Calculating and Comparing Security Returns is harder than you think: A Comparison between Logarithmic and Simple Returns

by

Robert S Hudson[1]

Newcastle University Business School

Newcastle University

Andros Gregoriou[2]

Hull University Business School

University of Hull

[1]Newcastle University Business School, 5 Barrack Road, Newcastle NE1 4SE United Kingdom.
Tel: +44 191 208 1642; E-mail address: robert.hudson@ncl.ac.uk

[2] Hull University Business School, University of Hull, Hull HU6 7RX United Kingdom
Tel: +44 (0)1482 463066; E-mail address: A.Gregoriou@hull.ac.uk

Calculating and Comparing Security Returns is harder than you think: A Comparison between Logarithmic and Simple Returns

Abstract

We show that there are some troubling differences between mean returns calculated using logarithmic returns and those calculated using simple returns. The mean of a set of returns calculated using logarithmic returns is less than the mean calculated using simple returns by an amount related to the variance of the set of returns. This implies that there is not a one-to-one relationship between mean logarithmic and mean simple returns. In addition, when logarithmic returns are used, ceteris paribus, higher variance will automatically reduce expected returns. We illustrate when these effects are important drawing on examples from the extant literature.

1. Introduction

Calculating the return on a security in a particular period as the difference between the natural logarithm of the security price at the end of the period and the natural logarithm of the security price at the beginning of the period (referred to as a logarithmic return) is a very commonly used procedure in finance even though this returns differs from the monetary growth which would be actually be achieved by an investment over that period. The logarithmic method of calculating returns is frequently preferred to the obvious alternative of using the return calculated on the basis of simple interest over the period in question which, of course, is the monetary return which would actually be achieved by an investment over that period[1].

There are a number of theoretical and practical advantages of using logarithmic returns and these are outlined in section 1. In many areas of academic finance the advantages of using logarithmic returns appear to have been tacitly accepted although a few papers have pointed out pitfalls in their use in particular fields of investigation. In the area of event studies Dissanaike and Le Fur (2003) point out problems with the use of cross-sectional averages of logarithmic returns. Kothari and Warner (1997) and Barber and Lyon (1997) show that logarithmic returns are negatively skewed such that test statistics are unlikely to be well specified. In the area of assessing investment returns over long periods of time there has been a debate over whether logarithmic or simple means are most appropriate to assess returns (Jacquier et a1, 2003)[2].

This paper contrasts the use of logarithmic and simple returns in calculating and comparing mean single period security returns. It shows that there are some troubling differences between mean returns calculated using the two methods. The mean of a set of returns calculated using logarithmic returns is less than the mean calculated using simple returns by an amount related the variance of the set of returns where the variance is relatively invariant whether it is measured using logarithmic or simple returns. This implies that there is not a one-to-one relationship between mean logarithmic and mean simple returns so it is not possible to extrapolate conclusions about terminal wealth from studies carried out using logarithmic returns[3]. In particular, if period 1 has a higher mean logarithmic return than period 2 this does not necessarily imply that the mean simple return in period 1 is higher than in period 2. Thus even the most basic qualitative conclusions derived from investigations using logarithmic returns may not hold for the

[1] In this paper we adopt the following notation:
$R_{Lt} = \ln(P_{t+1}) - \ln(P_t)$ where R_{Lt} is the log return for period t, P_{t+1} is the price of a security at time t+1 and P_t is the price of a security at time t.
$R_{St} = P_{t+1}/P_t - 1$ where R_{St} is the simple return for period t, P_{t+1} is the price of a security at time t+1 and P_t is the price of a security at time t.

[2] Jacquier et al use different terminology referring to geometric and arithmetic means as opposed to logarithmic and simple means.

[3] In mathematical terms there is a non-injective relationship between mean logarithmic and mean simple returns.

monetary returns of actual investments. In addition, given that the mean logarithmic return in a period depends on the variance of returns in that period, the risk and return in that period are not independent by construction which is troubling in the context of much finance theory.

This relationship between variance and return does, however, enable the derivation of an approximate method for converting between means calculated using logarithmic returns and those calculated using simple returns. This enables meaningful comparisons to made between past empirical studies made using alternative returns measures.

We undertake empirically studies to confirm the theoretical findings discussed above. We illustrate that the direct relationship between risk and return depends on how returns are measured by using a GARCH-M model. We then assess the relationship between risk and return for logarithmic and simple returns in periods of differing variance.

The paper further illustrates, by means of practical examples using cases drawn from the extant finance literature, that there can be a substantial empirical difference between results derived using logarithmic returns and those derived using simple returns and confirms the circumstances in which these differences are likely to be most important. Much of the literature in finance is, of course, related to calculations and comparisons of security returns so it is not possible to give a comprehensive range of examples. We do, however, cover a range of areas. First, we draw an example from the research into calendar based anomalies where a time series is divided into subsections based on a particular calendar effect such as the day of the week, month of the year, the day before a public holiday etc. Second, an example is taken from the research into trading rules in which a time series of security returns is divided into subsections that are expected to exhibit different returns using a particular trading rule. Third, an example is drawn from the literature on event studies where a time series is divided into subsections based on when specified events take place. Examples of such events include stock splits, IPO's, results declarations and other corporate events and other market events such as large drops in stock prices. Over these areas, there seems little consensus or indeed discussion in the literature regarding the best method of calculating returns and indeed many papers do not clearly specify which type of return is used. Even a moderate level of investigation, however, gives the conclusion that each of the literatures mentioned freely uses both logarithmic and simple returns. We provide evidence to support this assertion in section 4.

Section 5 considers how conclusions from finance research can depend crucially on the return measure used. We look at the conclusion from published research studies using daily data and also carry out an investigation showing how different return measures can have a very substantial effect on results when intraday data is used.

2. Rationale for the Use of Logarithmic Returns

A number of powerful arguments are put forward to justify the use of logarithmic returns:

i) Logarithmic returns can be interpreted as continuously compounded returns. This means that, for non-stochastic processes, such as the returns on risk-free fixed interest securities held to maturity, when logarithmic returns are used, the frequency of compounding does not matter and returns across assets can more easily be compared.

ii) Using continuously compounded (logarithmic) returns is advantageous when considering multi-period returns as the continuously compounded multi-period return is simply the sum of continuously compounded single period returns. Continuously compounded returns are time additive and it is easier to derive the time series properties of additive processes than multiplicative processes (see Campbell et al, 1997, p11). In this context some studies have shown that using simple returns to estimate returns over longer periods can be quite unsatisfactory (see, Roll, 1983 and Dissanaike, 1993).

iii) The use of logarithmic returns prevents security prices from becoming negative in models of security returns (see Jorion, 2001, p100).

iv) If a security price follows geometric Brownian motion[4] (a very popular model of security price movements used, for example, in the Black-Scholes option pricing model) then the logarithmic returns of the security are normally distributed.

v) For forecasting future cumulative returns, continuous compounding of the expected logarithimic return will give a better guide to median future cumulative returns (the return that investors are likely to realise) than compounding expected simple returns (Hughson et. al., 2006).

vi) Logarithmic returns are approximately equal to simple returns. Inspection of the formula connecting logarithmic and simple returns $R_{Lt} = \ln(1+ R_{St})$ shows that as long as R_{St} is not too large (Rozeff and Kinney, 1976, p380, suggest $R_{St} \leq 0.15$) then logarithmic and simple returns are very similar in size. Whilst this is true, it is important not to wrongly deduce from this that the mean of a set of returns measured using logarithmic returns is necessarily very similar to the mean of the same set of returns measured using simple returns. The theory behind this result is outlined in the next section and Appendix A

3. Undesirable Properties of Logarithmic Returns

There are some undesirable properties associated with logarithmic returns:

i) Logarithmic returns do not give a direct measure of the change in wealth of an investor over a particular period. By definition, the appropriate measure to use for this purpose is the simple return over that period. For non-stochastic systems this is a trivial difficulty as there is a one-to-one correspondence

[4] Also known as the multiplicative random walk see Cootner (1964) and Fama (1965).

between logarithmic returns and simple returns P[5]. The situation is much more problematic for stochastic systems as discussed in iii) below.

ii) The difference between the mean logarithmic return of a security over a given time period and the mean simple return over the same period depends on the variance of the returns as well as their expected mean simple return. A formal proof of this property is shown in Appendix A. In addition, the variance of the returns as almost independent of whether logarithmic or simple returns are used and a proof of this is also given in Appendix A. The relationship between variance and return outlined in this paragraph does however, enable the derivation of an approximate method for converting between means calculated using logarithmic returns and those calculated using simple returns and this derivation is outlined in Appendix A. The formula for this approximate method is:

$$\bar{x}_L = \bar{x}_S - 0.5\sigma_S^2 \qquad (1)$$

Where: mean sample return (logarithmic) is $\bar{x}_L$

mean sample return (simple) is $\bar{x}_S$

sample variance (simple) is σ_S^2

iii) Given that mean logarithmic returns are related to both the mean and variance of simple returns there cannot be a one to one relationship between mean logarithmic returns and mean simple returns for stochastic systems. A particular mean logarithmic return may result from multiple combinations of means and variances of simple returns. Appendix B illustrates this property. An important implication of this is that the mean of the logarithmic returns of a particular distribution cannot be used to infer the mean of the simple returns of that distribution.

