
Does Survivorship Bias really matter?
An Empirical Investigation into its Effects on the Mean
Reversion of Share Returns on the JSE Securities
Exchange (1984-2006)

EVAN GILBERT AND DAVE STRUGNELL

Stellenbosch Economic Working Papers: 19/08

KEYWORDS: SURVIVORSHIP BIAS, MEAN REVERSION, P/E RATIO
JEL: G10, G14

EVAN GILBERT
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: EGILBERT@SUN.AC.ZA

DAVE STRUGNELL
DEPARTMENT OF ACTUARIAL SCIENCES
UNIVERSITY OF CAPE TOWN
E-MAIL: DAVE.STRUGNELL@UCT.AC.ZA



UNIVERSITEIT
STELLENBOSCH
UNIVERSITY



A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

Does Survivorship Bias really matter? An Empirical Investigation into its Effects on the Mean Reversion of Share Returns on the JSE Securities Exchange (1984-2006)

EVAN GILBERT AND DAVE STRUGNELL¹

ABSTRACT

October, 2008

This paper tests for the impact of survivorship bias by building on the work of Cubbin, Eidne, Firer and Gilbert (2006), and Bailey and Gilbert (2007). The former paper confirmed the existence of mean reversion on the JSE Securities Exchange, because portfolios of shares with high Price to Earnings (P/E) ratios (being those which had tended to outperform recently) underperformed significantly over five years against portfolios of shares with low P/E ratios. The latter paper developed the economic validity of this conclusion by applying liquidity constraints to portfolio formation. This tended to slightly dampen the observed effects, but confirmed the significant presence of mean reversion. In both cases, extensive efforts were made to include all delisted shares in the study to avoid the effects of survivorship bias. This paper updates both studies by extending the period for a further 21 months, and then quantifies the impact of survivorship bias by comparing the results against those of an equivalent study based on a data set of currently listed shares only. The results of our study confirm that the effects of survivorship bias are present and material. While patterns of mean reversion are detected on both data sets, the returns earned on portfolios selected from currently listed shares are significantly higher than the corresponding returns on portfolios selected from all shares. Survivorship bias is therefore confirmed to be a significant issue in such studies, which researchers should be careful to avoid; although it does not necessarily affect the conclusion of the patterns of mean reversion revealed in the earlier studies.

Keywords: Survivorship bias, mean reversion, P/E ratio
JEL codes: G10, G14

¹ Department of Economics, University of Stellenbosch and Department of Actuarial Science, University of Cape Town, respectively. The authors would like to thank Paul Van Rensburg for suggesting the study; Greg Bailey for his comments on the methodology; Johan Wiid for his research contribution; I-Net Bridge for access to the data used; and especially Lesley Booysen for attempting to resolve our multitude of data-related queries.

1. Introduction

One of the most common challenges facing financial researchers in emerging markets is the lack of a clean and comprehensive data set of price and accounting data for listed firms. There may be some historical data available for currently listed companies – but historical data availability for delisted shares is always an issue.

It is well established in financial research that ignoring delisted companies when conducting historical research leads to the presence of survivorship bias. As Bain (1972: 105) asserts, when commenting on a paper by Wagner and Lau (1971): "the use of ex-post sampling will invariably produce an upward bias in the measurement of returns on risky securities". This bias results from the use of a data set that consists of the survivors over a period, not the full set of companies that were listed over this period. As the characteristics of the survivors are likely to differ systematically from those who have delisted, the results of such a study will be biased. Given that collecting data for delisted companies is a time-consuming and expensive process, obvious questions are does it really matter, and should researchers attempt to correct for this problem?

A review of the international evidence suggests that survivorship bias does matter. However, no attempt to answer this question has been made for South Africa. This paper attempts to address this gap by building on the empirical work reported in Cubbin, Eidne, Firer and Gilbert (2006), and Bailey and Gilbert (2007).

The former paper confirmed the existence of mean reversion on the JSE Securities Exchange, because portfolios of shares with high P/E ratios (being those which had tended to outperform recently) underperformed significantly over five years against portfolios of shares with low P/E ratios. The latter paper developed the economic validity of this conclusion by applying liquidity constraints to portfolio formation. This tended to slightly dampen the observed effects, but confirmed the significant presence of mean reversion.

In both cases, extensive efforts were made to include all delisted shares in the study to avoid survivorship bias. This paper updates both studies by extending the period for a further 21 months, and then quantifies the impact of survivorship bias by comparing the results against those of an equivalent study based on a data set of currently listed shares only.

The results of our study confirm that the effects of survivorship bias are present and material. While patterns of mean reversion are detected on both data sets, the returns earned on portfolios selected from currently listed shares are significantly higher than the corresponding returns on portfolios selected from all shares. Survivorship bias is therefore confirmed to be a significant issue in such studies, which researchers should be careful to avoid; although it does not affect the conclusion of the patterns of mean reversion revealed in the earlier studies.

Section two of this paper will review the evidence for the importance of survivorship bias in empirical studies. Section three will outline the methodology used in this study and section four will present the results. Section five concludes.

