Can profitable trading strategies be derived from investment best-sellers?

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Can Profitable Trading Strategies be derived from Investment Best-sellers?

Chris Brooks[1]
ISMA Centre, Department of Economics, University of Reading

William Chow
ISMA Centre, Department of Economics, University of Reading

Charles W. R. Ward
Department of Land Management, University of Reading

Abstract
A glance along the finance shelves at any bookshop reveals a large number of books that seek to show readers how to "make a million" or "beat the market" with allegedly highly profitable equity trading strategies. This paper investigates whether useful trading strategies can be derived from popular books of investment strategy, with "What Works on Wall Street" by James P. O'Shaughnessy, used as an example. Specifically, we test whether this strategy would have produced a similarly spectacular performance in the UK context as was demonstrated by the author for the US market. As part of our investigation, we highlight a general methodology for determining whether the observed superior performance of a trading rule could be attributed in part or in entirety to data mining. Overall, we find that the O' Shaughnessy rule performs reasonably well in the UK equity market, yielding higher returns than the FTSE All-Share Index, but lower returns than an equally weighted benchmark.

November 2000

1. Introduction

Many books have been written on how investors could enhance their wealth by adopting strategies suggested by authors from varied backgrounds. These authors of best-sellers are often key figures in large financial and to a lesser extent academic institutions. According to these authors, specific strategies have consistently beaten the market as evidenced by their personal experience or their own research.

A glance along the finance shelves at any bookshop reveals that there are a large number of such books, and many have become best-sellers, with more than half a million copies of each being sold world-wide. Although many purchasers of such books read them for pure enjoyment and have no intention of staking money along the lines that the authors suggest, equally there may be many purchasers who endeavour to use them as a means of learning more about the way financial markets work, and as an aid to developing their own trading strategies.

In spite of the obvious popular appeal of such texts, the vast majority of academics and probably many practitioners are extremely sceptical about whether such publications have genuine merit. The objectives of our proposed research are first, to survey the main best-sellers that embody information which could be viewed as aiding the investor's decision about what and how to trade, and second, to evaluate whether such indications can be reduced to a set of formal rules. Initial investigations revealed that such rules can indeed be formed from many of the publications, and hence we attempt to evaluate the rules in a consistent and coherent fashion on a single set of data. Some authors have conducted small studies of their own, but it is also important that the replicability of these results is considered using different markets, longer runs of data, or different frequencies of observation. To this end, in particular, we also implement these strategies using data from the UK market, while the majority of original authors base their evidence on US data and markets.

Peer-reviewed articles in academic journals are usually assumed to be rigorous in their application of statistical tests, and in their theoretical justification of the models considered. The same cannot be said of investment best-sellers, although by the same token, it is not appropriate to dismiss without investigation all the material that has not been subject to the same refereeing process as utterly worthless. The purpose of

this paper is therefore to conduct a formal evaluation of the merit of one of these popular texts, to explore whether the ideas and approaches underlying the principles in this best-selling text are soundly based on rigorous models well grounded in financial theory, and to test how robust the model are to changes in time period and market. As part of our investigation, we also highlight a general methodology for determining whether the observed superior performance of a trading rule could be attributed in part or in entirety to data mining. Data mining occurs when a given set of data is used more than once for the purposes of inference or model selection. As a result of such data reuse, superior results obtained may be due to chance alone rather than to any merit inherent in the method. The test that we employ for the effects of data mining is based upon the notion of nesting the selected rule within a universe of similar rules, and bootstrapping from this universe to determine whether the rule under consideration is significantly better than the others. The test, known as a "reality check", is due to White (see White, 2000 or Sullivan *et al*, 2000), and could have widespread applicability as a general method for evaluating trading rules.

The remainder of this paper is organised as follows. Section 2 presents a brief survey and classification of several investment best-sellers, and describes the trading rule described in O'Shaugnessey's book and how we test it. Section 3 describes the data employed and presents the results of our application to the London equity market. Section 4 describes the data snooping test, and outlines the results derived from its application in the context of O'Shaugnessey's rule. Finally, section 5 concludes.

2. Does "What Works on Wall Street" Necessarily Work in the City?

2.1 Survey of Best-sellers

The best-sellers used in this project were obtained from the best-sellers list compiled by *Amazon.com* as of 4 January 1999. *Amazon.com* is a leading firm that provides online purchasing of books through the internet. Two categories under the "*Investment*" field were considered. One focuses on "Stocks" alone while the list of "General" incorporates the whole range of items under investments. The US rankings of the books reviewed on 4 January 1999 are shown below.

AMAZON.COM "Investing/Stocks" best-seller list

Author	Title	Ranking
William J. **O'Neil**	How To Make Money in Stocks	#2
David N. **Dreman**	Contrarian Investment Strategies	#3
Peter **Lynch**	One Up on Wall Street	#4
Jeremy J. **Siegel**	Stocks for the Long Run	#5

AMAZON.COM "Investing/General" best-seller list

Author	Title	Ranking
Mary **Buffett**	Buffettology	#10
Burton G. **Malkiel**	A Random Walk Down Wall Street	#14
James P. **O'Shaughnessy**	What Works on Wall Street	#45

These books are as popular in the UK as they are in the US. To compare the popularity of these books in the UK, the Amazon best-seller lists for UK market are shown below.

AMAZON.UK.COM "Investing/Stocks" best-seller list

Author	Title	Ranking
William J. **O'Neil**	How To Make Money in Stocks	#1
David N. **Dreman**	Contrarian Investment Strategies	#2
Jeremy J. **Siegel**	Stocks for the Long Run	#5
Peter **Lynch**	One Up on Wall Street	#12

AMAZON.UK.COM "Investing/General" best-seller list

Author	Title	Ranking
Mary **Buffett**	Buffettology	#2
William J. **O'Neil**	How To Make Money in Stocks	#4
Jeremy J. **Siegel**	Stocks for the Long Run	#12
James P. **O'Shaughnessy**	What Works on Wall Street	#40
Peter **Lynch**	One Up on Wall Street	#48

The structure and authorship of these books are such that they appear to fall neatly into one of two broad categories. Best-sellers by the "academic" group (e.g., Malkiel, 1996; Siegel, 1998) are hesitant in providing specific trading rules and concentrate more on the theoretical aspects as expected. The "professionals" (e.g., O'Shaugnessey, 1998; Dreman, 1998; Buffett, 1997; Lynch 1989; and O'Neil, 1995), are bolder in this aspect as many provide testable trading rules. A further breakdown of the "professionals" category could be made by segregating best-sellers that have substantial statistical research supporting their trading rules and those without. Below are the authors of the best-sellers categorised into their relevant groups:

