EVENT STUDY METHODS AND EVIDENCE ON THEIR PERFORMANCE

Seth Armitage

University of Edinburgh

Abstract. The paper outlines widely used methods of estimating abnormal returns and testing their significance, highlights respects in which they differ conceptually, and reviews research comparing results they produce in various empirical contexts. Direct evidence on the performance of different methods is available from simulation experiments in which known levels of abnormal return are added. The market model is most commonly used to generate expected returns and no better alternative has yet been found despite the weak relationship between beta and actual returns. Choice of procedure for significance testing depends on the characteristics of the data. The evidence indicates that in many cases the best procedure is to standardise market model abnormal returns by their time series standard errors of regression and use the *t*-test. Alternatively a rank test appears to be at least as powerful. If errors are cross-correlated or increase in variance during the test period, other methods discussed should be used.

Keywords. Abnormal returns; event studies; expected returns; market model; significance tests; simulations.

1. Introduction

There exists a huge body of research in financial economics which seeks to measure unexpected or abnormal returns of shares. The return on a share is correlated to some extent with the return on the stock market of which it is part, and in the long term at least, riskier shares should earn higher returns if investors are risk averse. So a share's expected return over a period is the return on the market over that period, often with an adjustment for the share's risk. Its actual return will be affected by other factors but attempts to model returns using factors beyond market return and risk have not caught on in empirical work. The abnormal return for a period is simply the actual less the expected return.

Much of the research involving abnormal returns consists of tests of the efficient markets hypothesis that share prices reflect all available information, so there is no way greater (or less than) normal returns could be earned except by chance. Many studies have sought to estimate abnormal returns accruing from a set of investment decisions, for example following a systematic investment rule or the advice of investment analysts. Perhaps even more common are 'event studies' which estimate abnormal returns at and around the time of some event relating to the shares concerned, for example the announcement of a rights issue or a takeover bid. All of this research requires an estimate of expected or normal returns for shares or portfolios over some period of interest, the test or event period, which in many studies is of the same length for each share but over different calendar dates. In the case of a test of an investment rule (eg buy shares

0950-0804/95/01 0025-28 JOURNAL OF ECONOMIC SURVEYS Vol. 8, No. 4
© Basil Blackwell Ltd. 1995, 108 Cowley Rd., Oxford OX4 1JF, UK and 238 Main St, Cambridge, MA 02142, USA.

with a low price/earnings ratio), the test period starts when the rule is implemented; in an event study it is the date of the event and the surrounding period. The term 'event study methodology' has come to refer generally to procedures for estimating abnormal returns and testing their significance, though the application of these procedures is not limited to true event studies.

Studies are far from homogeneous in their use of such procedures. The reason is partly that there are some factors which should affect choice of method, for example whether abnormal returns are cross-correlated, as discussed below. But more of the variety can probably be attributed to a process of experimentation with different models and methods in different contexts, and of researchers becoming familiar with a widening range of possibilities. Indeed, there is a bewildering array of factors to consider in designing an abnormal returns study and appraising its results (Beaver, 1982, gives a 'partial list of selected research design issues' containing 42 items). Also empirical evidence on the effectiveness of alternative methods only became available during the 1980s.

The purpose of this paper is to provide a compact review of the main methods and of research concerning them. Bringing together and comparing these variants may be helpful both in reading the many studies which use them and in designing new empirical research. The paper does not derive or assess theories of expected returns but presents a number of ways in which abnormal returns have actually been estimated, explains how they differ and reviews evidence on their performance. The choice of method sometimes receives little or no discussion, perhaps because the underlying idea is so familiar, and in some cases because 'when the stock-price response to an event is large and concentrated in a few days, the way one estimates daily expected returns (normal returns) in calculating abnormal returns has little effect on inferences' (Fama, 1991, p. 1601). But the evidence is clear that the choice can affect the results in some circumstances. Given the very wide application of abnormal returns methodology, a review of the choices and the effects they can have seems worthwhile.