2. Literature Review

The presence and possible effects of survivorship bias has concerned researchers in this field, particularly in the United States (US). The primary data sources used in the US for finance-related research are the COMPUSTAT data base (for accounting data) and the data prepared by the University of Chicago's Centre for Research into Securities Prices (CRSP) (for price data). While the CRSP data set does not suffer from significant survivorship bias issues, the way the COMPUSTAT database is created does lead to this problem². Firstly, the accounting data for companies that are delisted are deleted from the database. Secondly, new companies added to the dataset are included with a full history, which means that companies that do not succeed are not included.

Researchers became aware of the presence and potential effects of survivorship bias in the early 1970s when looking at the question of using accounting data in explaining the cross-section of share returns. Initial tests were limited and focused on establishing whether the conclusions of previous studies were robust (Ball and Watts, 1979; Salamon and Smith, 1977). The first comprehensive examination of, and test for, the effects of this bias only happened later (McElreath and Wiggins, 1984; Banz and Breen, 1986, Davis 1994, 1996). The question of the impact of survivorship bias was also discussed in the context of establishing the persistence of portfolio manager returns (Grindblatt and Titman, 1989; Brown, Goetzmann, Ibbotson and Ross, 1992; Carpenter and Lynch, 1999).

Ball and Watts (1979) refer to a study they completed in 1972 where they critically evaluated the time series properties of Earnings Per Share (EPS) for listed companies. In their initial study, they recognised the potential for the existence of survivorship bias given the problems in the way COMPUSTAT data was created, but did not know how to correct for it. In their later, paper they present data that suggests that survivorship bias would not have had a significant impact on the conclusions of their earlier research. They reached this conclusion by randomly selecting 25 shares that were in existence in 1916 and 25 shares that were in existence in 1966. They compared the characteristics of the EPS time series for these shares with 25 shares which had been in existence from 1916 to 1966. They found no significant difference in the results for these two samples.

² Another problem with the COMPUSTAT database is the presence of the 'look ahead' bias. This occurs when data are recorded as being available to investors at a particular time, when it actually only becomes available at a later stage. For example, the annual financial statements (or announcement of the earnings) for a company may only be made public several months after its year end. However, in the COMPUSTAT database, these earnings would have been recorded as applying to the whole financial year, i.e., they were reported immediately.

McElreath and Wiggins (1984) explore more comprehensively the potential impact of survivorship bias in the COMPUSTAT data set. They reviewed the reasons for the delisting of the 330 firms that had left the New York Stock Exchange (NYSE) in the period 1970 to 1979 (approximately 1800 shares were listed on the NYSE at the end of the 1970s). They concluded that the likely size of the bias due to survivorship was not primarily due to the fact that 55% of these delisting were due to mergers. Also, bankruptcy and liquidations only accounted for 6% of the delisted firms. No further quantitative comparative analyses between this group of firms and those still listed was presented.

Banz and Breen (1986) present the first comprehensive, direct test of the size and impact of survivorship bias on the COMPUSTAT database. In a very similar vein to this paper, they compare the nature of the properties of two separate populations of firms – a complete and a partially complete (currently listed only) COMPUSTAT list of firms. For the period 1974 to 1981 they collected accounting data on a monthly basis for all listed firms on the NYSE and the American Stock Exchange (AMEX). They then compare the differences in return of similar equally weighted portfolios created from this list and the COMPUSTAT dataset. The portfolios were created on the basis of size (quintiles), and the firms with positive earnings then were ranked in terms of Earnings Yield (Earnings/Price)³. They use a Seemingly Unrelated Regression approach to test for significant differences in the returns of all 30 portfolios. The differences in returns were found to be significant at the 1% level for all portfolios whether the raw or risk-adjusted portfolio returns were compared.

They point out that this difference in returns is actually the combined effects of the survivorship and "look ahead" biases. To isolate the effects of the differences due to survivorship bias, they create a subset of firms that are included in the partial COMPUSTAT data series from their complete list. They then recreate the portfolios as explained above using their complete list and this subset of their complete list. In other words, the effects of the differences in the portfolio returns could only be due to survivorship bias. Again, the differences in returns are statistically significant. It is interesting to note, however, that the returns from the portfolios created from the complete list were systematically greater than those created from the smaller list. This is the opposite result to the expected effect as summarised by Bain (1972) above.

Banz and Breen (1986) also evaluate the effects of the survivorship bias on the results of studies investigating the presence of size and P/E effects on portfolio returns. They find that using the complete data series (i.e., corrected for the missing firms) leads to the rejection of earlier claims of a P/E effect on returns when size is controlled for. They conclude that survivorship bias does seem to matter.

Davis (1996) tests for the effects of survivorship bias on the results of his earlier study (Davis, 1994) that used COMPUSTAT data in part. He identifies, and then directly compares, the nature of the firms listed on the NYSE and the AMEX that were excluded from the COMPUSTAT data set to those in the set. He finds that

³ The firms with negative earnings were allocated to a sixth portfolio for each size quintile.

the nature of the excluded firms is systematically different in terms of relative size (the excluded firms are smaller) and monthly returns (the excluded firms have lower returns). While the inclusion of the shares listed in the Moody's database (but not in the COMPUSTAT database) did not change the conclusions of his previous study, it did lead to non-trivial difference in the regression results, both in terms of economic and statistical significance.

