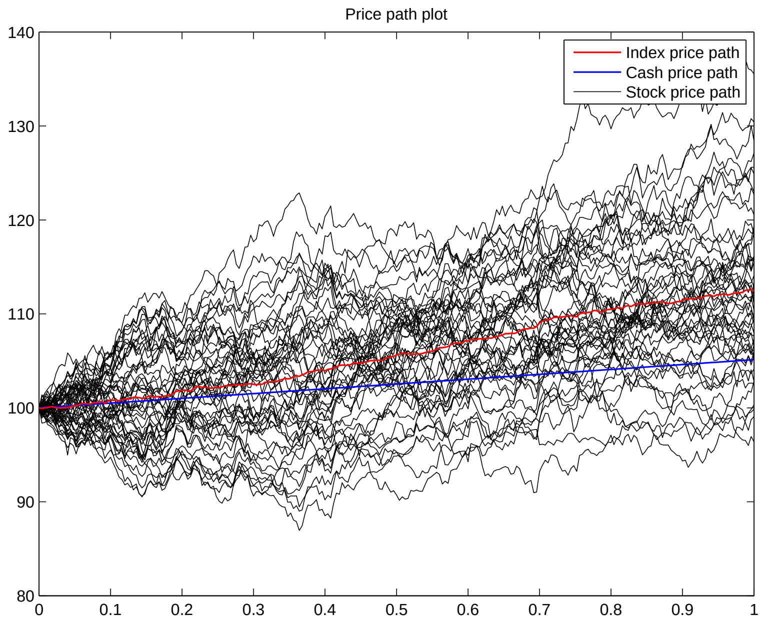


**Fig. A.2:** Price paths for Case 2

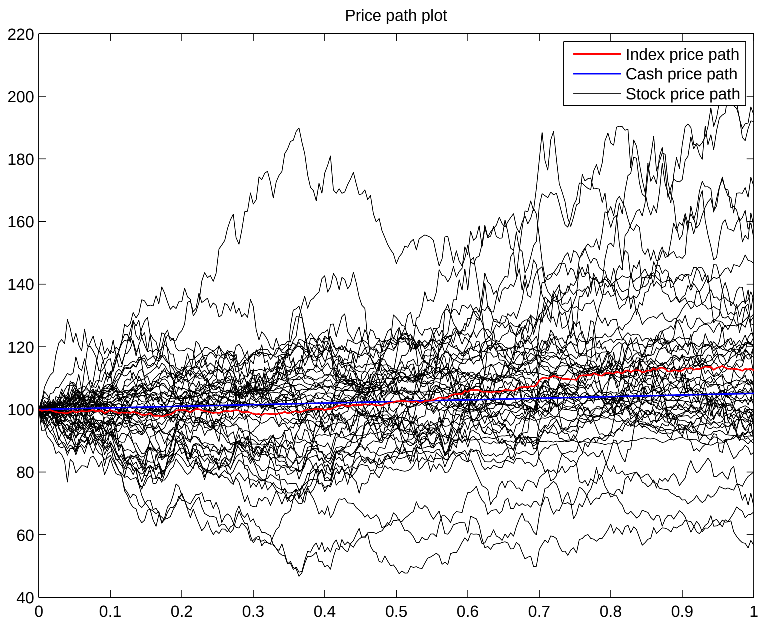


**Fig. A.3:** Price paths for Case 3

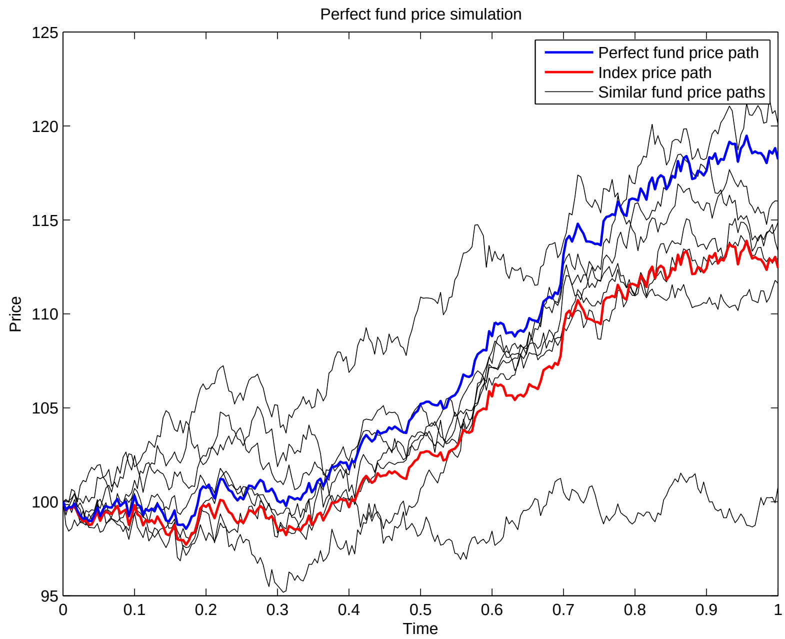


**Fig. A.4:** Price paths for the perfect fund compared to similar funds

# Appendix B

# Indices

**Tab. B.1:** Table of market sector indices

| Index Code | Full Name | Explanation |
|---|---|---|
| TOP40 | JSE Top 40 | JSE Equally Weighted Top 40 Index is a market capitalisation weighted index consisting of stocks in the FTSE/JSE Top 40 Index weighted equally at each quarterly review. |
| RESI20 | JSE Resource 20 Index | The FTSE/JSE Resources 20 Index consists of the 20 largest companies ranked by full market value, i.e. before the application of any investability weightings, as at the date of the review in the Resources, defined as the combination of Mining and Oil & Gas Sectors. |
| INDI25 | JSE Industrial 25 Index | Industrial 25 Index is a market capitalisation weighted index consisting of Industrial stocks in the FTSE/JSE All share Index weighted equally at each quarterly review. |
| FINI15 | JSE Financial 15 Index | Financial 15 Index is a market capitalisation weighted index consisting of Financial stocks in the FTSE/JSE All share Index weighted equally at each quarterly review. |
| STeFI | Short Term Fixed Interest Index | Designed to approximate the performance of money market instruments in the market. STeFI is a cash index. |
| ALBIP | All Bond Price index | The BEASSA All Bond Price index comprises of bonds from across the full range of maturities in the bond market and is a useful summary measure of the movement in the bond market. This index does not consider the reinvestment of coupons and therefore only considers the price or capital gain. |