The issues considered are grouped under two headings, choice of model and choice of significance test. On choice of model, the evidence suggests that the market model will perform as well as, if not better than, any alternative in most circumstances. On choice of significance test, the evidence favours a t-test or a rank test on average abnormal returns where the individual abnormal return for each share has been standardised by dividing it by the time series standard error of the regression used to estimate the market model coefficients. This test is preferable unless abnormal returns exhibit significant cross-correlation over time or increase in variance during the test period, in which case tests which allow for these circumstances will perform better. In simulations in which levels of abnormal return are known, performance is measured by both the rejection rate of the null hypothesis of no abnormal return when it is true (type 1 error) and the acceptance rate of the null when it is false (type 2 error). If type 1 errors are greater than would be expected given the significance level, the method is unreliable; if type 1 errors are acceptable, power is measured by the sensitivity of the method to the presence of abnormal returns.

© Basil Blackwell 1995

The contents of the paper are as follows. Section 2 presents models of expected returns whose performance has been tested and Section 3 outlines the methodology of simulation research into estimation of abnormal returns. Sections 4 and 5 consider choice of model and significance test respectively, Section 6 reviews certain special issues and Section 7 provides summary recommendations on choice of method in the light of the evidence. The text and notes include examples of the application of abnormal returns methodology.

2. Models of expected returns

Perhaps the simplest model is to assume that, over any period t, a share i will earn the market rate of return, R_{mt}. Then the abnormal return, AR_{it}, is the actual return, R_{it}, less R_{mt}. Call this the *index model*. It is used, for example, by Lakonishok and Vermaelen (1990) to measure abnormal returns from selling shares to companies which offer to repurchase them via tender offers.

Another simple model is to assume that the share earns the same return as it does on average during an estimation period before or around the test period. Then

$$AR_{it} = R_{it} - \bar{R}_i, \tag{1}$$

where $\bar{R}_i$ is the average return of the share during the estimation period. Call this the *average return model*. Masulis's (1980) investigation into the share price response to announcements of changes in gearing uses this model and the average daily return is calculated from returns for 60 days before and 60 after the event period.

The commonest approach is to estimate the relationship between a share's returns and returns on the market by ordinary least squares (OLS) regression and use this relationship to estimate expected returns, given returns on the market. This is the *market model*, a one factor OLS regression equation

$$R_{it} = \alpha_i + \beta_i R_{mt} + e_{it}, \tag{2}$$

where α_i and β_i are regression coefficients and e_{it} is the error term. Having calculated estimates of α_i and β_i with data from an estimation period, the expected return is given by inserting the estimated values of α_i and β_i together with the actual return on the market. So

$$AR_{it} = R_{it} - (\alpha_i + \beta_i R_{mt}). \tag{3}$$

Market model abnormal returns are prediction errors if the test period is distinct from the estimation period, but in many studies the test period is a subset of the estimation period, in which case the abnormal returns are given by the relevant subset of regression error terms, often referred to as residuals.

This model was used in Fama *et al.*'s (1969) examination of abnormal returns around the announcement of stock splits, in which a company increases the number of its shares in issue without raising new capital. This is the earliest and most influential event study. It uses monthly data and the estimation period for

© Basil Blackwell 1995

each share's market model coefficients is all months for 1926–1960 for which data is available, excluding the fifteen months preceding each stock split and, for splits followed by dividend increases, the fifteen months afterwards. The test or event period is the month of the stock split announcement together with the 29 months before and after, for which average cumulative abnormal returns are calculated.

Unlike the three models presented so far, which are not derived from theory except in an informal sense, the *capital asset pricing model* (CAPM) is a true theoretical model, resulting from reasoning from a set of assumptions. The derivation can be found in most finance textbooks, for example Sharpe and Alexander (1990, ch. 8). The model is