Structure of books reviewed

Academics
- Malkiel
- Siegel

Professionals

Statistical Backing
- O'Shaughnessy
- Dreman

No Research
- Buffett
- Lynch
- O'Neil
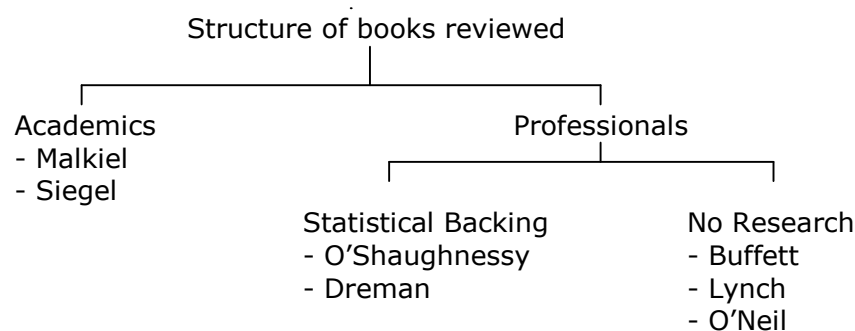
For the remainder of this paper, we investigate only O'Shaugnessey's book for illustrative purposes; other books will be the subjects of our future research.

2.2 O'Shaugnessey's Second Most Profitable Trading Rule

According to O'Shaughnessy's findings, the strategy of selecting the top 50 RPS (Relative Price Strength) ranked stocks with Earnings Price Ratio, EPR, of above 5% (i.e. Price Earnings Ratio, PER of less than 20x) turned an initial investment $10,000 into $12,570,451 at the end of 45 years. This return is a result of investing $10,000 in 1954 and rebalancing the portfolio annually until the end of 1996. For each year, the portfolio would consist of 50 equally weighted stocks. We shall refer to this strategy as the OSH2 strategy. This strategy generated a compounded return of 18.52%, a standard deviation of 24.48% and a Sharpe Ratio of 61. It is ranked second in terms of the highest absolute return amongst all the other strategies examined in the author's book.

Adjusting for simple biases

The author used stocks from the Standard & Poor's Compustat Active Research database. By doing so, the results are not flawed by survivorship bias. As the London Share Price Database (LSPD) is adjusted for dead stocks, our study also controls for survivorship bias in the sense that all the data for each period consists of all of the stocks that were actually trading at that moment in time.

To avoid look-ahead bias, O'Shaughnessey delayed forming portfolios until the end of the year when financial statements are released. Thus, the information used to form the strategy would have been publicly available at the time it was used. We again follow suit. For example, to build the EPR, we employed the EPS (Earnings per Share) as reported in the financial statement. The price used in calculating the EPR is the price at the end of the year following the financial fiscal year. Thus, the price is a year forward compared to the corresponding EPS. Stocks with negative and zero EPS have been excluded from our analysis.

The author selected stocks that had at least $150m (at end of sample price levels) in market capitalisation on the grounds that it was necessary to avoid small, illiquid stocks that are infrequently traded and could have considerably higher transactions costs. The author adjusted this figure over time by incorporating the inflation rate for the previous years. In the spirit of this approach, we have also adjusted the minimum market capitalisation figure over the sample period of 1969-97 by deflating £100m backwards instead of the author's $150m. We arrived at the £100m figure after converting US dollars to British pounds at the rate of 1.5 dollars to a pound. The value of £100m in the beginning of 1998 was £10.466m in 1969. Market capitalisation for each company is calculated by multiplying the price as of the end of the year by the number of ordinary shares in issue at the beginning of the year. This is done to strictly follow O'Shaughnessy's strategy.

Alternatively, the minimum market capitalisation could be derived by comparing the percentage of stocks in the NYSE that are above $150m and subsequently, using this percentage on the LSE to determine the appropriate minimum market capitalisation for this strategy. As of the 21st of October 1999, there were 1,972 stocks on the New York Stock Exchange (NYSE). After setting the criterion of $150m on Bloomberg's

Equity Search Result Criteria (ESRC) programme, the number of stocks that remained were 1,568. Thus, approximately 79.5% of the stocks in NYSE had a minimum market capitalisation of $150m. On the same day, the London Stock Exchange (LSE) had a total of 1,937 quoted stocks. The number of stocks that would constitute 79.5% of the LSE was 1,540. By setting a minimum market capitalisation of £5m, the total number of stocks turned out to be 1,517, which is close to the target of 1,540. As well as deflating £100m, we also test the OSH2 strategy by using £5m as the minimum market capitalisation.

The three elements of OSH2

The only difference in our analysis concerning the three elements of O'shaughnessy's strategy is that we have applied the deflated minimum market capitalisation of each year individually to the respective years. However, O'Shaughnessy applied the 5 year average deflated market capitalisation to the respective 5 years. No particular reason was given for his approach.

The Relative Price Strength (RPS) is simply the price index of the current price divided by the previous year's price. For example if the current year-end price is 20 and the previous year's price is 5, then the RPS is 4. The Earnings-Price Ratio (EPR) is constructed by dividing the financial year Earnings Per Share (EPS) with the stock's end-of-year price. After the initial selection of stocks with a deflated market capitalisation of £100m and an EPR of more than 5%, the portfolio would subsequently consist of the 50 highest RPS ranked stocks.

Excluded Stocks & Returns Definition

In this study we excluded secondary shares. Secondary shares or "class B" shares differ in their right to vote and receive dividends compared to Ordinary shares ("class A"). Secondary shares usually have limited voting privileges.

In addition, we eliminated firms that belong to the industry groups of Financial Trust, Investment Trust and Unit Trust since these are already portfolios of shares, only some constituents of which will satisfy the O'Shaughnessey rules. However, foreign shares traded on the London Stock Exchange were included in compliance with O'Shaughnessy's criteria. Returns calculated include both capital appreciation and

dividend received. The prices used for the return calculation are constructed as follows. Stocks were purchased using their actual transacted price. Should there be no quoted transacted price on the 31st of December, the asking price would be used. Stocks without either price would not be purchased. Stocks were sold at their transacted price. Failing that, the bidding price would be used in our analysis. If a stock has neither a transacted nor bidding price, the transacted price for the 13th month after purchase would be used.

Stocks that were no longer listed due to cessation of trading the year following their selection as part of the portfolio have been dealt with by assuming a total 100% loss. Since the OSH2 rule requires 50 equally weighted stocks, bankruptcy would reduce the portfolio value by 2%. There were 2 such firms picked by the rule over the 29-year period. A more common event was for a stock to cease listing due to a merger or takeover. For such stocks, returns were calculated using the last available quoted price for the old company. This study therefore assumes that the investor sells the shares outright rather than receiving shares in the acquiring or merged firm. This is rational since the acquiring or merged firm may not satisfy the OSH2 criteria, and should therefore not be part of the portfolio.