Kothari, Shanken and Sloan (1995) confirm the importance of the effects of survivorship. They firstly establish that the returns to the share excluded from the COMPUSTAT database (but on the CRSP data series) are on average 9 to 10 percentage points lower than the shares included in the database. This emphasises the systematic difference of the nature of the excluded shares. They also show that the significant Book to Market Value (B/M) result of Fama and French (1992) can, in part, be explained by the survivorship bias in the data set they used⁴.

When evaluating the persistence of portfolio manager's outperformance, Grindblatt and Titman (1989) review the relative performance of US portfolio managers for the period 1974 to 1984. They test for the potential size of survivorship bias by comparing the results of the entire universe of 274 funds with the sample of 157 surviving funds. They find that the effects are small, between 0.1% and 0.4% of gross returns per year, and statistically insignificant.

Brown, Goetzmann, Ibbotson and Ross (1992) conclude that survivorship bias will definitely have an effect on the results, but the size of this effect is an empirical question. Carpenter and Lynch (1999) indicate that 0.14% to 0.27% of gross management outperformance (alpha) returns can be attributed to survivorship bias.

Our study adds to this literature in two ways: firstly by testing for the effects of survivorship bias in a mean reversion context; and secondly, by using data for South Africa for the first time.

3. Research Methodology

De Bondt and Thaler (1985, 1987) first examined the question of the existence of mean reversion in returns for the US. They identified 'winner' and 'loser' portfolios based on past abnormal returns. They then tracked the relative performance of these portfolios. The research of Cubbins, Eidne, Firer and Gilbert (2006) was the first study to comprehensively establish the presence of mean reversion of share returns in South Africa. Their approach differed from that of De Bondt and Thaler (1985). Their 'winners' represented firms with high P/E ratios and their 'losers' represented firms with low P/E ratios.

⁴ The other reason offered for their finding that the B/M result of Fama and French (1992) was not valid was that their dataset was significantly longer (1927 to 1990) than that used by Fama and French (1941 to 1990). In other words, they believe that Fama and French's results were period specific.

Bailey and Gilbert (2007) extended the results of Cubbins *et al* (2006) by applying liquidity caps to 'High P/E' and 'Low P/E' portfolios in an attempt to evaluate the economic reality of the abnormal returns seemingly offered by the presence of mean reversion. They tested for effects of liquidity constraints on the presence of mean reversion in multiple portfolios by applying liquidity caps. Depending on the value of a portfolio, a share would only be considered for inclusion if its average monthly traded volume was sufficiently large. They concluded that, although dampened, mean reversion persists after application of liquidity constraints.

This paper builds on this work, by testing for the presence and effects of survivorship bias by applying the Bailey and Gilbert methodology to two separate groups of firms: the complete list of firms and the list of firms that are currently listed on the JSE Securities Exchange.

We replicated the methodology used by Gilbert and Bailey (2007) using historic share data collected from I-Net Bridge. The key advances are: firstly, we extended the period under consideration by 21 months; and secondly, we conducted the analysis on two separate groups of shares: the currently listed shares and the complete list of shares (i.e., including the delisted shares). By comparing the results of the two groups of shares, we can ascertain the effects of the lack of the delisted shares on the results of this study.

Historical data for shares traded on the JSE at each month end was obtained from I-Net Bridge. Month end price, Earnings Yield (from which P/E ratio was derived) and dividend yield data were collected for all shares listed on the All Share Index (ALSI) for the JSE for the period 31 October 1984 to 31 September 2007 – a 23 year period⁵.

A list of delisted shares was obtained from I-Net Bridge⁶ and all the data for these shares was included in the dataset for this period. In total, 1631 shares were included in the analysis, of which 841 had either been delisted or were removed due to some other type of corporate action.

The data presented in Figure 1 highlights the existence and size of the difference in the two groups of shares. The dark shaded area represents the shares that are currently listed (the 'current' list) and the light shaded area indicates all the shares that were listed (the 'complete' list). The area between the two parts represents the shares that would have been excluded from the analysis if only the currently listed shares were included in the analysis. The relative variation of the two lists ranged from 0% (in the final month – by definition) to 80% (in late 1992).

INSERT FIGURE 1 ABOUT HERE

⁵ We excluded preference shares ('P' series) from the analysis, because they have a different dividend risk profile.

⁶ We have used the old code list from I-net Bridge to identify the delisted shares. While this is the most complete and accurate list of 'dead' shares that we were able to obtain, the jumps in the number of firms covered by our analysis (see Figure 1) suggest that we still do not have a comprehensive list.

While I-Net Bridge was the most complete and accurate data source that we were able to use, we still encountered significant data completeness and quality issues. The presence of the discontinuous jumps in the number of total shares in September 1991, November 1992 and September 1997 in Figure 1 suggest that the list of 'dead' shares from I-Net Bridge is still not completely accurate. Moreover, the P/E Ratio data was found to be unreliable so we used the inverse of Earnings Yield (EY) ratio that had been recorded more extensively and more accurately in the dataset. We also encountered obvious errors in share price and dividend yield data. Where found, these observations were deleted. The reasons for these errors have not been fully resolved and one of the project's conclusions is that we need to consolidate and clean this historical database through comparisons with other sources (e.g., McGregors BFA Net, Datastream or Reuters).