The properties of logarithmic returns presented in i) to iii) above give rise to a conceptual complication in that finance theory views risk[6] and return as concepts that are linked only due to the fact that in a competitive market investors who take a particular level of risk are rewarded by an appropriate level of return. This view presents no difficulty when simple returns are considered as there is no direct and automatic mathematical connection between risk and return. However, the expected logarithmic return of a series of returns does depend on the variance of the series. Thus, to some extent, this mathematical artefact of the method used to calculate logarithmic returns obscures the relationship between risk and return. Appendix C shows an example where, in a particular period, the expected simple return is higher than the expected logarithmic return due to high spread of returns.

[5] $R_{Lt} = \ln(1+ R_{St})$, $R_{St} = \exp(R_{Lt}) - 1$

[6] In this paper, as is conventional in finance, we assume that risk is measured by variance.

Inspection of formula 1 shows when the differences between logarithmic and simple returns will be most acute. In numerical terms, the difference increases as σ_S^2 increases. In proportional terms, the ratio of $\bar{x}_S$ to $\bar{x}_L$ depends on the relative size of $\bar{x}_S$, $\bar{x}_L$ and σ_S^2 . It is also noteworthy that the size of conventional t tests also depend on the relative size of $\bar{x}_S$, $\bar{x}_L$ and σ_S^2 . When testing whether a mean differs from zero, simple returns will normally produce a higher significance level than log returns if the mean is positive and a lower significance level if the mean is negative although these rules may not always hold if the variance of the sample is particularly large relative to the mean. Tests of the equality of mean returns in periods of differing variance will also differ according to whether logarithmic or simple returns are used. Appendix D formally analyses the implications of different return measures for significance tests.

In Sections 3 and 4 we confirm that the theory discussed above is confirmed by empirical observations and that the formula derived above appears to be highly accurate in practice.

4. Empirical Evidence on the Relationship between Risk and Return for Simple and Logarithmic returns

4.1 The direct measurement of the Relationship between Risk and Return

Given that the relationship between mean logarithmic returns and mean simple returns depends on the variance of the simple returns this implies that one should expect a different relationship between risk and return depending on how returns are measured. To investigate this we have made maximum likelihood estimates of the parameters of the MA(1) GARCH-in-mean model introduced in Bali and Ping (2006) to directly measure the relationship between risk and variance:

$$R_r = \alpha + \beta\sigma_t^2 + \varepsilon_t + \theta\varepsilon_{t-1}$$
$$\sigma_t^2 = \delta_0 + \delta_1\varepsilon_{t-1}^2 + \delta_2\sigma_{t-1}^2$$

The estimates have been calculated using logarithmic and simple returns and are based on data from the full life of the S&P 500 index from 1/4/1950 to 12/21/2009. The results are shown in Panel A of Table 1 with the main parameter of interest being β. For both logarithmic and simple returns β is a positive and highly significant coefficient indicating a positive short term relationship between return and risk. When the results for simple returns are compared to those for logarithmic returns we see that the coefficient of β is substantially higher and more significant. Given we are using the same data set throughout this does not reflect any difference in the underlying relationship between risk and return but is an artefact of the way returns are being calculated.. From the approximate relationship between mean sample logarithmic and simple returns derived in Appendix A (formula 1), an increase in variance should reduce logarithmic returns compared to simple returns indicating that the coefficient of β should be lower for the equation calculated using logarithmic returns and this is indeed the case.

It can be shown that the existence of a relationship between risk and return in some sample periods depends on the way returns are measured. Panel B of Table 1 looks at the S&P 500 index from 1/2/1980 to 12/21/2009. If logarithmic returns are used β is not significantly different from 0 indicating no relationship between risk and return whereas, if simple returns are used, there is still a significant positive relationship.

4.2 The measurement of the Relationship between Risk and Return in periods of differing variance

Table 2 shows the empirical relationship between logarithmic and simple returns over subsets of data with different variance drawn from the entire history of the Dow Jones index from 2 January 1897 to 23 March 2009. The subsets with different variances are determined by reference to the absolute value of the simple return on each day in the sample[7]. The variance of the returns in each subset is closely related to the size of the absolute value of the returns.

The relationships previously derived theoretically are closely confirmed empirically. Over the entire sample period of 30,643 days, and in each of the subsets with different variance, the mean logarithmic return is less than the mean simple return. The ratio of the mean simple return to the mean logarithmic return is often considerable, exceeding 1.3 where the means are positive and being of the order of 0.7 or less when the means are negative. In contrast, as predicted, the difference in sample variances between the two return measures in each subset is very small usually less than 1%. The approximate method of converting from mean simple returns to mean logarithmic returns (Formula 1) also proves to be highly accurate with an error that is always less than 1%. This later finding gives very strong and direct empirical evidence that the difference between mean logarithmic returns and mean simple returns for actual stock data does depend on the variance of the simple returns.

5. Empirical Evidence on the Theoretical relationship in a range of situations

In this section empirical evidence is presented from several areas of finance where security returns are routinely calculated and compared.

Generally the primary interests of researchers are in calculating returns and determining how they differ between particular subsets of a population. However, variance might well also differ between the subsets and, as we have seen, this may easily effect the conclusions with differential effects on logarithmic and simple returns. In this section we outline three examples to illustrate this effect, one

[7] The results reported in this subsection are robust to the choice of a wide range of subsets which are not all reported to conserve space.

where the subsets have similar variances, one where there are moderate differences in variance and one where there are large differences in variance. To show wide applicability we look at examples from three different areas.

5.1 Calendar Based Anomalies (an example in which population subsets have similar variance)

There is a very substantial academic literature on calendar based anomalies where the investigations generally focus on whether returns in a particular calendar period differ from those not in that period. The literature in this area is far too large to summarise in anything less than a dedicated survey paper. However, important effects and the early, influential, papers that popularised them, include general seasonality, especially monthly effects, (Rozeff and Kinney, 1976), weekend and holiday effects (French, 1980), Monday/day-of-the week effects (Gibbons and Hess. 1981) and the turn-of-the month effect (Lakonishok and Smidt, 1988). Sullivan et al (2001) and Hansen et al (2005) provide relatively recent summaries of the large literature in this area and offer contrasting views on the extent to which the findings are driven by data mining. Studies in this area freely use both logarithmic and simple returns. For example, if the papers already mentioned are considered, Rozeff and Kinney (1976), French (1980) and Hansen et al (2005) explicitly state that they use logarithmic returns, Sullivan et al (2001) explicitly state that they use simple/arithmetic returns and the other papers are not very explicit on which returns they use although, from the context, simple returns seem the more likely.

In the current example we focus on the turn-of-the-month effect. This is partly because this effect has been empirically strong in most world markets for many years and is still the subject of active research (see McConnell and Xu, 2008 and Hudson and Atanasova, 2009) but mainly because, for this effect, the subsets under comparison have similar variances. It should be noted in passing that not all calendar effects give rise to subsets with similar variances, for example, when the holiday effect is considered, pre-holiday days and non pre-holidays have substantially different variances (Chong et al, 2005).