3. The Data & Results of Applying O'Shaughnessy's Strategy

3.1 Data

The London Share Price Database (LSPD) obtained from the London Business School, comprises financial particulars of 6,757 firms – i.e., all shares traded on the London Stock Exchange. The data spans the period 1969-1998. Portfolios are constructed annually from 1969 until 1997, and their performance measured one year later.

After excluding the shares of companies as described above, the total number of firms examined ranged between 1085 and 2130 (fourth column of Table 1). After setting the 3 requirements of OSH2 (EPR>5%, market capitalisation above deflated £100m and the 50 highest RPS ranked stocks), the sub-sample size ranged from 173 in 1971 to 1,360 in 1979.

3.2 Results of Applying O'Shaughnessy's Strategy

Besides the OSH2 strategy, the reciprocal strategy (selecting the bottom 50 RPS ranked stocks) is also examined to illustrate the potential effects of data-mining. The results of the 50 top RPS and the 50 bottom RPS portfolios are displayed in Table 2. The first benchmark used to measure the performance of the strategy is the FT All Share (FTA) index. From 1969-97, the FTA generated an average return of 14.31% as indicated in Table 2.

The strategy suggested by O'Shaughnessy, OSH2, out-performed the FTA by 2.59% per annum over 29 years. However, the average return of OSH2 is not significantly different from the return of the FTA All-Share (with a *t*-statistic of 0.567). To measure the risk-reward payoff, we employ the Sharpe Ratio. To calculate the Sharpe Ratio, we used the average arithmetic return and the risk-free rate of 9.93%, which was the annualised one-month Treasury bill rate. In addition to the higher return obtained by the OSH2 strategy, the risk involved as measured by standard deviation was slightly lower than the market's. The OSH2 strategy consequently recorded a Sharpe Ratio of 25.23 compared to the benchmark's 13.76.

We also employed another benchmark to better match the performance of the OSH2 strategy since this strategy involves forming an equally weighted portfolio. The second benchmark is produced by equally weighting the returns of all the quoted stocks above the minimum market capitalisation threshold for each respective year. Using a deflated minimum market capitalisation of £100m, the average returns were recorded at 20.66%. This would mean an under-performance of 3.76% by the OSH2 strategy. The yearly performance of the strategies and the benchmarks are presented in Table 3.

By contrast, the portfolio with the 50 lowest RPS-ranked stocks performed well. This portfolio significantly outperformed the FTA benchmark by a huge margin of 14.96% with a t-statistic of 2.56. Although the standard deviation is much higher, its spectacular average returns over the 29 years of study more than compensated for this additional risk as evidenced by its high Sharpe ratio of 47.37. Thus, constructing a portfolio that is in direct contrast to the one O'Shaughnessy suggested would yield the

highest average return over the 29 years of study compared with OSH2 and the two benchmarks.

An investor adopting the OSH2 strategy would turn an initial investment of £10,000 to £375,639 over the 29 years. Graph 1 charts the log returns which provides a clearer illustration of analysing the current cummulative returns (Graph 2) at each point in time for all the portfolios investigated. Graph 3 displays the terminal wealth of the respective portfolios. Building the portfolio that consists of the 50 lowest RPS ranked stocks would reap an incredible total wealth of £4,971,732 by the end of 1998. However, there is no economic rationale for the performance of a portfolio constructed using the obverse of O'Shaughnessy's strategy. This illustrates well the point that it is easy to find a highly profitable strategy that would have worked well in the past. However, all subsequent analysis will concentrate on OSH2 only.

Return difference from using different minimum market capitalisation
The return resulting from replication of the OSH2 strategy with a minimum deflated market capitalisation of £5m is only slightly higher (1%) than that obtained by using a minimum deflated market capitalisation of £100m, which we have done so far. The slightly higher return could be attributed to an increase in the number of smaller stocks for consideration. By lowering the deflated minimum market capitalisation to £5m, the total number of stocks in the sample increased by 284 firms, from the previous 22,728 to 23,012 firms. Due to foreseeable liquidity complications from using small stocks in a real application, we should be conservative and examine the rest of the results from using a minimum deflated market capitalisation of £100m. The results of using a minimum market capitalisation of £5m can be seen in Table 4.

After replicating OSH2 using the 5-year average deflated market capitalisation approach, we find that the returns are slightly higher (0.1%) than for the original results. The reason for the slightly higher result (shown in Table 4) is similar to the one described above after replicating the OSH2 strategy using a deflated market capitalisation of £5m, which is the inclusion of more smaller capitalised stocks (which were found *ex post* to exhibit a higher probability of higher returns). Again, due to its similarity with the previous results, we do not consider this variant further.

Consistency of Performance and Risk Analysis over time of the OSH2 Strategy

On a year-to-year basis, the OSH2 portfolio beat the FTA 15 times out of 29 years, which is equivalent to 52% of the time. Graph 4 shows the returns on a year-to-year basis. As can be seen, the FTA outperformed the equally weighted benchmark 12 times, hence 41% of the time examined.

Another measure of the consistency of the rule's performance over time would be obtained by analysing the returns of the portfolios on a rolling basis, meaning to hold the portfolios for various overlapping horizons. We find that over a three-year rolling period, the OSH2 strategy beat the FTA index 48% of the time. Thus, investors buying the OSH2 portfolio would out-perform the benchmark 48% of the time, with a return 11% over the FTA, provided that they implemented the strategy and held the stocks for three years.

Conducting a similar analysis but with a longer rolling period of ten years reaffirms the mediocre performance of the OSH2 strategy as it only beat the benchmark for 7 out of 20 periods. Over the whole sample period, the OSH2 strategy underperformed the benchmark by 24%. The performance of OSH2 clearly deteriorates as the holding periods increase.

One measure of the riskiness of the OSH2 strategy and its performance under different circumstances was obtained by examining the portfolio's performance during different states of the market (as measured by the FTA Index). The first state comprises the benchmark's worst return years, comprising all the negative returns. The remaining 3 states were defined by the benchmark's level of returns, that is from zero to 15%, from 15% to 25% and returns of above 25%. The average returns of each portfolio for each separate state are shown in Table 5.

During the worst return years, the OSH2 strategy exposed investors to less downside risk as it out-performed the benchmark by 9%, as visible from the fourth column of Table 5. The OSH2 strategy also out-performed during the intermediate states when the benchmark enjoyed returns of between 0% and 25%. As expected, the OSH2 strategy did relatively worse during the best years, since it had been less risky during poor market conditions. The average under-performance of OSH2 for the FTA's best

years is 16%. Tests of significance were performed on the under or over-performance of the OSH2 strategy over the FTA in each state. However, none of these results are statistically significant.