Following Bailey and Gilbert (2007), we adjusted for liquidity by applying a liquidity cap. We assumed an investor starts with a specific amount of money to be invested in the portfolio. For the sake of illustration, assume this is R1 billion. As every portfolio is made up of 35 equally weighted shares, the investment required in each share is equal to the value of the portfolio (R1 billion) divided by the number of shares in the portfolio (35). Approximately R28.5m is thus invested in each share. It is at this point that the liquidity cap comes into consideration. For the share to be considered for inclusion in the portfolio, 50% of the average monthly value of shares traded must exceed this threshold of R28.5m, suggesting that it would be possible for the portfolio manager to unwind a position in that share within a reasonable time period. This is calculated by taking the average volume of shares (which is based on the previous 12 months' volume traded) multiplied by the current share price and then dividing by two. If, for example, the average number of shares traded each month is 500 000 and the current share price is R10 then the liquidity cap for this share is $500\ 000 \times R10 \times 50\% = R2.5$ million.

In this example, the share would be filtered out due to its lack of liquidity for an investor of this size (R28.5m investment > R5,5m liquidity cap). However, if the investor had a portfolio of R10m, then this share would be considered available for investment ($R10m/35 = R0,285m$ holding per share which is less than the R2,5m liquidity cap).

Using the approach outlined above, a liquidity constraint can be derived from portfolio size, share price and average volume traded. As explained above, the impact of the liquidity cap is dependent on the size of the portfolio. Seven different portfolio sizes were evaluated: R0 (i.e., no liquidity constraint), R100 000, R1 million, R10 million, R100 million, R1 billion, and R10 billion.

These portfolio sizes reflect nominal values as at the end of the portfolio formation period of the study (September 2002). To ensure that portfolio sizes were comparable in real terms, the nominal size of portfolios was discounted by the monthly inflation rate for portfolios formed prior to September, 2002. The Consumer Price Index (CPI) for South Africa for the entire period of the study was

downloaded from International Financial Statistics website⁷ which is run by the International Monetary Fund.

The liquidity cap was calculated using this methodology for all shares for a given portfolio size. Those shares which did not meet the liquidity cap were excluded from the rest of the portfolio construction process.

The shares that met this liquidity requirement were then ranked by their P/E ratios at the portfolio formation date and portfolios consisting of the top 35 (high P/E) and bottom 35 (low P/E) shares were constructed⁸. These shares were weighted equally and the excess monthly returns relative to the ALSI index returns were then calculated for 5 years (60 months) from the portfolio formation date. The start date was then incremented one month forward and the process repeated. Applying this process over the entire period allowed for the creation of 5,184 portfolios and 311,040 monthly returns.

The monthly returns for each high P/E and low P/E portfolio were then combined and averaged to give the final average relative returns for each category. What this means is that the returns for month one of all the high P/E portfolios were combined to give the average high P/E portfolio return for month one. The same exercise was conducted for the low P/E portfolios to give the average returns for each month for these portfolios.

Following the approach adopted by Cubbins, Eidne, Firer and Gilbert (2005) and Gilbert and Bailey (2007), our monthly portfolio returns are calculated as the sum of the capital gain (change in price/starting price) and dividend yields (DY) (with the latter converted to an equivalent monthly rate of return). This was done because DY was recorded on a monthly basis, but it is obviously a proxy as the dividends are not received on an equal monthly basis over the 12 months of the year. Due to lack of data on the timing of the dividend payments, this is the next best solution.

One significant problem that we encountered with the implementation of this approach was the very high dividend yields associated with shares whose prices fell rapidly and approached zero. As the portfolios are equally weighted, the portfolios returns are the average of the individual component shares' returns. Thus if uncorrected, these outliers led to some portfolios having extremely high returns. To correct for this, we made all DY greater than 150% equal to zero. This minimised the effects of the real outliers on the portfolio returns⁹ reported in this analysis.

⁷ <http://www.imfstatistics.org/imf/>

⁸ There were some months when there were less than 70 shares in the overall portfolio (especially when liquidity constraints were introduced). The high P/E and low P/E portfolios were formed in this case by taking the list of shares available (ranked by P/E) and dividing the shares into two equally sized groups.

⁹ We recognise that the use of 150% as the cutoff value is an arbitrary decision. We arrived at this number by considering the tradeoff between the effects of the cutoff level on the number of shares affected and the reality of the portfolio earning dividend yields of such high values. At levels below this value, we found that a relatively large proportion of shares were affected.

4. Results

The above analysis was conducted for six different portfolio sizes. For reasons of brevity we shall report the results for all shares (i.e., no liquidity cap) and, where relevant, highlight the differences in the results for the other five portfolio sizes.

The pattern of average portfolio returns for the high and low P/E portfolios drawn from the complete and currently listed groups of shares is reported in Figure 2. These returns are the arithmetic average of portfolio returns relative to the ALSI return in the relevant month post portfolio formation.