## Appendix C

# Alpha tables

## C.1 Raw alpha

**Tab. C.1:** Raw alphas from OLS regression (Retail)

| Alpha Unadjusted | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **ALEQTYF** | 0.05815 | 0.08482 |
| **CORTP20** | -0.02777 | 0.06721 |
| **INVINDX** | 0.03633 | 0.08055 |
| **CORALSI** | -0.02659 | -0.00820 |
| **RMBEQTY** | 0.01251 | -0.05350 |
| **PRUOPTM** | 0.11792 | 0.04855 |

**Tab. C.2:** Raw alphas from OLS regression (Institutional)

| Alpha Unadjusted | | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Fund 1** | -0.04708 | -0.04831 |
| **Fund 2** | -0.00643 | -0.00708 |
| **Fund 3** | 0.00433 | 0.00615 |
| **Fund 4** | -0.00299 | -0.01502 |
| **Fund 5** | -0.00129 | -0.00329 |
| **Fund 6** | 0.00019 | 0.00200 |
| **Fund 7** | 0.02740 | 0.02809 |
| **Fund 8** | -0.00494 | 0.01607 |
| **Fund 9** | 0.00038 | 0.01542 |
| **Fund 10** | 0.04783 | 0.02293 |
| **Fund 11** | -0.01044 | -0.01059 |
| **Fund 12** | -0.00446 | -0.00424 |

## C.2 CSV adjusted alpha

**Tab. C.3:** Alphas from WOLS regression (Retail)

| | **Alpha Adjusted** | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **ALEQTYF** | 0.03925 | 0.09133 |
| **CORTP20** | -0.03074 | 0.07708 |
| **INVINDX** | 0.03026 | 0.08953 |
| **CORALSI** | -0.03499 | 0.00192 |
| **RMBEQTY** | 0.00606 | -0.04303 |
| **PRUOPTM** | 0.11690 | 0.05806 |

**Tab. C.4:** Alphas from WOLS regression (Institutional)

| | **Alpha Adjusted** | |
|---|---|---|
| | **Jan06-Jan08** | **Jan11-Jan13** |
| **Fund 1** | -0.10239 | -0.09987 |
| **Fund 2** | -0.01756 | -0.01983 |
| **Fund 3** | 0.01149 | -0.00460 |
| **Fund 4** | -0.00730 | -0.03835 |
| **Fund 5** | -0.01219 | -0.02834 |
| **Fund 6** | -0.00570 | -0.00322 |
| **Fund 7** | 0.04739 | 0.08697 |
| **Fund 8** | -0.00144 | 0.02097 |
| **Fund 9** | -0.00013 | 0.01293 |
| **Fund 10** | 0.10631 | 0.02577 |
| **Fund 11** | -0.03000 | -0.00635 |
| **Fund 12** | -0.00877 | -0.01260 |