4. The Data Snooping Test

Data-snooping or data-mining involves searching through databases for correlations and patterns, and is considered a notorious bias in assessing a trading strategy. When a given set of data is reused over and over again, the probability of finding a satisfactory result is high. As stated by Jensen and Bennington (1970), "…given enough computer time, we are sure that we can find a mechanical trading rule which works on a table of random numbers – provided of course that we are allowed to test the rule on the same table of numbers".

The pitfalls of data-snooping were again highlighted in an article by Washington (1998). She gave an example of David Leinweber's (managing director of First Quadrant, LP, Pasadena, California) reported search for the best predictor of the S&P500 using the United Nations CD-ROM. As it turned out, Leinweber discovered that butter production in Bangladesh was historically the best predictor.

It is difficult to determine whether an apparently significant relationship or profitable opportunity (from back-testing) occurred as a consequence of chance alone or whether the relationship has genuine predictive capability. Data-snooping could arise from researchers torturing a set of data as well as a consequence of survivorship bias operating on the entire universe of technical trading rules that have been considered historically. The latter, also known as inter-generational data snooping, would occur over time as profitable trading rules become popular and remain within the investment community while the bad ones fade away. This is particularly apparent in the case of technical trading rules. Technical trading rules, without any fundamental basis, are susceptible to inter-generation data-snooping. Investors would apply a certain technical rule without much thought if they know that it had worked in the past for the creator of that rule. Technical trading rules are easy to operate and to improvise, thus the probability of finding a variant to an original rule that works would be very high.

The negative effects of data-snooping have been acknowledged by Lo and MacKinlay (1990). In their research, they found that tests of financial asset pricing models (e.g., the CAPM) may yield misleading inferences when properties of the data are used to construct the test statistics. The properties mentioned relate to the construction of portfolios based on some empirically motivated characteristic of the stock such as market capitalisation. The test of financial asset pricing models performed on the returns of these portfolios may create potentially significant biases in the test statistics. In this case, data-snooping occurs when researchers sort portfolios according to how this was done in previous studies, thus the term "empirically motivated characteristic". When the authors sorted the portfolios based on market capitalisation (empirically motivated), the null hypothesis of zero-intercept (excess return over market premium) was rejected whereas sorting on theoretically relevant characteristics such as dividend yield did not result in rejection. To this end, the authors suggest that data-instigated grouping procedures should be employed cautiously.

Kahn (1998) suggests four guidelines on how we can safeguard against data mining biases. The *intuition* guideline demands the investigation of only those strategies that are plausible and with some *ex ante* expectation of success. Research work should never involve free-ranging searches for patterns without regard for intuition. The *restraint* guideline attempts to minimize the number of strategies investigated. Researchers should decide *ex ante* exactly which strategies to investigate, run their tests, and look at the answers. They should not go back and continually refine their investigations. The *sensibility* guideline deletes results that seem improbably successful. Extremely large observed *t*-statistics may signal database errors or an improper methodology rather than a new and plausible strategy. Finally, *out-of-sample testing* is the statistician's answer to the curse of data mining. Coincidences observed over one data set are quite unlikely to reoccur in another independent data set. Most current studies on investment strategies do implement out-of-sample testing in their research.

Sullivan, Timmermann and White (1999) (hereafter STW) argue that genuinely meaningful out-of-sample testing is difficult to carry out as the exact identity and functional form of the model to be tested is not suggested by theory; for example, the

identification of factors and anomalies in cross-sectional tests of asset pricing models. STW suggest a bootstrap methodology that would be able to quantify the data-snooping bias.

4.1 How the Bootstrap Procedure Works

Why use the bootstrap? In simple situations, we can gauge the uncertainty of an estimate by analytical calculation based on an assumed probability model. However, in more complicated problems, this approach would be difficult and possibly inoperable. Its results may potentially be misleading if inappropriate assumptions or simplifications have been made. Bootstrapping essentially resamples the original data to create replicate datasets, from which the variability of the quantities of interest can be assessed without long-winded and error-prone analytical calculation. A wide range of statistical problems can be tackled by bootstrapping, thus we can avoid having to oversimplify complex problems.

For instance, assumptions required to validly construct t-ratios (normal, stationary and time-independent distributions) can be relaxed, which is clearly desirable since stock returns exhibit leptokurtosis, autocorrelation, conditional heteroscedasticity and changing conditional means. Thus, the bootstrap can accommodate fat tails, jumps and other departures from the normal distribution. It should be noted that the bootstrap method assumes that entities being resampled (the returns) are independent. This can be overcome using a technique known as the moving block bootstrap, although we do not explore this issue further here. A further potential limitation of the bootstrap is that the bootstrapped distributions may be a poor approximation to the actual distributions when utilising a small sample size.

In carrying out their research, STW utilised the technical trading rules from Brock, Lakonishok and LeBaron's (1992) (hereafter BLL) study. BLL had been unable to comprehensively account for data-snooping biases. They mitigated the problem by emphasising the robustness of results across various non-overlapping sub-periods, reporting results from 26 trading rules and utilising a very long data series (1897 to 1986). STW developed the Reality Check bootstrap methodology that would enable researchers to conduct a comprehensive test across rules (any number) in a manner not previously possible. This test would evaluate the performance of the trading rules

in terms of whether any apparent superior performance is a result of superior economic content or just due to luck.

Since data-snooping could arise from survivorship bias of the trading rules from the entire universe, STW decided to expand BLL's 26 rules into a universe of nearly 7,846 trading rules. STW's results would be deemed more plausible as their research is not based solely on the subset of surviving trading rules that may be misleading since it does not account for the full set of initial trading rules (failed and unpopular).

The Reality Check bootstrap is basically a nonparametric bootstrap. Briefly, nonparametric analysis assumes that $X_1,\ldots,X_n$ are independently and identically distributed according to an unknown distribution function F. Equal probabilities n^{-1} are placed on each sample value, thus each simulated sample point, X^*, is independently sampled at random from those data values. Thus the simulated sample $X_1^*,\ldots,X_n^*$ is a random sample taken with replacement from the data.

The objective of the reality check is to evaluate the distribution of a suitable performance measure (the absolute return or Sharpe ratio) giving consideration to the full set of models that led to the best-performing trading rule. The test is based on the following l x 1 performance statistic:

$$f = n^{-1} \sum_{t=R}^{T} \hat{f}_{t+1} \qquad (1)$$

where :

$\hat{f}_{t+1} = f(Z_t, \hat{\beta}_t)$, Z_t = vector of dependent and predictor variables, $\hat{\beta}_{t+1}$ = vector of estimated parameters, n = number of prediction periods, thus, $n = T - R + 1$, l = number of technical trading rules.