INSERT FIGURE 2 ABOUT HERE

Three important observations can be drawn from the data presented in this figure. Firstly, the outperformance of the low P/E shares relative to the high P/E shares holds for both groups of portfolios. Secondly, the pattern of the returns of the portfolios drawn from currently listed shares systematically differs from those drawn from the complete list of shares. Finally, the returns of the portfolios drawn from the currently listed shares only seem to systematically exceed the returns of the portfolios drawn from the complete set of shares.

The first observation suggests that mean reversion is present for portfolios drawn from both groups of shares. The second and third observations indicate that survivorship bias is present in this study. These conclusions are examined in more detail below.

INSERT TABLE 1 ABOUT HERE

The geometric excess mean returns (annualised) for each group of portfolios are summarised in Table 1. This represents the average annual portfolio return in excess of the equivalent ALSI return for the same period. The same geometric means for all five liquidity caps are presented in Figure 3.

INSERT FIGURE 3 ABOUT HERE

Looking at the data in Table 1, we can see that the mean excess returns for high P/E portfolios are significantly lower than the equivalent returns for the low P/E portfolios for both groups¹⁰. While not reported, here similar results holds for the differences in mean returns for the portfolios for all five liquidity caps considered. This strongly supports the conclusion that mean reversion is present in shares which are, or have been, listed on the JSE Securities Exchange.

Anything above this number would lead to an unrealistically large upward bias in the measured return for the portfolio containing such a share.

¹⁰ It can be argued that the data for the current portfolios provides only partial support for the presence of mean reversion conclusion. While the excess returns of the low P/E portfolios are consistently higher than the equivalent high P/E portfolios, the fact that the geometric mean return of the high P/E portfolios is positive indicates that they still outperform, rather than revert to the mean.

The dampening effect of liquidity caps on the size of the mean reversion effect reported in Bailey and Gilbert (2007) is also confirmed here. As is illustrated in Figure 3, there is a strong negative relationship between the geometric mean return for all the portfolios and the liquidity cap level, indicating that liquidity constraints do, as expected, limit portfolio managers' attempts to outperform the market.

The mean reversion result can also be clearly seen if the differences in excess returns between low and high P/E portfolio shares are plotted for the two groups of shares (Figure 4). For the complete list of shares, the average monthly excess return for the low P/E portfolio exceeds that of the high P/E portfolio for 55 out of the 60 months considered. For currently listed shares, the same result holds for all 60 months.

INSERT FIGURE 4 ABOUT HERE

The robustness of the mean reversion result is further supported by the lack of any significant statistical difference in the pattern of outperformance of low P/E portfolios relative to high P/E portfolios illustrated in Figure 4. The differences are remarkably consistent – no matter what shares the portfolios are drawn from¹¹.

The existence of survivorship bias can be seen more clearly if we plot the differences between the mean portfolio returns for high P/E portfolios drawn from the two sets of shares, and the same for low P/E portfolios. This is illustrated in Figure 5.

If there are systematic differences in the firms delisted as compared to those currently listed, then there should be a consistent difference in the returns to high (and low) P/E portfolios drawn from the two groups. In particular, we would expect there to be a positive difference between the portfolios drawn from the currently listed shares and those drawn from the complete list.

This is almost exactly what we find illustrated in Figure 5. The difference between the returns for the current portfolios is positive for 50 months out of 60 for the high P/E portfolios and 43 months for the low P/E portfolios.

The geometric averages of the differences in mean returns are summarised in Table 2. This represents the difference in high (and low) P/E portfolios for the current and complete lists of shares.

INSERT FIGURE 5 ABOUT HERE

¹¹ Note that no attempt was made to adjust these results for the risks of the portfolios. It is possible (though we believe unlikely) that Betas could explain this effect. Zarowin (1989, 1990), for example, finds that risk adjustments of this sort do explain the presence of mean reversion in share returns in the United States as identified by De Bondt and Thaler (1985, 1987). However, these results have been questioned (Albert and Henderson, 1995; Gropp 2004). This is an obvious area for further study, however.

The data presented in Table 2 shows that the positive differences in the portfolio returns for high (and low) P/E portfolios are significantly different from zero. While not reported here, these results hold true for all liquidity adjusted portfolios considered (the maximum p value is 0.0008 which is reported in Table 2).

INSERT TABLE 2 ABOUT HERE

The differences in the geometric average excess returns generated by the high and low P/E portfolios drawn from the two lists are reported in Figure 6 for the five different liquidity caps. This shows that the positive difference between the two is not affected by the changes in liquidity.

INSERT FIGURE 6 ABOUT HERE

This data indicates that survivorship bias is clearly present in the returns of the shares listed on the JSE in the period 1980 to 2007.

5. Conclusion

The data and analysis presented in this paper leads to a clear conclusion regarding the presence of survivorship bias, but implications for researchers are nuanced.

On the one hand, the presence, and importance, of survivor bias is clearly demonstrated in the difference of returns between the complete and listed groups of firms. The mean levels of returns from the currently listed portfolios (both high P/E and low P/E) are significantly higher than those generated from the portfolios of the complete data set. This is exactly the sort of bias that would be expected if the characteristics of the delisted firms are systematically different from those that are listed at each point in time. This highlights the need for researchers to take the survivorship bias problem seriously – you will get incorrect answers if you do not.