$$E(R_{it}) = R_{ft} + \beta_i[E(R_{mt}) - R_{ft}], \tag{4}$$

where $E(R_{it})$ is the expected or normal return on share i for time t, R_{ft} is some measure of 'the risk-free rate of interest', $E(R_{mt})$ is some measure of the expected return on the appropriate stock market and β_i is the covariance of R_{it} with R_{mt} over some estimation period ($\text{cov}[R_{it}, R_{mt}]$) divided by the variance of R_{mt} over that period ($s^2[R_{mt}]$). The theory is that ex ante, expected values of R_{mt} and β_{it} determine $E(R_{it})$, but ex post values are usually substituted as proxies. Beta is the model's measure of the risk of the share, and investors are assumed to require and expect a higher return for more risk. The abnormal return on share i for time t is estimated by subtracting the actual return, R_{it}, from the expected return $E(R_{it})$. The market model is a simple regression model, but it can be viewed as a version of the CAPM by interpreting α_i in equation (3) as an estimate of $R_{ft}(1 - \beta_i)$. The formula for β_i, $\text{cov}(R_{it}, R_{mt})/s^2(R_{mt})$, is exactly the same as that required by the CAPM.

Studies of the performance of managed funds commonly use the CAPM as a benchmark, as does Dimson and Marsh's (1984) study of the accuracy of analysts' forecasts of share returns. In the latter, betas are estimated by equation (2) using monthly returns over five years before the start of the test period, the length of which varies with each forecast (usually one year). Analysts were asked to forecast abnormal returns and these forecasts are then compared with actual abnormal returns as measured by $R_{it} - E(R_{it})$, where $E(R_{it})$ is given by equation (4).[1]

A variant of the CAPM which was developed by Fama and MacBeth (1973) starts with shares of different betas and then, for each month, regresses the share returns for that month against the betas. This has become known as the *Fama–MacBeth model*. Expected returns on shares for a given time are

$$E(R_{it}) = \alpha_{1t} + \alpha_{2t}\beta_{it}, \tag{5}$$

where α_{1t} and α_{2t} are cross-sectional regression coefficients for time t of returns against beta and β_{it} is the actual beta of share i at time t. In this model, α_{1t} is interpreted as the return on a zero-beta portfolio, which theoretically should equal the risk-free rate but was not found to. Further details of this model are given in footnote 4.

© Basil Blackwell 1995

In the *control portfolio model*, returns of a test portfolio are compared with those of a control portfolio designed to have the same risk, measured by beta. Abnormal returns are measured by subtracting control portfolio returns from those of the test portfolio during the test period. For example, Reinganum (1981) constructs portfolios, all designed to have a beta of one, either from shares reporting unexpectedly high earnings per share (EPS) or from shares with unexpectedly low EPS, and calculates the difference in subsequent returns between high and low EPS portfolios. It is well known that ex post betas do not explain much of the difference between actual returns on shares or portfolios. For example, 'Fama and French (1992) find that the relation between β and average returns on ... stocks for 1963–1990 is feeble, even when β is the only explanatory variable' (Fama, 1991, p. 1592). This has motivated the continuing search for alternatives such as multifactor models and consumption β (the slope in the regression of a share's returns on growth rates of per capita consumption), and some of these alternatives may have more explanatory power. However, none have gained acceptance in empirical work, which if anything has become increasingly standardised on use of the market model. The trouble seems to be that no other model has been found for which there is both a theoretical case and consistent supporting evidence.

For example, there is substantial evidence for a long term 'small companies effect' — after adjusting for risk, share returns of smaller companies are higher than of larger ones, so it is argued that expected returns should reflect this too. But this is not merely unsupported by any theory: it is an efficient markets 'anomaly'. Identification of the size effect is credited to Banz (1981) and papers which discuss its impact include Dimson and Marsh (1986) and Chopra, Lakonishok and Ritter (1992). In a summary of a paper on the stock market response to announcements of dividend cuts, Marsh writes 'abnormal returns are defined simply as actual returns (capital gain plus dividends) less the returns on a control group of companies of similar size' (*Financial Times*, 12 August 1992). This remains an unusual definition. For a review of anomalies and models of expected returns, see Fama (1991).