The f equation above differs from White's statistic of $\overline{f} = P^{-1} \sum_{t=R}^{T} f_{t+\tau}(\hat{\beta}_t)$ because estimated parameters are not needed for this application. Returns are generated through the parameterisations of the trading rules ($\hat{\beta}_k$, k=1, ... l). Technical trading rules, k, are assessed by the following equation:

$$f_{k,t+1} = \ln[1 + y_{t+1}S_k(\chi_t, \beta_k)] - \ln[1 + y_{t+1}S_0(\chi_t, \beta_0)], \ k = 1,....l \qquad (2)$$

where $\chi_t = \{X_{t-i}\}_{i=0}^{R}$ is the Original Dow Jones price series (the FTA series in our

case), $Y_{t+1} = \dfrac{(X_{t+1} - X_t)}{X_t}$ is the per period return, $S_k(\cdot)$ and $S_0(\cdot)$ are signal functions

that convert price index information, χ_t into market positions.

The null hypothesis to test whether there exists a superior technical trading rule is that the performance of the best rule from the universe is no better than the performance of the benchmark. Rejecting the null hypothesis below means that the best rule achieved significantly superior performance to the benchmark.

$$H_o : \max_{k=1,..l}\{E(f_k)\} \leq 0$$

The null hypothesis can be evaluated by applying the bootstrap to the observed values of returns, $f_{k,t}$. Resampling the returns from the trading rules will yield B

bootstrapped values of $\bar{f}_k$, denoted as $\overline{f}^{*}_{k,i}$. The statistics that measure performance for the actual rules in the universe and for the bootstrapped rules are respectively is constructed as

$$\overline{V}_l = \max_{k=1,..l}\left\{\sqrt{n}\left(\bar{f}_k\right)\right\} \qquad (3)$$

$$\overline{V}^{*}_{l,i} = \max_{k=1,..l}\left\{\sqrt{n}\left(\bar{f}^{*}_{k,i} - \bar{f}_k\right)\right\}, i = 1,..,B \qquad (4)$$

To summarise, the effects of data-snooping operating over time and across many investors and researchers, can only be quantified when one considers the performance of the best trading rule in the context of a full universe of trading rules form that this rule originated. Thus, the former equation (3) selects the trading rule with the highest average return. Equation (4), the order statistic would select the rule with the largest difference between te average bootstrapped value and the actual return for each i bootstrap series.

White's Reality Check p-value for the null hypothesis can be obtained by comparing $\overline{V}_l$ to the quantiles of $\overline{V}^{*}_{l,i}$ (the order statistic). This is achieved by sorting the order

statistics as $\overline{V}_1^*, \overline{V}_2^*, .., \overline{V}_B^*$. Then find b such that $\overline{V}_b^* \leq \overline{V} < \overline{V}_{b+1}^*$. Hence, the Reality Check p-value is written as

$$P_{RC} = 1 - \frac{b}{B} \tag{5}$$

The Reality Check p-value incorporates the effects of data-snooping since it accounted for the maximum value over all the l trading rules.

4.2. Steps Involved in the Implementation of the Bootstrap Procedure for OSH2 and Related Rules

1. Set an observation period, n, for $t = R$, …, T. In our application, there are 29 observation periods, starting with the year 1970 to 1998.

2. Next, the percentage returns, incorporating both capital gains and dividends are calculated for each stock. Percentage returns are used because it would allow the computation of average returns of a portfolio of stocks for each time space. For the benchmark, we utilised the average returns of all the listed stocks with available return values. Thus, both the returns of individual stock and the equally weighted benchmark are determined for each t. Using the specified model specifications, $f_{k,t+1}$, derive the over or under-performance of each l trading rules over the benchmark for n periods.

3. Construct and average the performance criterion to obtain, $\overline{f}_k = n^{-1} \sum_{t=R}^{T} f_{k,t+1}$

4. Develop the universe of trading rules. We have expanded the suggested trading rule by O'Shaughnessy of selecting the 50 highest RPS stocks with a market capitalisation below a deflated £100m and an EPR of more than 5% into 3,864 variants of the original trading rule. As with the original rule, we applied the deflated market capitalisation to a pre-set range of £5m to £1bn for the universe. The EPR range from 0 to 20% and the portfolio size examined varies from 1 stock to 100 stocks. The variable elements responsible for the creation of the universe are displayed below.

EPR	Portfolio Size	Market cap*
0.000	1	5000
0.005	5	25000
0.010	10	50000
0.015	15	75000

0.020	20	100000
0.025	25	200000
0.030	30	500000
0.035	35	1000000
0.040	40	
0.045	45	
0.050	50	
0.055	55	
0.060	60	
0.065	65	
0.070	70	
0.075	75	
0.080	80	
0.085	85	
0.090	90	
0.095	95	
0.100	100	
0.150		
0.200		

In thousands of pounds

5. Specify the number of bootstrap resamples, *B*. In our case we set *B* = 500. White developed an *observation index* before constructing the bootstrap values for each trading rule. The *observation index* is created so that every 3,864 trading rules would be bootstrapped by the same sequence of random picks. Thus, for each trading rule, the bootstrap will be performed using the same sequence of the 29 random observations over 500 times. Since random selection is done by replacement, repetition of a certain observation could exist for a particular *i* series. Instead of creating the *observation index*, we fixed the random number seed in our program. This procedure would have the equivalent result as building the *observation index* as the same set of random numbers are used which will generate the same set of index numbers for each rule. For each *i* series, find the average of the bootstrap values for each trading rule, $\overline{f}^{*}_{k,i}$.

6. For each trading rule, set $\overline{V}_k = \sqrt{n}\,\overline{f}_k$ and $\overline{V}^{*}_{k,i} = \sqrt{n}\left(\overline{f}^{*}_{k,i} - \overline{f}_k\right)$, the latter being the average of each *i* series less the average return of that trading rule.

7. When we have all the values for all the trading rules, we could obtain the Reality Check *p*-values as described in the section above.

4.2 Results from Application of the Data Snooping Test

For the full sample period between 1970 and 1998, the best rule from the universe of 3,864 trading rules recorded a mean return of 25.38% over the FTA benchmark. This trading rule suggested the purchase of the top 20 ranked RPS stocks with a market capitalisation of more than £200m and an earnings yield, EPR, of above 20% (PER<5x). Since the best rule recorded a Reality Check p-value of 0.346 (fourth column of Table 6), we do not reject the null hypothesis and conclude that the performance of the best trading rule has no statistically significant predictive superiority over the given benchmark.