However, the results of this study do not challenge the conclusions of Cubbins *et al* (2006), and Bailey and Gilbert (2007) that mean reversion of share returns exists on the JSE. In other words, in this case, the right overall conclusion would have been reached even if a non survivorship bias corrected data set had been used. This suggests that mean reversion of returns is a robust phenomenon as it applies to both groups of shares that we examined. Given that there is no reason why mean reversion should not apply to all shares equally, this is a reassuring result.

The conclusions are subject to the following caveats. Firstly, we are not convinced that we have a complete data set going back in time. As illustrated in Figure 1, there are clear jumps in the number of firms covered by the data provider. This strongly suggests that the data source we used is still not complete and thus free from survivorship bias. Secondly, in the process of creating the portfolios for this analysis, we noticed anomalies in the data which raises questions regarding the

quality of the entire data set. This suggests the need for a data cleaning exercise on this data set. Thirdly, when calculating portfolio returns, we have treated delisted shares as delivering a zero return in the month following their delisting. This is correct if the company is in the process of being liquidated. However, this is obviously not correct if companies are delisted for other reasons and investors receive the value of their shares¹². Finally, we have not made any risk adjustments when comparing portfolio returns. There is some literature (Zarowin, 1989, 1990; Clare and Thomas, 1995) that suggests that this may well affect the conclusions regarding the presence of mean reversion of returns. This offers an interesting opportunity to extend this work.

In conclusion, our analysis shows that any research that excludes delisted shares is likely to be subject to survivorship bias. This may not materially affect the outcomes of the studies (as in this case), but our work suggests that including data for delisted shares is likely to have a significant effect on the results reached. Researchers should be aware of this and attempt to include such data in any empirical analysis of this sort.

¹² This is equivalent to assuming that the return on the delisted share is -100%. This would overstate the loss (and understate the return) if some value is received for the share on its delisting. To test for the sensitivity of our results to this assumption, we re-ran the analysis assuming a -50% return to the shares in the month following their delisting. There was not material change to our results.

Reference

- Albert, RL and GV Henderson, 1995. Firm size, overreaction and return reversals. *Quarterly Journal of Business and Economics*, 34(4): 60-80.
- Bailey, G and E Gilbert. 2007. The impact of liquidity on mean reversion of share returns of the JSE. *Investment Analysts Journal*, 66: 1-11.
- Bain, JS. 1972. Letters to the Editor. *Financial Analysts Journal*, November/December: 104-106.
- Ball R and R Watts. 1979. Some additional evidence on survival biases. *Journal of Finance*, 34: 197-206.
- Banz, RW and WJ Breen. 1986. Sample-dependent results using accounting and market data: some evidence. *The Journal of Finance*, 41(4): 779-793.
- Brown, SJ, W. Goetzman, RG Ibbotson and SA Ross, 1992. Survivorship bias in performance studies. *Review of Financial Studies*, 5(4): 553-580.
- Carpenter, JN and AW Lynch, 1999. Survivorship bias and attrition effects in measures of performance persistence, *Journal of Financial Economics*, 54(3): 337-374.
- Chan LKC, N Jegadeesh and J Lakonishok. 1995. Evaluating the performance of value versus glamour stocks - the impact of selection bias. *Journal of Financial Economics*, (38)3: 269-296.
- Cubbins E, M Eidne, C Firer, and E Gilbert. 2006. Mean reversion on the JSE Securities Exchange. *Investment Analysts Journal*, 63: 1-17.
- Clare, A. and S Thomas. 1995. The overreaction hypothesis and the UK stockmarket. *Journal of Business Finance & Accounting*, 22(7): 961-973.
- Davis, JL. 1994. The cross-section of realized stock returns: the pre-COMPUSTAT evidence. *Journal Of Finance*, 49(5): 1579-1593.
- Davis, JL. 1996. The cross-section of stock returns and survivorship bias: evidence from delisted stocks. *The Quarterly Review of Economics and Finance*, 36(3): 365-375.
- De Bondt WFM and RH Thaler 1985. Does the Stock Market Overreact? *Journal of Finance*, 40(3): 793-806.
- De Bondt WFM and RH Thaler. 1987. Further Evidence On Investor Overreaction and Stock Market Seasonality. *Journal of Finance*, 42(3): 557-582.
- Fama, E and K French, 1992. The cross-section of expected stock returns *Journal of Finance*, 47(2): 427-465.

Grindblatt, M and S. Titman, 1989. Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings, *Journal of Business*, 62(3): 313-416.

Gropp J 2004. Mean reversion of industry stock returns in the U.S., 1926–1998. *Journal of Empirical Finance*, 11(4): 537-551.

Kothari, SP, Shanken, J and RG Sloan, 1995. Another look at the cross-section of expected stock returns, *Journal of Finance*, 50(1): 185-224.

McElreath RB and CD Wiggins. 1984. Using the COMPUSTAT tapes in financial research: problems and solutions. *Financial Analysts Journal*, January-February: 71-76.

Salamon GL and ED Smith. 1977. Additional evidence on the time series properties of reported earnings per share: comment. *Journal of Finance*, 32(5): 1795-1801.