The results of investigating the existence of the Dow Jones Index from 2 January 1897 to 23 March 2009 are reported in table 3. The results have been set out in a similar format to that used in the recent paper on the effect by McConnell and Xu (2008). In broad terms, the hypothesis is that the four days around the turn of each calendar month comprising the last day of the preceding month and the first three days of the subsequent month should experience higher average returns than all other trading days. In accordance with the recent literature, the table shows that this is indeed the case with returns around the turn of the month being higher than other days with very high levels of significance whether returns are measured using logarithmic or simple returns. When logarithmic and simple returns are compared, as expected, the mean logarithmic returns are lower for each subset of days. Again, as expected, when each subset of days is examined individually the variance of logarithmic returns is very similar to the variance of simple returns. Formula (1) the approximate formula for converting from mean simple to mean logarithmic returns again proves highly accurate. An important point to note is that when the different subsets of days are compared they all have similar variance i.e. Days -1 to +3 and the other days all have similar variance. This implies that the difference between mean simple and mean logarithmic returns will not be distorted by differing variances and this can be seen in the last line of the table where the differences are quite constant. An implication of this, based on the theory in Appendix D, is that the t-statistic testing the difference between the means of days around the turn of the month and other days should be little affected by whether logarithmic or simple returns are used and this is indeed the case.

5.2 Trading Rules (An example in which population subsets have moderately different variances)

The use of technical trading rules in stock markets is an established practice amongst many practitioners and has also generated a substantial academic literature. The rules seek to identify subsets of a data set where returns are expected to be higher. A huge number of rules are used in practice and an overview of the area is provided in Lo et al, 2000.

An important paper in this area is by Brock et al (1992) who find that simple moving average and trading range break-out rules have predictive ability on the Dow Jones Index from 1897 to 1986. These results have generated quite a number of subsequent investigations. Several studies such as Hudson et al (1996) and Ratner and Leal (1999) have confirmed that the trading rules are predictive in other equity markets and for individual stocks (Bokhari et al, 2005). Sullivan et al (1999) find that the results are robust to data-snooping. Studies in this area have used both logarithmic and simple returns. For example, if the papers already mentioned are considered, Brock et al (1992) and Hudson et al (1996) use logarithmic returns and the other papers use simple returns (adjusted for inflation in the case of Ratner and Leal (1999)).

For the example on trading rules we focus on the rules used by Brock et al (1992). This is partly because these rules are very well known and have been extensively investigated but also because they produce data subsets for comparison where there are moderate differences in variances.

We investigate the rules on an updated version of the data set used by Brock et al (1992), the Dow Jones Index from 1897 to 2009. Two technical trading rules, moving average and trading range breakout, are used in the investigations. For moving average rules, buy (sell) signals occur when the short run moving average over period x, is above (below) the long run moving average over period y by an amount larger than a band z. The buy (sell) return on each individual day in the sample is calculated according to these signals. For trading range breakout rules a buy (sell) signal is triggered if the stock price, averaged over period x, moves above (below) a 'resistance' (support) level defined as the maximum (minimum) price achieved by the stock over a previous period, y, by an amount larger than a band z. A 10-day holding period return is calculated following each signal. For both types of rule returns are calculated using logarithmic and simple returns.

The results of the investigations are shown in table 4 and table 5. The moving average rules are covered in table 4. As found in previous studies in this area the rules do have predictive ability with the mean return on a day when there is a buy signal being significantly greater than the average daily return over the whole sample and the mean return on a day when there is a sell signal being significantly smaller than average and in fact actually negative. When logarithmic and simple returns are compared, as expected, the mean logarithmic returns are lower for each subset of days. The significance of the rules in this sample period is such that even with the lower mean return associated with logarithmic returns the rules are still highly significant. When each subset of days is examined individually the variance of logarithmic returns is very similar to the variance of simple returns. Formula (1), the approximate formula for converting from mean simple to mean logarithmic returns, again proves highly accurate. In the case of this example it is interesting to note that when the different subsets of days are compared the subset of days associated with sell signals has a substantially larger variance than the subset associated with buy signals. This implies theoretically that the difference between mean simple and mean logarithmic returns will be differentially affected by differing variances for the buy and sell subsets. This is confirmed empirically with the difference between simple and logarithmic returns being much larger for the subset associated with sell signals. Additionally the theory developed in Appendix D would indicate that the t-statistics testing the significance of the difference between buy and sell periods will differ depending on whether logarithmic or simple returns are used., This is again confirmed empirically with logarithmic returns giving higher t-statistics in this case.

The trading range breakout rules are covered in table 5. The rules do have some predictive ability. The mean return on a day when there is a buy signal is significantly greater than the average daily return over the whole sample when both logarithmic and simple returns are considered. By contrast, the mean return on a day when there is a sell signal is not significant for either logarithmic or simple returns. When the difference between buy and sell returns is considered there is some significance with this being substantially stronger when the returns are measured using logarithmic returns. For the TRB(1,50,0) there is significance at the 10% level with logarithmic returns but no significance with simple returns. For the TRB(1,50,0.01) there is significance at the 5% level with logarithmic returns but only at the 10% level with simple returns. When each subset of days is examined individually the variance of logarithmic returns is very similar to the variance of simple returns. As for the moving average rules, when the different subsets of days are compared the subset of days associated with sell signals has a substantially larger variance than the subset associated with buy signals. This implies theoretically that the difference between mean simple and mean logarithmic returns will be differentially affected by differing variances for the buy and sell subsets. This is confirmed empirically with the difference between simple and logarithmic returns being much larger for the subset associated with sell signals. As predicted by the theory in Appendix D, the t statistics are for the difference between the buy and sell subsets are also substantially affected by the use of different return measures.

5.3 Event Studies (An example in which population subsets have substantially different variances)

Event studies cover a wide range of research studies where a time series is divided into subsections based on when specified events take place. Such events include stock splits, IPO's, corporate events and other market events such as large drops in stock prices. As discussed in Dissanaike and Le Fur (2003) both logarithmic and simple returns have been used extensively in this area. In this sub-section we focus on the work on large drops in stock prices which has particularly interesting properties in the context of this investigation as outlined below.

Many papers have investigated the short term reaction of individual stocks to large preceding price movements (for example, Brown et al. (1988, 1993), Atkins and Dyl (1990), Bremer and Sweeney (1991), Cox and Peterson (1994), Park (1995), Pritamani and Singal (2001), Mazouz et al (2009)). Similarly a substantial number of papers have investigated how stock indices react to large preceding price movements (see, Ferri and Min (1996), Schnusenberg and Madura (2001), Hudson et al. (2001), Lasfer et al. (2003), Michayluk and Neuhauser (2006) and Spyrou et al (2006)). The research in this area generally finds strong elements of short term predictability amongst securities following large one-day price changes although the rationale for this predictability is unresolved with market microstructure effects, rational responses to changes in risk and irrational behavior by investors all being suggested as explanations for the phenomenon. Studies in this area have used both logarithmic and simple returns although simple returns probably predominate. For example, if the papers already mentioned are considered, Mazouz et al (2009) explicitly state they use logarithmic returns whereas Bremer and Sweeney (1991), Schnusenberg and Madura (2001) and Spyrou et al (2006) explicitly state they use simple returns, perhaps strangely, the other papers do not explicitly state which type of returns they use.

In the context of this paper this research is interesting because it focuses on the very time periods, around large price movements, when variance will be highest and there will be the greatest potential differences between logarithmic and simple returns and their significance levels. To investigate this we again focus on the Dow Jones Index from 1897 to 2009. In accordance with the general approach in the literature we consider the index returns after daily market changes of various sizes. The results are shown in Table 6. Panel A shows returns after large positive price changes[8]. It is clear that the variance on the day after a price change is positively related to the size of the prior change. As theoretically expected there is little difference between observed variance whether measured using logarithmic or simple returns. By contrast, there is a substantial difference between mean logarithmic and mean simple returns. The proportionate difference ranges up to over 35% for the sub-sample of days where the previous simple return was greater than 2.5% which is clearly of potential economic significance. As expected, statistical significance can also be affected with the significance of expected returns calculated using simple returns always being greater than when calculated using logarithmic returns. For example in the sub-sample of days where the previous simple return was greater than 2% the simple mean is significant at the conventional 5% level whereas the logarithmic mean is only significant at the 10% level. Formula (1), the approximate method for converting from mean simple to mean logarithmic returns, proves highly accurate for the sub-periods reported in this panel although the results are not reported to conserve space.

Panel B shows returns after large negative price changes. As seen for positive price changes, the variance on the day after a price change is positively related to the absolute size of the prior change and there is little difference between observed variance whether measured using logarithmic or simple returns. There is again a substantial difference between mean logarithmic and mean simple returns. Measures of statistical significance are also affected. For example in the sub-samples of days where the previous simple return was less than -2.5% and less than -2% the simple means is significant at the 10% level whereas the logarithmic mean is not significant. Formula (1), the approximate method for converting from mean simple to mean logarithmic returns, again proves highly accurate for the sub-periods reported in panel B although the results are not reported to conserve space.