3. Simulation experiments

In this section, we outline the construction of studies which use simulated data to test different methods, and which will be referred to in the following sections. The most well-known studies are probably those by Brown and Warner (1980, 1985). In both, they construct 250 sample portfolios each of 50 shares randomly selected with replacement. The 1980 study uses monthly data. For each share a hypothetical event month is randomly selected from months between June 1944 and March 1968 and the same level of positive abnormal monthly return (eg 1%) is artificially added to the actual return of each share in that month. Market model coefficients are estimated using 79 monthly observations of the share's returns from the University of Chicago's Centre for Research in Securities Prices (CRSP) and of returns on the CRSP equally weighted index from month −89 to

© Basil Blackwell 1995

month −11, with the event month being 0. The 1985 study is similar but uses daily CRSP returns. Hypothetical events could occur between 2 July 1962 and 31 December 1979. With the event day as 0, the estimation period is day −244 to −6 (239 observations, if available) and there is an event period from −5 to +5.

Both studies test the performance of the average return, index and market models for estimating expected returns with, for each model, several adjustments and significance tests. We refer to each model-adjustment-test permutation as a method. The control portfolio and Fama–MacBeth models are also tested in the monthly returns study (1980). Performance is tested under a variety of circumstances; with smaller sizes of sample portfolio than 50 shares, different tests of significance, event day uncertainty, all shares in a portfolio having the same event date ('clustering') instead of different ones, a value weighted instead of an equally weighted index, adjustments for non-synchronous or thin trading and for autocorrelation of errors.

Dyckman, Philbrick and Stephan (1984) carried out similar experiments independently of Brown and Warner, using CRSP daily data and 250 sample portfolios to test the first three models under a similar array of circumstances. The most notable difference is that their estimation period is divided and includes observations after the test period (for days +60 to +120) as well as before (−120 to −60). Also model coefficients are estimated using returns over consecutive three or five day periods as well as daily returns.

Another major study, by Collins and Dent (1984), is different in approach. Instead of starting with real returns and adding abnormal performance, weekly returns of hypothetical shares in portfolios of ten are artificially generated as follows. Consider the right hand side of equation (2), $\alpha_i + \beta_i R_{mt} + e_{it}$. First, artificial values are inserted: $\alpha_i = 0.001$ for each share; β_i ranges from 0.6 to 1.5 in increments of 0.1. Then R_{mt} is drawn from actual weekly closing values of the New York Stock Exchange (NYSE) index (a value weighted index) from July 1975 to August 1977. e_{it} is randomly generated within parameters for variance and covariance which are preset. This cocktail produces hypothetical observations of R_{it} which are then used to calculate new least squares estimates of α_i, β_i and the variance of the error, s_i^2.

Collins and Dent's aim is not to test the performance of different estimation models but to test four procedures for calculating test statistics for average abnormal returns, all using the market model. 100 simulations are carried out and in each case the estimation period is the first 100 'observations' and the test period the next five. As well as introducing abnormal returns in the test period, Collins and Dent introduce various levels of error term cross-correlation and higher levels of error variance in the event period than in the estimation period.

4. Choice of model

4.1. *Comparison of models*

Table 1 summarises the results of tests by various researchers of detection of abnormal returns using the market model compared with others models of share

© Basil Blackwell 1995

returns. In Brown and Warner's experiments, results for the index model and market model are similar, while the average return model performs badly when event dates are the same, but otherwise comparably. The overall conclusion from their monthly returns paper is that 'beyond a simple, one-factor market model, there is no evidence that more complicated methodologies convey any benefit' (p. 249), and this is confirmed for daily returns in their 1985 paper. It is also confirmed by Dyckman *et al.* who report that 'the market model performs significantly better (at the 5% level)' than either the index or average return models, but that 'the difference does not appear to be important since the rejection percentages are quite close in all cases' (p. 15). Their results are not affected by estimating coefficients from returns over three or five day periods rather than single days.

Krueger and Johnson (1991) use the F-ratio instead of the t-statistic to test the significance of abnormal returns arising from three efficient market anomalies using the index and beta-adjusted models, the latter being the market model without the α term, i.e. $AR_{it} = R_{it} - \beta_i R_{mt}$. The results from the two are similar but by no means identical. However, it is impossible to say which model is performing 'better' in this context, since the true extent of abnormal returns is unknown. This comment is applicable to comparisons of methods in any actual event study, as opposed to a simulation.