Wagner WH and SC Lau. 1971. The Effects of Diversification on Risk. *Financial Analysts Journal*, November/December: 48-53.

Zarowin P. 1989. Does the Stock Market Overreact to Corporate Earnings Information? *Journal of Finance*, 44(5): 1385-1399.

Zarowin P. 1990. Size, Seasonality, and Stock Market Overreaction. *Journal of Financial and Quantitative Analysis*, 25(1): 113-125.

Table 1: Geometric average excess returns for high and low P/E portfolios (annualised) – no liquidity cap

Portfolio	Geometric Mean Return	p-values (paired t-test for difference in means)
High P/E (Complete)	-1.05%	0.0000
Low P/E (Complete)	7.08%	
High P/E (Current)	2.32%	0.0000
Low P/E (Current)	10.47%	

Table 2: Geometric average difference in returns for high and low P/E portfolios

Portfolios compared (Current – complete)	Difference in geometric means	p-values (one tailed paired t-test for difference from zero)
High P/E portfolios	3.39%	0.0000
Low P/E portfolios	3.13%	0.0008

Figure 1: Number of listed and delisted firms included in the analysis

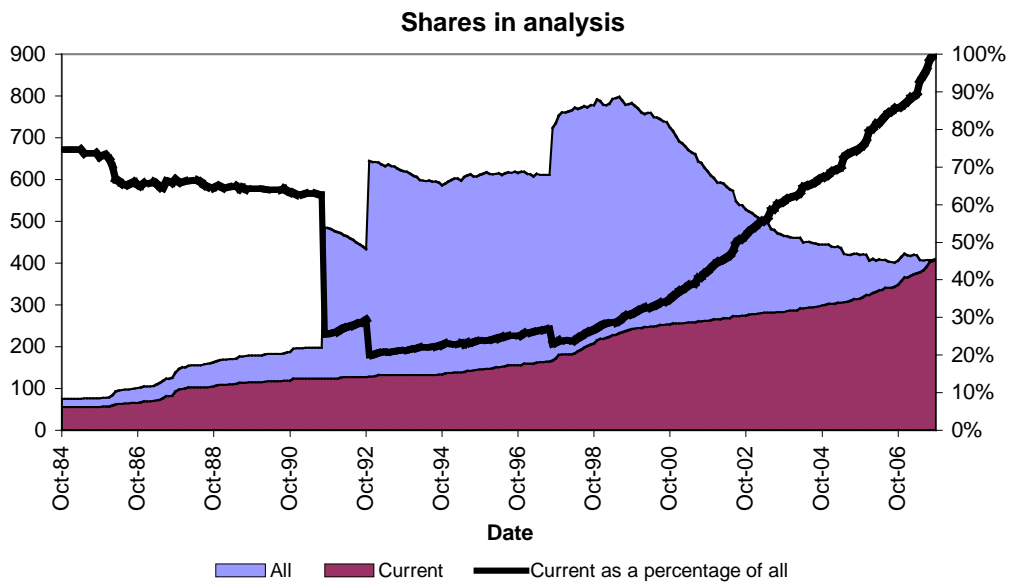


Figure 2: Returns for high and low P/E portfolios (relative to the ALSI) drawn from currently listed and complete populations – no liquidity cap

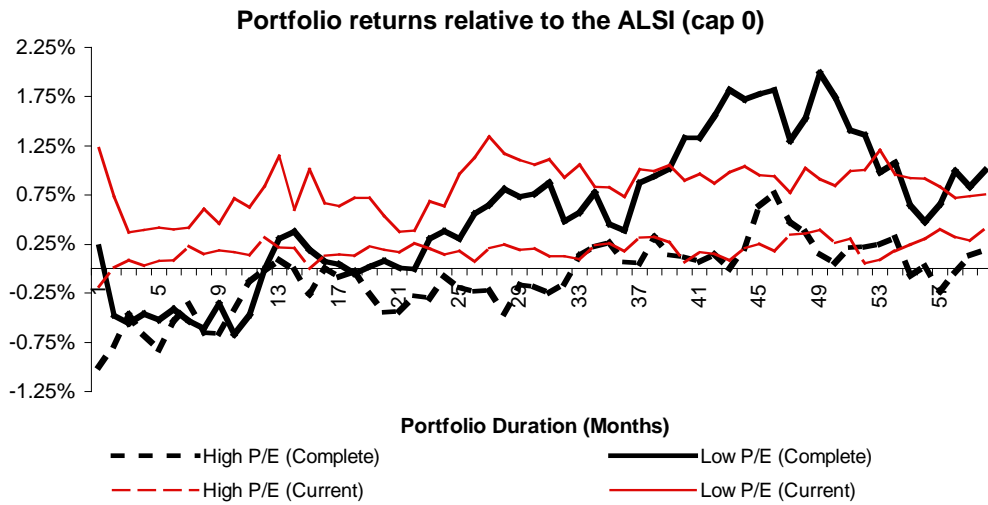


Figure 3: Impact of liquidity caps on the geometric mean excess returns of high and low P/E portfolios