6. Implications of our Findings for Financial Investigations

It might be thought that although the forgoing sections have presented some theoretical issues and illustrated them with practical examples there is no material issue as logarithmic returns and simple returns would give very similar results for financial investigations. This, however, is not generally the case. Regardless of the observation interval being used can make a difference to the results at the margin. When short observation intervals are being used different return measures are likely to lead to very substantial differences in results

In this section we initially review a sample of prior studies using daily data and show that the method of return calculation is too important to be ignored. Secondly we present a study using intraday data which shows that the method of return calculation can have a dramatic effect on results.

[8] The results reported in this subsection are robust to a wide variety of different definitions of large price movements. The full set of results has not been presented to conserve space.

6.1 Prior Studies using Daily Data

It is clear from the previous discussions that, at the margins, the choice of return measure will certainly make a difference. Many hundreds of papers in the finance literature adopt at least some of the methodology covered above and these papers certainly report thousands of mean returns and the associated t statistics. Clearly a very substantial number of these results have significance which is marginal enough to be potentially affected by the choice of return. Doing any sort of systematic investigation of this is an impossibly large task but to give a preliminary indication of the size of the issue we review a few papers in the calendar effect area. There is no intention to criticise the papers covered. They have just been chosen as clear and convenient examples from amongst hundreds of possibilities in the finance literature. We apply formula (1) to the results published in the paper concerned to try to determine how many of the results might have a changed level of significance had a different return measure been used[9]. Arsad and Coutts (1997) explicitly use logarithmic returns and report a table showing returns on different days of the week over many subsets of the UK FT 30 index from 1935 to 1994. In their table 1 they report 65 separate returns of which 29 are significant at least at the 10% level (6 are significant at the 10% level, 10 at the 5% level and 13 at the 1% level). Applying formula (1) would indicate that, if simple interest had been used, several of the reported levels of significance would change, one previously insignificant return would become significant at the 10% level, two returns would increase in significance from the 10% level to the 5% and 1% levels respectively and one return would reduce in significance from the 5% to the 10% level. Wang et al (1997) test the Monday effect on the NYSE-AMEX and Nasdaq indices on various data subsets between 1962 and 1993. Their table 1 reports 63 returns of which 37 are significant at least at the 10% level (5 are significant at the 10% level and the others at the 5% level). The authors do not make it entirely clear whether they are using logarithmic returns but, if this were the case, applying formula (1) would change the significance of one of the results from the 5% to the 10% level. Chong et al (2005) examine the pre-holiday effect in three international stock markets. Although it is not explicitly stated in the paper simple returns were used in this investigation. Tables 4, 5 and 6 in their paper deal with data from the Hong Kong, US and UK markets respectively. In total the significance of the difference between pre-holiday returns and non-pre holiday returns is tested on 15 different subsets of the data. The difference is significant for 6 of these subsets (2 are significant at the 10% level, 2 at the 5% level and 2 at the 1% level). If the exercise had been carried out using logarithmic returns, applying formula (1) indicates that one result would increase in significance from the 10% level to the 5% level and one result which was significant at the 10% level would become insignificant.

[9] Details of the calculations are available from the authors.

These specimen examples show how the results of many studies would be affected if returns were calculated in a different way. Basically results with fairly marginal statistical significance are likely to lose or gain significance depending on how returns are measured and this could affect the whole set of qualitative conclusion to be drawn from a study.

6.2 Effects of Intraday Data

The examples covered so far use daily data but the effect can be much more dramatic in studies dealing with market microstructure. As outlined in section 1n the ratio of $\bar{x}_S$ to $\bar{x}_L$ depends on the relative size of $\bar{x}_S$, $\bar{x}_L$ and σ_S^2 . Now it is well known empirically that as stock return horizons become shorter the standard deviation of returns becomes larger relative to their expected return (see, for example, Kritzman, 1994). In fact, if a process follows a log-normal distribution, then the sum of the outcomes of the process over a number of time periods increase directly in line with the number of periods whereas the standard deviation of the sum increases proportionately to the root of the number of periods. This leads us to expect that, as the investment holding period decreases, the size of the standard deviation of returns becomes larger relative to expected returns. Thus the difference between simple and logarithmic returns may be particularly large in microstructure work.

Many papers in the microstructure literature, as in the rest of finance, are primarily concerned with the calculation of expected returns. For instance, many papers deal with the price impact of block trades, that is, the amount that the price of a stock moves around the purchase or sale of a large block of equity in that stock. Papers in this branch of the literature routinely use both simple and logarithmic returns. For example, Kraus and Stoll (1972), Dann et al (1977) and Frino et al (2005) use simple returns and Holthausen et al (1987), Gemmill (1996), Madhaven and Cheng (1997) and Gregoriou (2008) use logarithmic returns. The essence of work in this area is in measuring the size and sign of returns but at the microstucture level the choice of return can be of considerable significance. To illustrate this we have considered a relevant example decomposing the bid-ask spread into its information and non-information components. The details of the example are given in Appendix E. As expected the effect of using different return measures is dramatic in the intraday domain. In this example, the main result of interest, the mean percentage error of the model, varies by almost 300% depending on the return measure used.

7. Conclusions

This paper demonstrates that comparing means calculated using logarithmic and simple returns in financial calculations raises some troubling issues. Theoretical proofs have established that the mean of a set of returns calculated using logarithmic returns is less than the mean calculated using simple returns by an amount related the variance of the set of returns, where the variance is relatively invariant whether it is measured using logarithmic or simple returns. This implies that there is not a one-to-one relationship between mean logarithmic and mean simple returns so it is difficult to draw conclusions about expected terminal wealth from studies carried out using logarithmic returns. In addition, calculations of the relationship between risk and return calculated using logarithmic returns will systematically differ from those calculated using simple returns. Indeed when logarithmic returns are used, ceteris paribus, higher variance will automatically reduce expected returns as a matter of basic algebra. Thus the relationship between risk and return in any financial situation depends on how returns are measured.

Empirical examples draw from several areas of academic practice confirm our theoretical findings. They also show that the effect can be material in practical studies with substantial differences in the magnitude of calculated returns being easily discernable and also instances where the statistical significance of results change. In general, for any empirical study where statistical significance is fairly marginal altering the return measure is likely to alter significance. The effect of altering the return measure can be quite dramatic where intraday data is involved.

Clearly the return under consideration in any research exercise could be defined as either the logarithmic return or the simple return and each of these would give an internally consistent logical framework to address the problem. Nonetheless, in the context of investigations into the terminal wealth of investors, it seems clear that simple interest is the most appropriate measure to use.

In the light of the foregoing it may be appropriate in research studies of returns to give greater consideration to whether mean returns are calculated simple or logarithmic returns. Calculating returns using either, or both, methods is generally a trivial task for future research given modern computer systems. There are, however, numerous papers in the literature where only one type of return has been reported and in some cases it might be desirable and convenient to revisit the conclusions without getting involved in reopening old calculations. Accordingly, we give an approximate method for adjusting means calculated using logarithmic returns so meaningful conclusions about terminal wealth can be drawn from studies using these returns.

In summary, although each method of calculating returns has advantages, the methods may give results that are surprisingly different. It is worthwhile to be aware of this and so not to draw inappropriate conclusions from empirical studies.

Table 1

Daily Risk-Return Relationships Based on Daily GARCH-in-Mean Estimates from the S&P 500

The figures in this table are the maximum likelihood estimates of the parameter (β) based on a MA(1) GARCH-in-mean model

$$R_r = \alpha + \beta\sigma_t^2 + \varepsilon_t + \theta\varepsilon_{t-1}$$
$$\sigma_t^2 = \delta_0 + \delta_1\varepsilon_{t-1}^2 + \delta_2\sigma_{t-1}^2$$

The estimates have been calculated using log and simple returns. The t-statistics are obtained using Bollerslev-Wooldridge robust standard errors.

Panel A

Data from the full life of the S&P index from 1/4/1950 to 12/21/2009.

	B
Log Returns	2.985700
	(2.625340)***
Simple Returns	3.434317
	(2.926564)***

Panel B

Data from the S&P index from 1/2/1980 to 12/21/2009.