The magnitude of the data-snooping bias could be clearly identified once we produce the nominal *p*-value on the best trading rule. The nominal or the "data-mined" *p*-value can be obtained by applying the bootstrap methodology to the best rule only. By excluding the other rules we would ignore the effects of data-snooping. The nominal *p*-value computed for the best trading rule decreased tremendously as expected. The best trading rule would appear to be potentially credible with its low nominal *p*-value of 0.032. However, if we adjust for data-snooping biases, through the use of the Reality Check *p*-value, we can conclude outright that the best trading rule is not significantly superior. The difference between the two p-values represents the magnitude of the data-snooping bias on the performance measure. Using the Reality Check *p*-value, we can quantify precisely the evidence against the null hypothesis in a manner previously impossible. We can also quantify the extent to which data-mining can mislead the unsuspecting researcher.

The strategy suggested by O'Shaughnessy, OSH2, recorded a very insignificantly superior performance, with a nominal *p*-value of 0.3, and thus in actual fact we need not bother proceeding with computing the Reality Check *p*-value for the universe in this case.

Table 6 also displays the best trading rule for the two almost equally long sub-periods; sub-period 1 from 1970 to 1984 and sub-period 2 from 1985 to 1998. For both the sub-periods, rules that outperformed the benchmark could be constructed, but again these failed to deliver low Reality Check *p*-values. Thus in both sub-periods, the

performance of the respective best trading rules were not significantly superior to the benchmark.

4.3 Can a trader possibly gain from using the best trading rule?

Suppose now that a trader conducts a historical search to find the best rule with the greatest cumulative return and applies it in the subsequent year. By doing so, a trader could implement the current best strategy without the knowledge of how well it will do in the subsequent year whereby money is committed. Hence, we would analyse 28 years instead of 29 years of data due to this time lag.

Two different analyses were conducted, as there were occasions where more than one best rule existed. Consequently, the performance in the following year could differ between these trading rules. Thus, we examine the lowest returns (which we call the "Pessimistic scenario") and the highest returns (termed the "Optimistic scenario") from the range of past best trading rules.

From the second column of Table 7, it can be seen that the cumulative wealth rule underperformed the benchmark by 4% under the Pessimistic scenario and by 3% under the Optimistic scenario. Hence, even with the knowledge of the best cumulative return rule in one year, an investor will not be able to exploit it in the following year.

Suppose an investor implements the rule with the best previous year's returns. Would this make money? As can be seen in the fourth column in Table 7, this strategy will still underperform the benchmark by 3% under the Pessimistic scenario. However, this is a slight improvement from the cummulative wealth rule. The only trading rule that would produce excess return is the Optimistic scenario of the Previous Year's best rule. Investors could potentially earn excess return of 11%, although this return is not statistically significant.

5. Conclusions

This paper has investigated whether profitable trading strategies can be derived from popular books of investment strategy, with "What Works on Wall Street" by James P. O'Shaughnessy, used as an example. With an abundance of evidence on the effectiveness of momentum and low PER strategies in the US and the UK, the OSH2

strategy would be expected to perform well. To a certain extent, OSH2 was indeed successful, out-performing the FTA All-Share index over the past 30 years.

But how successful was the strategy in the UK compared with that delivered in the US? In the US context, OSH2 generated an average return of 18.52% per annum, which is 8.57% more than the Dow Jones Industrial Average and 7.97% more than the Standard and Poor's 500 benchmark. By contrast, OSH2 only managed to beat the FTA All-share by 2.59%, and viewed from this perspective, OSH2 would not have been nearly as successful in the UK as in the US.

We also applied a new test for the effects of data mining on the results from applying the trading rules. This test is based on a bootstrap procedure, and involves nesting the rule under test within a universe of similar rules based on the same criteria. We were able to determine that no trading rules from within the O'Shaughnessey family are able to yield statistically significantly higher (or significantly lower) returns than the benchmark. We believe that the approach taken here should be viewed as a general, easily interpreted and widely applicable approach to auditing new investment trading strategies.

References

Brock, W., Lakonishok, J. and LeBaron, B., 1992, Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, *Journal of Finance* XLVII, 1731-1764.

Buffett, M., 1997, Buffettology. Rawson Associates, New York.

Dreman, D., 1998, Contrarian investment strategies: the next generation. Simon & Schuster, New York.

Efron, B, 1982, The Jackknife, the bootstrap and Other Resampling Plans, Number 38 in CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia

Kahn, Ronald N, 1998, Data Mining is Easy, BARRA Newsletter, publication no.165, Winter, www.barra.com/Newsletter/NL165/SevIns5NL165.asp

Lo, A.W., and MacKinlay, C.A., 1990, Data-Snooping Biases in Tests of Financial Asset Pricing Models, *The Review of Financial Studies*, 3, 431-467.

Lynch, P., 1989, One up on Wall Street. Pengiun

Malkiel, B., 1996, A random walk down Wall Street, Norton, New York

O'Neil, W., 1995, How to make money in Stocks. Mcgraw-Hill, New York.

O'Shaughnessy, J.P., 1998, What works on Wall Street. Mcgraw-Hill, New York

Seigel, J.J., 1998, Stocks for the long run. Mcgraw-Hill, New York.

Sullivan, R., Timmermann, A. and White, H., 1999, Data-Snooping, Technical Trading Rule Performance, and the Bootstrap, *Journal of Finance* 54, 1647-1691.

Washington, L., 1998, What's the Stock Market got to do with the Production of Butter in Bangladesh?, Money, March, Vol.27, No.3 (www.pathfinder.com/money/archive/magarticle/0,4015,7818,00.html)

White, H., 2000, A Reality Check for Data Snooping, Econometrica, 68, 1097-1126

Table 1: Number of Firms Examined

Year	Sample firms	Rejected firms	Total firms after first filter*
69	250	901	1,151
70	271	814	1,085
71	173	1,255	1,428
72	207	1,223	1,430
73	374	1,012	1,386
74	589	764	1,353
75	454	1,794	2,248
76	1,042	1,109	2,151
77	785	1,261	2,046
78	1,112	914	2,026
79	1,360	679	2,039
80	1,263	513	1,776
81	1,075	682	1,757
82	914	810	1,724
83	855	847	1,702
84	847	878	1,725
85	914	797	1,711
86	793	971	1,764
87	813	980	1,793
88	921	890	1,811
89	1,045	771	1,816
90	1,122	668	1,790
91	1,041	684	1,725
92	803	851	1,654
93	544	1,142	1,686
94	670	1,115	1,785
95	789	1,129	1,918
96	854	1,189	2,043
97	848	1,282	2,130

Excludes secondary shares and shares belonging to the Trust Industry

Table 2: The Average Return of OSH2, Reciprocal Strategy & the Benchmarks.