# Appendix D

# Alpha Rankings

## D.1 Retail Rankings

**Tab. D.1:** Table of retail alpha rankings for Time Window 1

| Retail W1 Rankings | | | | |
|---|---|---|---|---|
| Fund Name | Alpha | CSV adjusted alpha | Alpha proxy benchmark | CSV adjusted alpha proxy benchmark |
| ALEQTYF | 2 | 2 | 4 | 4 |
| CORTP20 | 6 | 5 | 5 | 5 |
| INVINDX | 3 | 3 | 2 | 2 |
| CORALSI | 5 | 6 | 6 | 6 |
| RMBEQTY | 4 | 4 | 3 | 3 |
| PRUOPTM | 1 | 1 | 1 | 1 |

**Tab. D.2:** Table of retail alpha rankings for Time Window 2

| Retail W2 Rankings | | | | |
|---|---|---|---|---|
| Fund Name | Alpha | CSV adjusted alpha | Alpha proxy benchmark | CSV adjusted alpha proxy benchmark |
| ALEQTYF | 1 | 1 | 1 | 1 |
| CORTP20 | 3 | 3 | 4 | 4 |
| INVINDX | 2 | 2 | 5 | 5 |
| CORALSI | 5 | 5 | 3 | 3 |
| RMBEQTY | 6 | 6 | 6 | 6 |
| PRUOPTM | 4 | 4 | 2 | 2 |

## D.2 Institutional Rankings

**Tab. D.3:** Table of institutional alpha rankings for Time Window 2

| Institutional W1 Rankings | | | | |
|---|---|---|---|---|
| Fund Number | zero-sum alpha | CSV adjusted alpha | Alpha proxy benchmark | CSV adjusted alpha proxy benchmark |
| 1 | 12 | 12 | 12 | 10 |
| 2 | 10 | 10 | 7 | 7 |
| 3 | 3 | 3 | 2 | 2 |
| 4 | 7 | 7 | 8 | 9 |
| 5 | 6 | 9 | 5 | 6 |
| 6 | 5 | 6 | 4 | 4 |
| 7 | 2 | 2 | 1 | 1 |
| 8 | 9 | 5 | 9 | 8 |
| 9 | 4 | 4 | 10 | 11 |
| 10 | 1 | 1 | 3 | 3 |
| 11 | 11 | 11 | 6 | 5 |
| 12 | 8 | 8 | 11 | 12 |

**Tab. D.4:** Table of institutional alpha rankings for Time Window 1

| Institutional W2 Rankings | | | | |
|---|---|---|---|---|
| Fund Number | zero-sum alpha | CSV adjusted alpha | Alpha proxy benchmark | CSV adjusted alpha proxy benchmark |
| 1 | 12 | 12 | 12 | 12 |
| 2 | 9 | 9 | 7 | 7 |
| 3 | 5 | 6 | 8 | 8 |
| 4 | 11 | 11 | 9 | 9 |
| 5 | 7 | 10 | 2 | 2 |
| 6 | 6 | 5 | 6 | 6 |
| 7 | 1 | 1 | 5 | 4 |
| 8 | 3 | 3 | 3 | 3 |
| 9 | 4 | 4 | 4 | 5 |
| 10 | 2 | 2 | 1 | 1 |
| 11 | 10 | 7 | 11 | 11 |
| 12 | 8 | 8 | 10 | 10 |

# Bibliography


Ankrim, E. M. and Ding, Z. (2002). Cross-sectional volatility and return dispersion, *Financial Analysts Journal* **58**(5): pp. 67–73.
**URL:** *http://www.jstor.org/stable/4480418*

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation, *Econometrica: Journal of the Econometric Society* pp. 1287–1294.

Campbell, J. Y., Lettau, M., Malkiel, B. G. and Xu, Y. (2001). Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk, *The Journal of Finance* **56**(1): pp. 1–43.
**URL:** *http://www.jstor.org/stable/222462*

de Silva, H., Sapra, S. and Thorley, S. (2001). Return dispersion and active management, *Financial Analysts Journal* **57**(5): pp. 29–42.
**URL:** *http://www.jstor.org/stable/4480342*

Raubenheimer, H. (2012). *Managing portfolio managers: the impacts of market concentration, cross-sectional return dispersion and restrictions on short sales*, PhD thesis, Stellenbosch: Stellenbosch University.

Strong, N. (1992). Modelling abnormal returns: a review article, *Journal of Business Finance & Accounting* **19**(4): 533–553.

Wasserman, L. (2006). *All of nonparametric statistics*, Springer.

Wild, C. and Seber, G. (1999). *Chance Encounters: A First Course in Data Analysis and Inference*, Wiley.
**URL:** *http://books.google.co.za/books?id=r5uiQgAACAAJ*