	B
Log Returns	2.485555
	(1.672631)
Simple Returns	2.958536
	(2.018502)**

,* significant at 5% or 1% respectively

Table 2 Empirical relationship between log and simple returns in periods of differing variance
Calculations on Dow Jones Index. Investigation Period 2 January 1897 to 23 March 2009

All Days (30643 observations)				
	Mean	Variance	t	
log return	0.00018	0.00012	2.916***	
simple return	0.00024	0.00012	3.877***	
Ratio simple/log	1.32713	0.99602		
Expected mean log[1]	0.00018		Ratio actual/expected mean log	0.99913
Days where Absolute value of Simple Ret > 5% (138 observations)				
	Mean	Variance	t	
log return	-0.00461	0.00524	-0.747	
simple return	-0.00201	0.00519	-0.327	
Ratio simple/log	0.43596	0.98929		
Expected mean log[1]	-0.00460		Ratio actual/expected mean log	1.00091
Days where Absolute value of Simple Ret > 2% (1753 observations)				
	Mean	Variance	t	
log return	-0.00218	0.00117	-2.669**	
simple return	-0.00160	0.00117	-1.957*	
Ratio simple/log	0.73119	0.99492		
Expected mean log[1]	-0.00218		Ratio actual/expected mean log	1.00154
Days where Absolute value of Simple Ret > 1% (7071 observations)				
	Mean	Variance		t
log return	-0.00022	0.00044		-0.866
simple return	0.00000	0.00044		0.011
Ratio simple/log	-0.01281	0.99556		
Expected mean log[1]	-0.00021		Ratio actual/expected mean log	1.00258
Days where Absolute value of Simple Ret > 0.5% (15347 observations)				
	Mean	Variance		t
log return	0.00023	0.00023		1.845*
simple return	0.00034	0.00023		2.790***
Ratio simple/log	1.50887	0.99582		
Expected mean log[1]	0.00023		Ratio actual/expected mean log	0.99861

*, **,*** significant at 10%,5% or 1% respectively [1]mean simple – 0.5 variance

Table 3

The Turn-of –the- Month Effect

Calculations on Dow Jones Index Investigation Period 2 January 1897 to 23 March 2009

Log Returns	Day -1	Day +1	Day +2	Day +3	Day (-1.+3)	Other days	Difference
Mean Daily Ret %	0.0957	0.1163	0.1350	0.1151	0.1155	-0.0025	0.1181
Variance	0.01022	0.01189	0.01082	0.01095	0.01097	0.01208	
Number	1343	1344	1344	1344	5375	25268	
t-stat	3.4690	3.9085	4.7593	4.0326	8.0882	-0.3652	7.4389

Simple Returns	Day -1	Day +1	Day +2	Day +3	Day (-1.+3)	Other days	Difference
Mean Daily Ret %	0.1008	0.1221	0.1405	0.1207	0.1210	0.0035	0.1175
Variance	0.01020	0.01182	0.01076	0.01106	0.01096	0.01203	
Number	1343	1344	1344	1344	5375	25268	
t-stat	3.6578	4.1177	4.9671	4.2066	8.4777	0.5081	7.4121

All t-stats are significant different from 0 at the 1% level except those for ‘Other days’ which are not significant

Expected mean log (mean simple – 0.5 variance) (Formula. 1)						
	0.0957	0.1162	0.1351	0.1152	0.1155	-0.0025
Difference mean simple – mean log						
	0.0051	0.0058	0.0055	0.0056	0.0055	0.006

Table 4 Moving Average Rules								
Calculations on Dow Jones Index. Investigation Period 2 January 1897 to 23 March 2009.								
Parameters (x,y,z)		**N(Buy)**	**N(Sell)**	**Mean Buy %**	**Mean Sell %**	**Buy variance**	**Sell variance**	**Buy-Sell %**
1,50,0	Log	17927	12666	0.04380	-0.0183	0.007555	0.018066	0.0621
				(2.5049)***	(-3.1560)***			(4.5698)***
	Simple	17927	12666	0.047574	-0.0093	0.007551	0.017960	0.0569
				(2.2983)**	(-2.8958)***			(4.1944)***
Exp. Mean Log Formula. 1				0.0438	-0.0183			
	Simple – Log			0.0038	0.0090			
1,50,0.01	Log	14684	9646	0.05318	-0.0219	0.007619	0.021339	0.0621
				(2.9340)***	(-3.0380)***			(4.5425)***
	Simple			0.05670	-0.01125	0.0076222	0.0212040	0.0680
				(2.7173)***	(-2.7286)***			(4.1397)***
Exp. Mean Log Formula. 1				0.0529	-0.0219			
	Simple – Log			0.0035	0.0107			
*, **,*** significant at 10%,5% or 1% respectively								

Table 5 Trading Range Breakout Rules								
Calculations on Dow Jones Index. Investigation Period 2 January 1897 to 23 March 2009.								
Parameters (x,y,z)		**N(Buy)**	**N(Sell)**	**Mean Buy %**	**Mean Sell %**	**Buy variance**	**Sell variance**	**Buy-Sell %**
1,50,0	Log	866	478	0.4330	0.0073	0.089086	0.22736	0.4257
				(2.6595)***	(0.028590)			(1.7698)*
	Simple			0.47100	0.13333	0.089016	0.21897	0.33767
				(2.9280)***	(0.672406)			(1.4258)
Exp. Mean Log Formula. 1				0.4265	0.0238			
	Simple – Log			0.0380	0.1260			
1,50,0.01	Log	308	297	0.7200	-0.1044	0.13078	0.29.157	0.8244
				(2.2340)***	(-0.325673)			(2.1984)**
	Simple			0.7750	0.0691	0.13101	0.27790	0.7059
				(2.4463)***	(0.2109)			(1.9137)*
Exp. Mean Log Formula. 1				0.7095	-0.0699			
	Simple – Log			0.0550	0.1735			
*, **,*** significant at 10%,5% or 1% respectively								

Table 6 Returns after large price changes

Calculations on Dow Jones Index Investigation Period 2 January 1897 to 23 March 2009.

	Mean	Variance	t
Panel A – Positive Prior Price Changes			
Days where Previous Simple Ret > 5% (60 observations)			
	Mean	Variance	t
Mean log return	0.00302	0.00102	0.730
Mean simple return	0.00353	0.00105	0.844
Ratio simple/log	1.16987	1.02469	
Days where Previous Simple Ret > 2.5% (447 observations)			
	Mean	Variance	t
Mean log return	0.00060	0.00042	0.616
Mean simple return	0.00081	0.00042	0.826
Ratio simple/log	1.35248	1.01560	
Days where Previous Simple Ret > 2% (810 observations)			
	Mean	Variance	t
Mean log return	0.00122	0.00035	1.844*
Mean simple return	0.00140	0.00036	2.092**
Ratio simple/log	1.14657	1.02193	
Panel B – Negative Prior Price Changes			
Days where Previous Simple Ret < -5% (79 observations)			
	Mean	Variance	t
Mean log return	0.00803	0.00174	1.714*
Mean simple return	0.00893	0.00176	1.891*
Ratio simple/log	1.11138	1.01382	
Days where Previous Simple Ret < -2.5% (544 observations))			
	Mean	Variance	t
Mean log return	0.00194	0.00079	1.611
Mean simple return	0.00234	0.00077	1.964*
Ratio simple/log	1.20222	0.97191	
Days where Previous Simple Ret < -2% (945 observations)			
	Mean	Variance	t
Mean log return	0.00115	0.00058	1.474
Mean simple return	0.00144	0.00057	1.861*
Ratio simple/log	1.24949	0.97910	

*, **,*** significant at 10%,5% or 1% respectively

Appendix A - Proof That the Difference between Mean Log Returns and Mean Simple Returns depends on the Variance of Simple Returns

Mean sample return (simple)

$$= \frac{1}{n}\left[(1+r_1-1)+(1+r_2-1)+.......+(1+r_n-1)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} r_i$$

Mean sample return (log)

$$= \frac{1}{n}\left[(\ln(1+r_1)-\ln(1))+(\ln(1+r_2)-\ln(1))+.......+(\ln(1+r_n)-\ln(1))\right]$$

$$= \frac{1}{n}\left[\ln(1+r_1)+\ln(1+r_2)+.......+\ln(1+r_n)\right]$$

Now it is possible to expand ln(1 + x) using Taylor's series

$\ln(1 + x) = x – x^2/2 + x^3/3 – x^4/4 + …….$

Thus mean sample return (log)

$$= \frac{1}{n}\left[\left(r_1-\frac{r_1^2}{2}+...\right)+\left(r_2-\frac{r_2^2}{2}+...\right)+.......+\left(r_n-\frac{r_n^2}{2}+...\right)\right]$$