Brenner (1979) tests five models including the market, index and capital asset pricing models. He conducts a stock split study similar to that of Fama *et al.* (1969) and compares the results of the five models each using the same data. For

Table 1. Summary of Tests of Market Model Against Other Models of Share Returns

Market model: $AR_{it} = R_{it} - (\alpha_i + \beta_i R_{mt})$

Compared with:

Index model: $AR_{it} = R_{it} - R_{mt}$

Results similar; market model slightly more powerful.

Average return: $AR_{it} = R_{it} - \bar{R}_{it}$

Market model more powerful when shares have same event dates. Biased results during bull and bear markets.

CAPM: $AR_{it} = R_{it} - (R_{ft} + \beta_i [R_{mt} - R_{ft}])$

Results different.

Fama–MacBeth model: $AR_{it} = R_{it} - (\alpha_{1t} + \alpha_{2t}\beta_{it})$,

where α_{1t} and α_{2t} are coefficients of a regression of returns against betas for time t. Results different (Brenner); results similar (Brick *et al.*)

Control portfolio: $AR_{pt} = R_{pt} - R_{ct}$,

where R_{ct} is the return on control portfolio c, designed to have the same beta as p. Market model more powerful in some circumstances.

Note: See Section 4.1 for sources

© Basil Blackwell 1995

the second year after the split only, there are 'relatively small' but statistically significant differences between cumulative monthly abnormal returns as measured by the market and index models compared with the other three. No explanation is offered for this, but Brenner speculates that in some studies, 'the differences [between models] may be large enough to make definite conclusions about market efficiency dependent upon the ... model that is used' (p. 927).

This possibility is realised in Brick, Statman and Weaver's (1989) study of abnormal returns available from following the trades of insiders from June 1976 to July 1979. They use five estimation models; the Fama–MacBeth, market, index and average return models and the CAPM in the form of Jensen's (1968) performance index.[2] The data used are CRSP monthly share and value weighted index returns. The models tend to produce abnormal returns of the same sign but these are not statistically significant for the CAPM and average return model whereas they are for the other three. For example, cumulative abnormal returns over eleven months starting the month after insiders sold shares are −1.87% using the CAPM (t-statistic not given but not significant) but −6.30% with a t-statistic of −10.99 using the market model. Seyhun (1986), however, finds that the absolute values of abnormal returns following insider trading are greater using the CAPM than the market model.

Though Brenner and Brick *et al.* suggest that choice of model matters more than do Brown and Warner and Dyckman *et al.*, where they test the same models these studies are largely consistent. All four find that the market and index models give similar results. This is not surprising since the index model is a special case of the market model with $\alpha = 0$ and $\beta = 1$, and most portfolios not selected with reference to a factor associated with risk will have a beta fairly close to one.

As regards the average return model, this is not tested by Brenner while Brick *et al.* note that their results (insignificant abnormal returns) are consistent with those of Brown and Warner because of this estimator's poor performance in detecting abnormal returns when events share the same date. The superiority of the market over the average return model in this context is because market model errors of shares in each sample portfolio are much less cross-correlated than are discrepancies from the mean, so the variance of market model portfolio residuals is lower than the variance of portfolio returns. Klein and Rosenfeld (1987) use simulation to show that the average return model, but not index or market model, produces upwardly biased abnormal returns during a bull market and downwardly biased abnormal returns during a bear market, so this is a further reason for caution regarding the average return model.

Both Brenner and Brick *et al.* report statistically significant differences between results using the market model and the CAPM, which is not tested by Brown and Warner or Dyckman *et al.* This is the main reason for the difference of opinion regarding importance of choice of model. Results of the studies also diverge for the Fama–MacBeth model. Brown and Warner (1980) report little difference between results using this and the market model under a variety of conditions, and this is consistent with Brick *et al.* However, Brenner reports a small but statistically significant difference between the market and Fama–MacBeth models

© Basil Blackwell 1995

and little difference between the latter and the CAPM, which is inconsistent with Brick *et al.*[3]