	50 Lowest RPS	50 Highest RPS	FTSE All Share*	Equally-Weighted Benchmark
Average Return	0.3177	0.1690	0.1431	0.2066
Std Deviation	0.4628	0.2707	0.3097	0.3118
Sharpe Ratio	47.37	25.23	13.76	34.16

Datastream ICV; Figures are annualised and expressed as a proportion of the initial investment

Table 3:Replication Results & the Benchmark Performance

Performance Year	50 Lowest RPS	50 Highest RPS (OSH2)	FTSE All Share	Equally-Weighted Benchmark
70	-0.1238	0.0825	-0.0752	-0.0202
71	0.6616	0.7879	0.4193	0.5512
72	0.3427	0.4517	0.1282	0.2019
73	-0.2500	-0.1969	-0.3136	-0.3075
74	-0.5422	-0.4534	-0.5534	-0.5035
75	1.1468	0.5035	1.3633	1.0923
76	0.0860	0.2146	-0.0387	0.0242
77	1.0812	0.6539	0.4118	0.7914
78	0.3186	0.2552	0.0265	0.2411
79	-0.0334	0.0166	0.0435	0.0636
80	0.0742	0.0109	0.2717	0.0856
81	0.1160	-0.0422	0.0715	0.1639
82	0.1666	0.3505	0.2207	0.2445
83	0.5720	0.1505	0.2310	0.4262
84	0.1718	0.1763	0.2602	0.1904
85	0.1144	0.1264	0.1518	0.1988
86	0.2902	0.5758	0.2234	0.3267
87	0.3882	0.1795	0.0416	0.1597
88	0.0517	0.0602	0.0648	0.0992
89	0.1105	-0.0567	0.3001	0.1184
90	-0.3291	-0.2493	-0.1431	-0.2134
91	0.3189	0.4206	0.1506	0.2468
92	0.3649	0.1046	0.1483	0.1300
93	0.9867	0.3720	0.2335	0.6213
94	0.8844	-0.0526	-0.0955	0.1103
95	0.1099	0.1283	0.1851	0.3439
96	0.8779	0.1520	0.1168	0.4588
97	1.3581	0.0974	0.1973	0.1633
98	-0.1025	0.0800	0.1091	-0.0187

Table 4: Return Difference from using different minimum market capitalisation and averaging over 5 years.

Replicating OSH2 using deflated Market Capitalisation of £5m

	Highest 50 RPS
Average Returns	0.1797
Std Deviation	0.2752
Sharpe Ratio	0.2920
Compounded Return	0.1471

Replicating the OSH2 using the 5 year average deflated market capitalisation of £100m

	Highest 50 RPS
Average Returns	0.1702
Std Deviation	0.2727
Sharpe Ratio	0.2599
Compounded Return	0.1380

Table 5: Portfolio Performance during different States of FTA

Different States	OSH2	FTA	OSH2-FTSE	Respective Years	t-statistic at 5% Level
< 0% (worst returns)	-0.1092	-0.2033	0.0941	74,73,90,94,70,76	1.9116
0% - 15% returns	0.1397	0.0833	0.0564	78,87,79,88,81,98,96,72,92	1.1777
15% - 25% returns	0.2777	0.1992	0.0785	91,85,95,97,82,86,83,93	1.3005
> 25% (best returns)	0.3460	0.5044	-0.1584	84,80,89,77,71,75	-0.8741

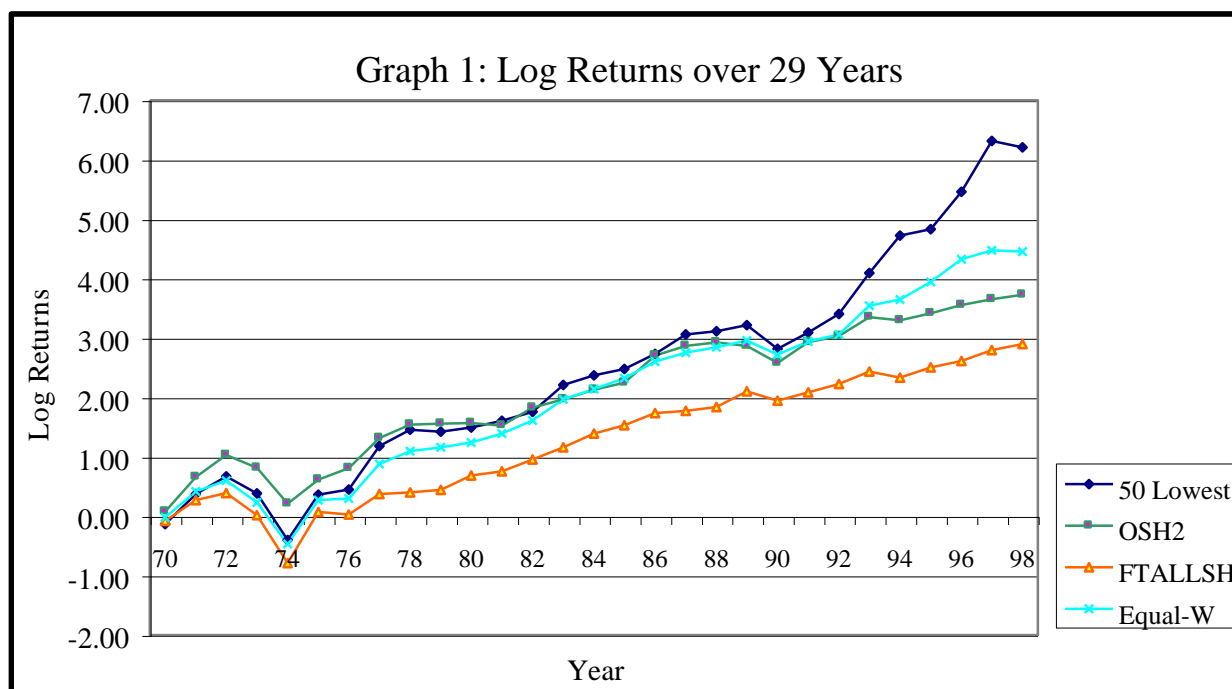**Table 6: Summary Statistics for the Performance of the Rules in the Universe**

Best performing rule :	Subperiod 1970-84	Subperiod 1985-98	Full Sample 1970-98
Market cap above	£1,000m	£200m	£200m
Portfolio size	1 stock	20 stocks	20 stocks
EPR above	20%	20%	20%
Mean over benchmark	0.1852	0.4454	0.2538
Reality Check p-value	0.6520	0.2440	0.3460
Nominal p-value	0.0720	0.0440	0.0320

**Table 7: Potential Profits from implementing the universe of rules**

	Cummulative Wealth Rule		Implementing Previous Year's Best Rule	
	Pessimistic	Optimistic	Pessimistic	Optimistic
Mean over benchmark	-0.0382	-0.0268	-0.0172	0.1126
Reality Check p-value	n/a*	n/a*	n/a*	n/a*
Nominal p-value	0.6000	0.5580	0.5580	0.1160