If cubic and higher terms in r_i can be neglected

Thus mean sample return (log)

$$= \frac{1}{n}\left[\left(r_1-\frac{r_1^2}{2}\right)+\left(r_2-\frac{r_2^2}{2}\right)+.......+\left(r_n-\frac{r_n^2}{2}\right)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} r_i - \frac{1}{2n}\sum_{i=1}^{n} r_i^2$$

Now sample variance (simple)

$$= \frac{1}{n}\left[r_1^2+r_2^2+......+r_n^2\right] - \frac{1}{n^2}\left[r_1+r_2+......+r_n\right]^2$$

If n is large the second term tends to zero so sample variance (simple)

$$\approx \frac{1}{n}\sum_{i=1}^{n} r_i^2$$

Thus

$$\bar{x}_L = \bar{x}_S - 0.5\sigma_S^2$$

Where: mean sample return (logarithmic) is $\bar{x}_L$

mean sample return (simple) is $\bar{x}_S$

sample variance (simple) is σ_S^2

Proof that the Sample Variance of Log Returns and the Sample Variance of Simple Returns are approximately equal

Sample variance (simple)

$$= \frac{1}{n}\left[r_1^2 + r_2^2 + \dots\dots + r_n^2\right] - \frac{1}{n}\left[r_1 + r_2 + \dots\dots + r_n\right]^2$$

Sample variance (log)

$$= \frac{1}{n}\left[\ln(1+r_1)^2 + \ln(1+r_2)^2 + \dots\dots + \ln(1+r_n)^2\right]$$
$$- \frac{1}{n}\left[\ln(1+r_1) + \ln(1+r_2) + \dots\dots + \ln(1+r_n)\right]^2$$

Now again use Taylor's expansion

$$= \frac{1}{n}\left[\left(r_1 - \frac{r_1^2}{2} + \dots\right)^2 + \left(r_2 - \frac{r_2^2}{2} + \dots\right)^2 + \dots\dots + \left(r_n - \frac{r_n^2}{2} + \dots\right)^2\right]$$
$$- \frac{1}{n}\left[\left(r_1 - \frac{r_1^2}{2} + \dots\right) + \left(r_2 - \frac{r_2^2}{2} + \dots\right) + \dots\dots + \left(r_n - \frac{r_n^2}{2} + \dots\right)\right]^2$$

Again neglecting cubic and higher terms in r_i

$$= \frac{1}{n}\left[r_1^2 + r_2^2 + \dots\dots + r_n^2\right] - \frac{1}{n}\left[r_1 + r_2 + \dots\dots + r_n\right]^2$$

Thus sample variance (simple) approximately equal to sample variance (log)

Appendix B

Example of How a Particular Expected Log Return Equates to Different Combinations of Mean and Variance of Simple Returns

The proofs in Appendices A do not depend on returns following any particular statistical distribution. In this Appendix a particular distribution has been chosen, without loss of generality, to generate an example of how a particular expected log return equates to different combinations of the mean of the simple returns and the variance of the distribution (where variance is approximately equal whether measured by log or simple returns). The lognormal distribution has been chosen because it is frequently assumed to hold in empirical investigations and because it is analytically tractable. However, the property of there being no one to one correspondence between expected log returns and expected simple returns holds for distributions in general.

In the case of the lognormal model, if the mean and variance of the log return r are μ and σ^2 respectively then the expected return and variance of simple returns are given by the formulae below (Campbell et al, 1997, p15):

$$E[R] = \exp(\mu + \sigma^2/2) - 1$$

$$V[R] = \exp(2\mu + \sigma^2).(\exp(\sigma^2) - 1)$$

Table 1 shows how for a constant expected log return the expected simple return may vary dramatically with the variance of the log return. As the variance of the log return increases the expected simple return increases.

In addition, by comparing Panel B with Panel A it can be seen that a distribution with a lower expected log return can have a higher expected simple return if the variance of log return is sufficiently high.

Appendix B - Relationship between expected simple return and expected value and variance of log returns assuming lognormal model.					
Expected log return	Variance of log return	Expected Simple return	Variance of simple return	Simple/log return	Simple/log variance
Panel A – Expected log return of 0.0001					
0.0001	0.00001	0.000105	0.00001	1.0501	1.0002
0.0001	0.00005	0.000125	0.00005	1.2501	1.0003
0.0001	0.00010	0.000150	0.00010	1.5001	1.0004
0.0001	0.0002	0.000200	0.00020	2.0002	1.0005
0.0001	0.0005	0.000350	0.00050	3.5006	1.0010
0.0001	0.001	0.000600	0.00100	6.0018	1.0017
0.0001	0.01	0.005113	0.01015	51.1303	1.0153
0.0001	0.02	0.010151	0.02061	101.5118	1.0307
0.0001	0.05	0.025418	0.05391	254.1766	1.0782
Panel B – Expected log return of 0.0002					
0.0002	0.00001	0.000205	0.00001	1.0251	1.0004
0.0002	0.00005	0.000225	0.00005	1.1251	1.0005
0.0002	0.0001	0.00025	0.00010	1.2502	1.0006
0.0002	0.0002	0.0003	0.00020	1.5002	1.0007
0.0002	0.0005	0.00045	0.00050	2.2505	1.0012
0.0002	0.001	0.0007	0.00100	3.5012	1.0019
0.0002	0.01	0.005214	0.01016	26.0677	1.0155
0.0002	0.02	0.010252	0.02062	51.2610	1.0309
0.0002	0.05	0.02552	0.05392	127.6010	1.0784

Appendix C Example of how in a Particular Period the Simple Return is Higher than the Log Return Due to a High Spread of Returns

Consider some return sets from Appendix B.

Expected log return	Variance of log return	Expected Simple return	Variance of simple return
Return set (a)			
0.0001	0.05000	0.025418	0.05391
Return set (b)			
0.0002	0.02	0.010252	0.02062

Return Set (b) has the higher expected log return. Return Set (a) has the higher expected simple return

Appendix D

The Implications of Using Different Return Measures for Significance Testing

Testing Whether Expected Returns Differ from Zero

The sample t-statistic is conventionally given by,

$$\frac{\bar{x}}{\left(\sigma^{2} / N\right)^{0.5}} \qquad (D1)$$

where $\bar{x}$ is the mean sample return, σ^2 is the sample variance and N is the number of observations. In the literature generally no attention is given to whether returns are calculated using logarithmic or simple returns.

Now from formula (1), if logarithmic returns are used, $D1$ is approximately equal to

$$\frac{\bar{x}_s - 0.5\sigma_s^2}{\left(\sigma_s{}^{2} / N\right)^{0.5}} \qquad (D2)$$

Where the subscript s denotes simple returns.
Now if simple returns are used, $D1$ is equal to

$$\frac{\bar{x}_s}{\left(\sigma_s{}^{2} / N\right)^{0.5}} \qquad (D3)$$

The implications for the size of the t-statistics depends on the relative size of $\bar{x}_S$ and σ_S^2. $D2$ is always smaller than $D3$. So if $\bar{x}_S$ is positive calculating t-statistics using simple returns will show higher significance levels. Conversely, if $\bar{x}_S$ is negative, calculating t-statistics using logarithmic returns will show higher significance levels.

<u>Testing Whether the Expected Returns in two Populations Differ</u>

The sample t-statistic is conventionally given by,

$$\frac{\bar{x}_1 - \bar{x}_2}{\left(\sigma_1{}^2 / N_1 - \sigma_2{}^2 / N_2\right)^{0.5}} \qquad (D4)$$

where $\bar{x}_i$ is the mean sample return of population i, σ_i^2 is the sample variance of population i and N_i is the number of observations in population i. In the literature generally no attention is given to whether returns are calculated using logarithmic or simple returns.

Now from formula (1) if logarithmic returns are used, *D*4 is approximately equal to

$$\frac{\bar{x}_{s1} - 0.5\sigma_{s1}^2 - \bar{x}_{s2} - 0.5\sigma_{s2}^2}{\left(\sigma_{s1}{}^2 / N_{s1} - \sigma_{s2}{}^2 / N_{s2}\right)^{0.5}} \qquad (D5)$$

Now if simple returns are used, *D*4 is approximately equal to

$$\frac{\bar{x}_{s1} - \bar{x}_{s2}}{\left(\sigma_{s1}{}^2 / N_{s1} - \sigma_{s2}{}^2 / N_{s2}\right)^{0.5}} \qquad (D6)$$

Now *D*6 may be smaller or larger than *D*5 depending on the relative size of σ_1^2 and σ_2^2, the variances of the two populations. Thus tests of significance for differences in expected returns between two populations will differ depending on the type of return used if the two populations have different variances.