The important point from Brenner and Brick *et al.* is the apparent difference in results for the two most widely used models, the market model and the CAPM. Ball and Brown (1968) used both models in their seminal test for monthly abnormal returns around announcements of unexpected changes in profit, but report that they produced results 'essentially the same' (p. 164). It is perhaps a pity that Brown and Warner or Dyckman *et al.* did not include the CAPM in their experiments. Its use as a benchmark has been deemed inappropriate in some circumstances. Dimson and Marsh (1986) and Seyhun (1986) note that market model abnormal returns are much less prone to bias due to the size effect (because a share with high returns has a high regression constant). Chopra *et al.* (1992) show how the finding of 'numerous empirical studies' that the return per unit of beta risk is less than $R_{mt} - R_{ft}$ required by the CAPM can distort results when portfolios under study have very different betas.

Finally we consider the control portfolio model. This is tested by Brown and Warner (1980) by constructing portfolios with an estimated beta of one and comparing their returns with returns on the market. Results from this method are similar to those from the market model using an equally weighted index, but exhibit over-rejection of the null hypothesis when a value weighted index is used. Performance is also poor when shares are clustered by risk, both for high and low beta portfolios. (All the models perform better with low than with high betas because error variances are lower with low betas. This at least partly explains the finding of earlier research that abnormal returns are negatively related to risk.)

The main conclusions up to this point are that the different models produce similar but not identical results and that the market model is the most reliable in the sense that, across each of the range of circumstances tested, it is always at least as powerful as the best alternative. This is consistent with Beaver's (1981) comparison of the econometric properties of several models. The average return model, though simple, should be avoided when event dates are shared or when events are in bull or bear markets.

4.2. *Choice of market index*

There is some evidence in Brown and Warner's studies that choice of market index affects the performance of the market model. It generally performs slightly better with the equally weighted than with the value weighted index, and the difference is significant when event dates are shared because of lower cross-correlation of residuals with the equally weighted index (see Section 5.1). However, Krueger and Johnson (1991), testing efficient market anomalies, obtain similar results using the CRSP equally weighted index and the NYSE composite index (value weighted); there are differences, but they believe that 'anomaly research findings are generally robust to market surrogate selection' (p. 579).

Thompson (1988) uses simulation with daily data to compare the standard market model with a version in which the appropriate industry index is substituted

© Basil Blackwell 1995

for the market index and with a two-factor version in which the industry index is included with the market index. The results are sufficiently similar for Thompson to state that 'it doesn't seem to matter which model is used' (p. 80), though he does not examine the case in which event dates coincide. He also reports that it makes little difference whether simple or compound returns are used (as do Ball and Brown, 1968, and Brown and Warner, 1985), and whether returns associated with extraneous events (to the one in question) are deleted.

4.3. *Estimation and event periods*

In practice, estimation periods range from 100 to 300 days for daily studies and from 24 to 60 months for monthly studies (Peterson, 1989, p. 38). Lengthening the estimation period involves a trade-off between greater precision of estimation of α_i and β_i and these coefficients becoming more 'out of date'. In some studies the estimation period comes on both sides of or after the event period if the pre-event period is unusual in some way (eg Mikkelson and Partch, 1986; see footnote 7) or if pre-event and post-event model coefficients are expected to differ, for example with a change in gearing (eg Masulis, 1980). Corrado and Zivney (1992) compare, by simulation, results from three test statistics using pre-event estimation periods of 239, 89 and 39 days. These are the t-test given in equation (8) (p. 19 below) and a sign and rank test (pp. 31–32). They were 'virtually unaffected' by an estimation period of 89 instead of 239 days, and a 39 day period produced 'only a slight deterioration of performance' (p. 477), so 100 days or more seems safe.

Identification of the correct event date is, of course, crucial in event studies. Choice of surrounding event period will depend very much on what is being studied. Two day event windows are common if the event date can be determined with precision, supplemented by cumulative abnormal returns for longer periods before and after. Not surprisingly, the shorter the event period, the easier it is to identify any abnormal return present (Dyckman *et al.*, 1984; Glascock *et al*, 1991). If there is uncertainty as to the event date, Dyckman *et al.* find that accumulating abnormal returns is preferable to choosing one of the days at random as the event date.