* Does not apply because the Cummulative Wealth Rule & the Previous Year's Best Rule are not the best trading rule ex-post.
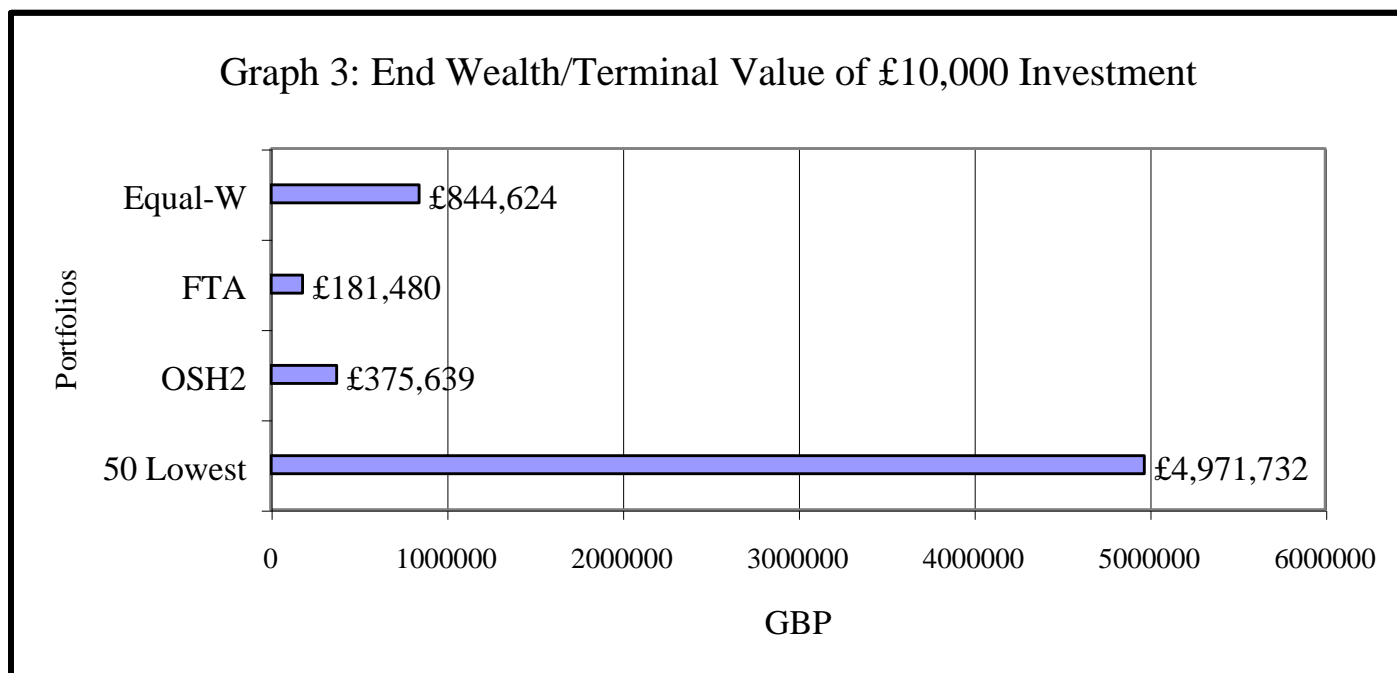
Graph 1: Cumulative Log Returns of OSH2 Strategy

Graph 1: Log Returns over 29 Years

Graph 2: Cummulative Returns

Graph 2: Cummulative Returns over 29 Years

26
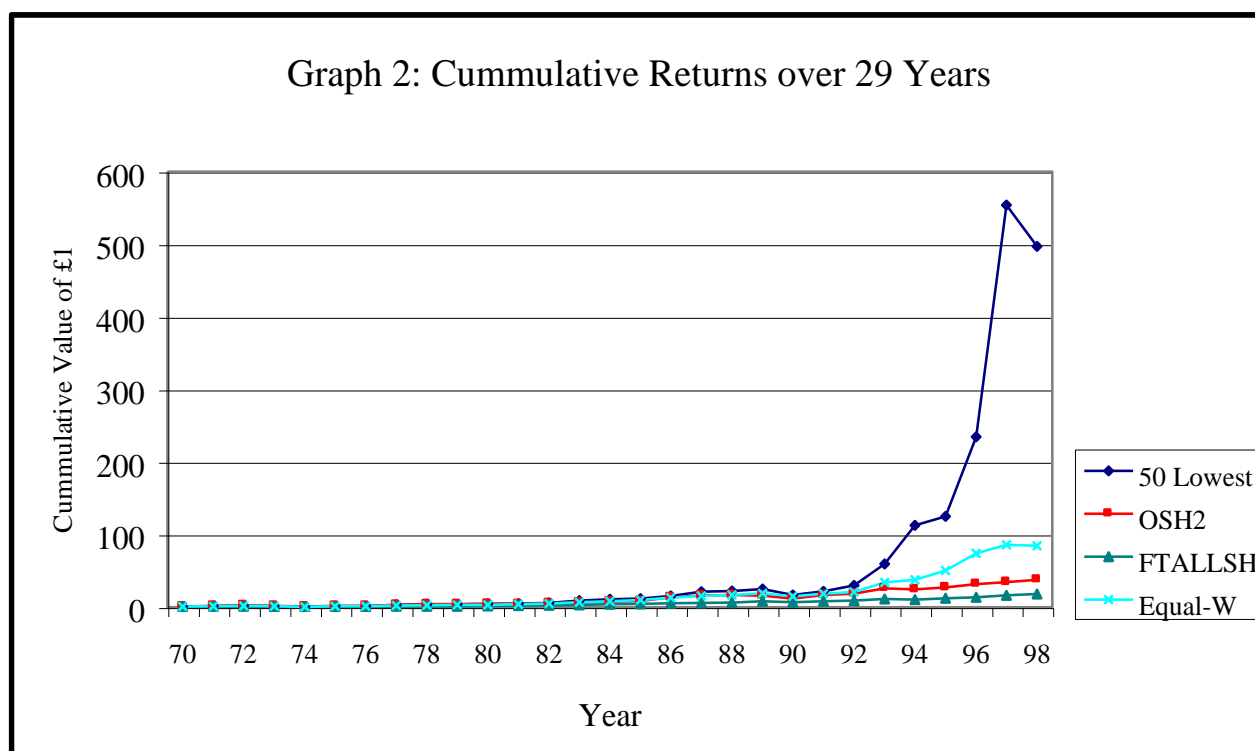
Graph 3: Terminal Value of the OSH2 Portfolio

Graph 3: End Wealth/Terminal Value of £10,000 Investment

Graph 4: Yearly Performance of the OSH2 portfolio relative to the Equally – Weighted Benchmark & the FTA