Appendix E

An example decomposing the bid-ask spread into its information and non-information components

Our sample consists of institutional trades of the 20 major companies listed in the London Stock Exchange (LSE) over the time period of 2008-2010.[10] We have a balanced panel of 1006 observations for each company providing an overall sample of 20120 daily observations. These institutional trades account for about 80% of total LSE volume and market capitalisation. Institutional trades are defined by the LSE as a block of 10,000 shares or greater. We obtain the complete record of manager and non manager institutional transactions from Directors Ltd.[11]

Extensive theoretical literature (Huang and Stoll, 1997; Lin et al, 1995; Madhavan, Richardson and Roomans, 1997; to name but a few) decomposes market liquidity proxied by the bid-ask spread into its non-information (non manager trades) and information (manager trades) components. The non-information component comprises the direct costs of inventory holding and order processing while the information component is associated with the costs of asymmetric information. The latter is commonly known as the adverse selection costs of trading. Its isolation and use in modelling market liquidity reveals the magnitude of the influence of asymmetric information on trading costs.

We compute the adverse selection component of total trading costs following the methods of Madhavan, Richardson and Roomans (1997, henceforth MRR) and Huang and Stoll (1997, henceforth HS). The MRR propose the following model for equity price changes:

$$\Delta p_t = \alpha + (\phi + \theta) Q_t - (\phi + \rho\theta) Q_{t-1} + u_t \tag{E1}$$

Where, Δ is the first difference operator and p_t denotes the transaction price of security at time t. Q_t is a trade initiation indicator variable such that $Q_t = +1$ implies buyer initiated trade; $Q_t = -1$ implies seller initiated trade and $Q_t = 0$ denotes pre-negotiated trades (crosses) which occur within bid-ask spread. [12] The constant, α , represents the drift in prices; and u_t, a random error term, embeds the noises associated with price discreteness. ϕ measures market-makers' direct cost of supplying liquidity per share (transaction costs component). Theta (θ) is the information asymmetry parameter which measures the magnitude of the adverse selection cost. The rho (ρ) is the autocorrelation coefficient of order flow which can also be defined as $\rho = 2\gamma - (1-\beta)$; where the parameters γ and β respectively denote the probabilities of trade flow continuation and mid-quote execution.[13] Equation (E1) expresses changes in

[10] Appendix E - Table 1 lists the 20 companies whose institutional trades are analyzed in this paper.
[11] Managers are defined as individuals that own a minimum of 5% of the outstanding shares of a company.
[12] The LSE provides information on the direction of the institutional trade, therefore we do not require a procedure such as the tick rule in order to identify the trade purchases and sales.
[13] For a detailed exposition of this price evolution mechanism readers are refereed to MRR (1997).

security price as a function of order (buy and sell) flows, transaction costs, adverse selection costs and the noises associated with price discreteness. MRR suggest estimating the price formation equation by Generalized Method of Moments (GMM) under the following moment restrictions:

$$E\left[Q_tQ_{t-1} - Q^2{}_t\rho\right] = 0,\ E\left[\left|Q_t\right| - (1-\theta)\right] = 0,\ \ E\left[u_t - \alpha\right] = 0,$$
$$E\left[(u_t - \alpha)Q_t\right] = 0,\ \ E\left[(u_t - \alpha)Q_{t-1}\right] = 0 \tag{E2}$$

The first moment defines the autocorrelation in trade initiation of trades, the second moment is the crossing probability, the third moment defines the drift term, α , as the average pricing error. The last two moments are OLS normal equations. We estimate the parameters of Equation (E2) by Generalized Method of Moments (GMM) estimator, subject to the moment restrictions given in (2), for each company of our sample. The MRR adverse selection component (AS_{MRR}) is calculated as:

$$AS_{MRR} = \frac{\hat{\theta}}{(\hat{\phi} + \hat{\theta})} \tag{E3}$$

The implied expected spread is given by $2(\hat{\phi} + \hat{\theta})$ and the implied effective spread by $\left(1-\hat{\lambda}\right)2(\hat{\phi} + \hat{\theta})$. The HS adverse selection component is computed by estimating the following regression by ordinary least squares at firm level:

$$\Delta p_t = \beta_1 Q_t + \beta_2 Q_{t-1} + \beta_3 Q_{A,t-1} + e_t \tag{E4}$$

Where Δp_t represents the change in the transaction price prior to the quoted spread at time t; Q_t equals 1 (-1) if the trade is a sell (buy) at time t. Following Heflin and Shaw (2000) we use a "combined" buy/sell indicator, $Q_{A,t-1}$, which equals 1 (-1, 0) if the sum of Q_{t-1} across all the trades is positive (negative, zero) to capture the market-wide pressure on the inventory cost component of the bid-ask spread. Assuming that the number of trade purchases and sales are equal, the estimated information cost component of the bid-ask spread is equal to $2\left(\beta_2 + \beta_1\right)$.[14]

The objective of this study is to derive the mean percentage error using the MRR and HS methodology of the distinction between the manager and non manager trades. This is calculated by comparing each method's prediction of a manager (non manager) trade with the actual institutional ownership trade data obtained from Directors Ltd. The percentage error is then averaged over the 20 firms

[14] Huang and Stoll (1997) develop a technique using an estimated trade reversal probability as an alternative to the aggregate buy/sell indicator but the problem with this measure is that it may produce negative empirical estimates of the information cost component of the bid-ask spread. Hence we decompose spread utilizing the aggregate buy/sell indicator.

over the two year data period in our sample. In model 1 we compute the change in the transaction price using the logarithmic transformation and in model 2 we obtain the change in the transaction price using level differences. The results can be seen in Appendix E - Table 2.[15]

Appendix E - Table 1: Companies Used in Our Sample

Argos Limited, Asda Stores Ltd, Blockbuster Entertainment Ltd, BP, Bradford and Bingley, British Airways, Cadburry Schweppes, Glaxo, Harrods Ltd, HSBC, House of Fraser, Lewis Partnership, J Sainsbury, Ladbrokes, Prudential, Royal Bank of Scotland, SKY BSB, Tesco, Whitbread, William Hill

Source: Hemscott (www.hemscott.com)

Appendix E - Table 2

Model **1 - Log Returns** **2 – Simple Returns**	**Methodology**	**Mean Percentage Error**
1	MRR	10.1%
1	HS	10.8%
2	MRR	29.2%
2	HS	30.1%

[15] Results appear quantitatively similar when we use the Lin et al (1995) methodology. These results are not reported here to save space but are available from the authors upon request. Van Ness et al., (2001) also report that all spread decomposition models provide similar empirical results.

References

Arsad, Z., and J. A. Coutts (1997) 'Security Price Anomalies in the London International Stock Exchange' *Applied Financial economics*, 7, 455 - 464.

Atkins, A. B., and E. A. Dyl (1990) 'Price Reversals, Bid-Ask Spreads, and Market Efficiency' *Journal of Financial and Quantitative Analysis*, Vol. 25, No. 4, 535 - 547.

Bali, T. G. and L. Peng (2006) 'Is there a Risk-Return Trade-off? Evidence form High-Frequency Data', *Journal of Applied Econometrics*, 21, 1169-1198.

Barber, B. M. and J. D. Lyon (1997) 'Detecting Long-run Abnormal Returns: The Empirical Power and Specification of Test Statistics', *Journal of Financial Economics*,. Vol. 43, No. 3, 341-72.

Bokhari, J., C. Cai, R. Hudson and K. Keasey (2005) 'The Predictive Ability and Profitability of Technical Trading Rules: Does Company Size Matter?' *Economics Letters*, 86, 21-27.

Bremer, M., and R. J Sweeney (1991) 'The Reversal of Large Stock-Price Decreases' *The Journal of Finance*, Vol. XLVI, No. 2, 747 - 754.

Brown, K. C., W. V. Harlow and S. M Tinic (1988) 'Risk Aversion, Uncertain Information, and Market Efficiency', *Journal of Financial Economics*, 22, 355 - 385.

Brown, K. C., W. V. Harlow and S. M Tinic (1993) 'The Risk and Required Return of Common Stock following Major Price Innovations', *Journal of Financial and Quantitative Analysis*, Vol. 28, No. 1, 101 - 116.