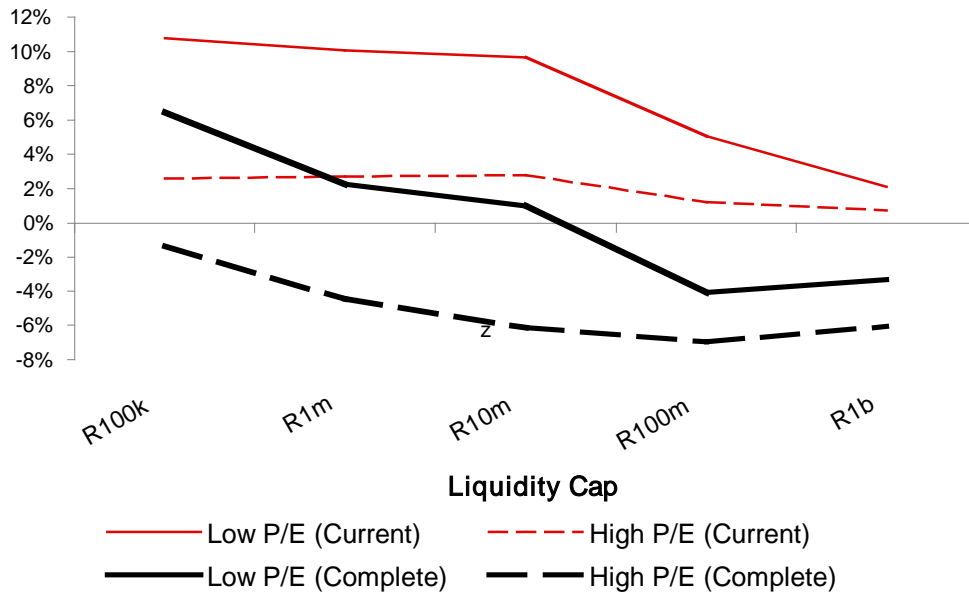


Figure 4: Difference between low P/E and high P/E portfolio excess returns – no liquidity cap

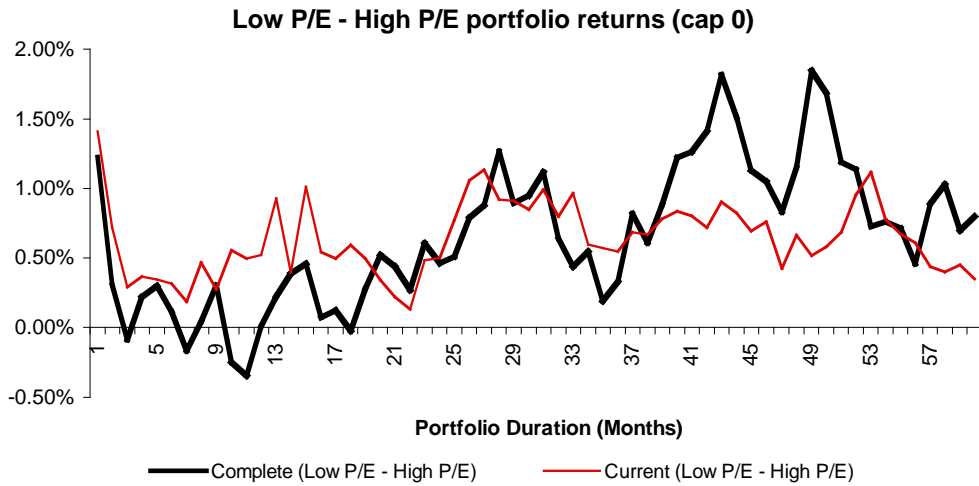


Figure 5: Differences in mean portfolio returns (Current - Complete) for high and low P/E portfolios

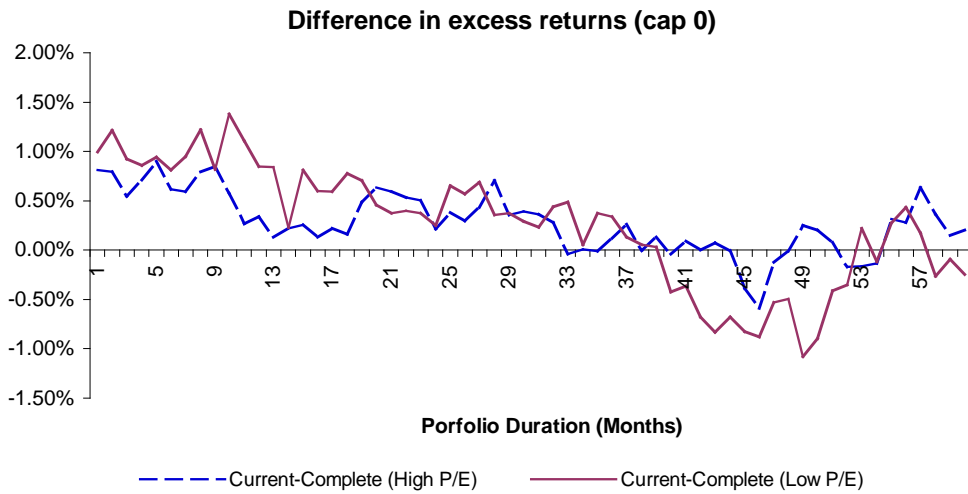


Figure 6: Differences in geometric average portfolio returns between current and complete portfolios