Campbell, J. Y., A. W. Lo and A. C. Mackinlay (1997) *The Econometrics of Financial Markets*, New Jersey, Princeton University Press.

Chong, K., R. Hudson, K. Keasey and K. Littler (2005) 'Pre-Holiday Effects: International Evidence on the Decline and Reversal of a Stock Market Anomaly', *Journal of International Money and Finance*, 24, 1226-1236.

Cootner (1964) *The Random-character of Stock Market Prices*, Cambridge, M.I.T. Press.

Cox, D. R. and D. R. Peterson (1994) 'Stock Returns following Large One - Day Declines: Evidence on Short - Term Reversals and Longer - Term Performance' *The Journal of Finance*, Vol. XLIX, No. 1, 255 - 267.

Dann, L. Y., D. Mayers and R. J. Raab, Jr. (1977) 'Trading Rules, Large Blocks and the Speed of Price Adjustment', *Journal of Financial Economics,* 4, 3-22.

Dissanaike, G. (1994) 'On the Computation of Returns in Tests of the Stock Market Overreaction Hypothesis' *Journal of Banking and Finance*. 18, 1083-1094.

Dissanaike, G. and A. Le Fur (2003) 'On the Use of the Log CAR Measure in Event Studies' *Journal of Business Finance and Accounting*. 30(7) & (8), 1165-1170.

Fama, E. F. (1965) 'The Behavior of Stock Market Prices.' *The Journal of Business*, 37, 34-105.

Ferri, M., G. and C. Min (1996) 'Evidence that the Stock Market Overreacts and Adjusts.' *The Journal of Portfolio Management*, 22, 3, 71-76.

Frino, A., E. Jarnecic, D. Johnstone and A. Lepone (2005) 'Bid-Ask Bounce and the Measurement of Price Behavior around Block Trades on the Australian Stock Exchange.' *Pacific-Basin Finance Journal*, 13, 247-262.

French, K., (1980) 'Stock Returns and the Weekend Effect', *Journal of Financial Economics*, 8, 55-69.

Gemmill, G. (1996) 'Transparency and Liquidity: A Study of Block Trades on the London Stock Exchange under Different Publication Rules' *The Journal of Finance*, 51, No. 5, 1765 - 1790.

Gibbons, M.R. and P. Hess, (1981) 'Day-of-the-week effects and asset returns', *Journal of Business*, 54, 579-596.

Gregoriou, A (2008) 'The Asymmetry of the Price Impact of Block Trades and the Bid-Ask Spread', *Journal of Economic Studies*, 35, 2, 191-199.

Hansen, P. R., A. Lunde and J.M. Nason (2005) 'Testing the Significance of Calendar Effects', Federal Bank of Atlanta, Working Paper Series.

Heflin, F, and K.W. Shaw (2000), 'Blockholder Ownership and Market Liquidity', *Journal of Financial and Quantitative Analysis*, 35, 621-633.

Hotlhausen, R.W., R. W. Leftwich and D. Mayers (1987) 'The Effect of Large Block Transactions on Security Prices', *Journal of Financial Economics*, 19, 237-267.

Huang, R.D, and H.R. Stoll (1997), 'The Components of the Bid-Ask Spread: A General Approach', *Review of Financial Studies*, 10, 995-1034.

Hudson, R. S and C. V. Atanasova (2009) 'Equity Returns at the Turn of the Month: Further Confirmation and Insights' *Financial Analysts Journal*, 65 (July/August), 14-16.

Hudson, R, M. Dempsey and K. Keasey (1996) 'A note on the weak form efficiency of capital markets: The application of simple technical trading rules to UK stock prices - 1935 to 1994' *Journal of Banking & Finance*, 20, 1121-1132.

Hudson, R., K. Keasey and K. Littler (2001) 'The risk and return of UK equities following price innovations: a case of market inefficiency?' *Applied Financial Economics*, 11, 187 - 196.

Hughson, E., M. Stutzer and C. Yung (2006) 'The Misuse of Expected Returns', *Financial Analysts Journal*, 62, (Nov/Dec): 88-96.

Jacquier, E., A. Kane and A. J. Marcus (2003) 'Geometric or Arithmetic Mean: A Reconsideration', *Financial Analysts Journal*, 59, (Nov/Dec): 46-53.

Jorion, P. (2001) *Value at Risk*, New York, McGraw-Hill.

Kothari, S. P. and J. Warner (1997) 'Measuring Long-horizon Security Price Performance', *Journal of Financial Economics*,. Vol. 43, No. 3, 301-39.

Kraus, A. and H. R. Stoll (1972) 'Price Impacts of Block Trading on the New York Stock Exchange' *The Journal of Finance*, 27, No. 3, 569 - 588.

Kritzman, M. (1994) 'What Practitioners Need to Know… About Time Diversification ', *Financial Analysts Journal*,. Vol. 50, No. 1, Jan-Feb, 14-18.

Lasfer, M. A., A. Melnik and D. C. Thomas (2003) 'Short-term reaction of stock markets in stressful circumstances' *Journal of Banking and Finance* 27, 1959-1977

Lakonishok, J. and S. Smidt (1988) 'Are Seasonal Anomalies Real? A Ninety-Year Perspective', *Review of Financial Studies*, Vol. 1, No. 4, 403-425.

Lin, J, G. Sanger, and G.G. Booth (1995) 'Trade Size and Components of the Bid-Ask Spread', *Review of Financial Studies*, Vol. 8, 1153-1183.

Lo, A. W., H. Mamaysky and J. Wang (2000) 'Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation', *Journal of Finance*, LV, 4, 1705-1765.

Madhavan, A. and M. Cheng (1997) ‘In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets’, *The Review of Financial Studies*, 10, 1, 175-203.

Madhavan, A, M. Richardson, and M. Roomans (1997) ‘Why do Security Prices Change? A transaction-level analysis of NYSE stocks’, *The Review of Financial Studies*, Vol. 10, 1035-1064.

Mazouz, K., N.L. Joseph, and J. Joulmer (2009) ‘Stock price reactions following large one-day price changes: UK evidence’, *Journal of Banking and Finance* 33: 1481-1493.

McConnell, J. and W. Xu (2008) ‘Equity Returns at the Turn of the Month’, *Financial Analysts Journal*, 64 (March/April), 49-64.

Michayluk, D. and K.L. Neuhauser (2006) ‘Investor Overreaction during Market Declines: Evidence from the 1997 Asian Financial Crisis’ *The Journal of Financial Research*, 29, 2, 217-234.

Park, I. (1995) ‘A Market Microstructure Explanation for Predictable Variations in Stock Returns following Large Price Changes’ *Journal of Financial and Quantitative Analysis*, Vol 30, No. 2, 241 - 256.

Pritamani, M., V. Singal (2001) ‘Return predictability following large price changes and information releases’ *Journal of Banking and Finance*, 25, 631-656.

Ratner, M. and R.P.C. Leal (1999) ‘Tests of technical trading strategies in the emerging equity markets of Latin America and Asia’, *Journal of Banking and Finance*, 23, 1887-1905.

Roll, R. (1983) ‘On Computing Mean Returns and the Small Firm Premium’, *Journal of Financial Economics*, 12, 371-386.

Rozeff, M and W. Kinney (1976) ‘Capital Market Seasonality: The Case of Stock Returns’, *Journal of Financial Economics*, 3, 379-402.

Schnusenberg, O. and J. Madura (2001) ‘Do U.S. Stock Market Indexes Over- Or Underreact?’ *The Journal of Financial Research* 24, 2, 179-204.

Spyrou, S., K. Kassimatis and E. Galariotis (2007) ‘Short-term Overrreaction, Underreaction and Efficient Reaction: Evidence from the London Stock Exchange’ *Applied Financial Economics*, Vol. 17, issue 3, 221-235.

Sullivan, R., A. Timmermann and H. White (1999) ‘Data-Snooping, Technical Trading Rule Performance, and the Bootstrap’, *Journal of Finance*, LIV, 5, 1647-1691.

Sullivan, R., A. Timmermann and H. White (2001) ‘Dangers of Data Mining: The Case of Calendar Effects in Stock Returns’, *Journal of Econometrics*, 105, 249-286.

Van Ness, B.F, R.A. Van Ness and R.S.Warr, (2001) ‘How Well Do Adverse Selection Components Measure Adverse Selection?’ *Financial Management*, Vol. 30, 77-98.

Wang, K., Y. Li and J. Erickson (1997) ‘A New Look at the Monday Effect’, *Journal of Finance*, 52, 5, 2171-2186.