There is the further problem of what to do about 'contaminating' events within event periods, i.e. news distinct from the event in question which may affect the share price. Unrelated news should have a zero effect on average over a large sample, as Thompson (1988) finds. But some events may be systematically linked with others, for example news of an acquisition together with a share or bond issue or new loan. The usual treatment is to report results for both full (contaminated) and uncontaminated samples. But stock markets have many sources of information, while checks for contaminating news are usually restricted to a search of the *Wall Street Journal* or *Financial Times* index, so decontamination may be partial. (See Wright and Groff, 1986, and Thompson, Olsen and Dietrich, 1987, for investigations of the information dissemination process in the USA.)

© Basil Blackwell 1995

5. Choice of estimation procedure and significance test

A wide variety of approaches can be and have been taken to measuring the significance of estimated abnormal returns. We shall present some of these and indicate which lead to improvements in performance and in what circumstances. In this section abnormal returns are assumed to be estimated using the market model.

5.1. *Initial comparison of methods*

Most studies are concerned with the effect of some event or investment rule on many shares, so it is usually the case that average abnormal returns (errors) for one or more portfolios of N shares are calculated and tested for significance by the *t*-test. The question is how the test statistic is arrived at. One approach is to standardise each share's abnormal return by its estimation period standard error of regression, s_i, resulting in a standardised error, SE:

$$SE_{it} = e_{it}/s_i, \tag{6}$$

where

$$s_i = \sqrt{\left([1/T-2]\sum_{t=1}^{T} e_{it}^2\right)} \tag{7}$$

and T is the number of observations (days, months etc) in the estimation period. Standardised errors are comparable in terms of significance; each has an expected mean of zero and standard deviation of (very close to) one. The more volatile a share, the larger an abnormal return has to be to reach a given level of significance. The average standardised error for time t (ASE_t) has, applying the Central Limit Theorem, a standard deviation of $s(SE)/\sqrt{N}$, where s(SE) is the standard deviation of the SEs. But as this equals one, the test statistic is simply

$$\frac{ASE_t - 0}{s(SE)/\sqrt{N}} = \sqrt{N}(ASE_t), \text{ or} \tag{8}$$

$$= (1/N)\sum_{i=1}^{N} SE_{it}/(1/\sqrt{N}) = \sum_{i=1}^{N} SE_{it}/\sqrt{N},$$

Call this the *share time series* method. It is used by Dyckman *et al.* and is similar to the main test in Brown and Warner (1980).

Second, sample or *portfolio* errors, the average of the constituent shares' errors for each time t, could be treated as observations in their own right to calculate a time series standard deviation, $s(\bar{e})$. On this view,

$$s(\bar{e}) = \sqrt{\left([1/T-1]\sum_{t=1}^{T}[\bar{e}_t - \mu]^2\right)}, \tag{9}$$

© Basil Blackwell 1995

where μ is the mean of the portfolio residuals over estimation period T and

$$\bar{e}_t = (1/N) \sum_{i=1}^{N} e_{it}. \tag{10}$$

Since $\bar{e}_t$, the portfolio error, is treated as an observation rather than a mean of a sample, it is simply divided by the standard deviation to produce the test statistic $\bar{e}_t / s(\bar{e})$. Call this the *portfolio time series* method. It is the main test in Brown and Warner (1985) and, for example, Jaffe's (1974)[4] and Seyhun's (1986) studies on returns to trades by insiders use similar methods.

Third, observations at times other than t could be ignored, in which case the shares' errors at time t are treated as a sample without further ado, of which $\bar{e}_t$ is the mean and s_t is the cross-sectional standard deviation, so

$$s_t = \sqrt{\left([1/N - 1] \sum_{i=1}^{N} [e_{it} - \bar{e}_t]^2 \right)} \tag{11}$$

and the test statistic is

$$\frac{\bar{e}_t.}{s_t / \sqrt{N}}$$

Call this the *cross-sectional* method. It is used, for example, by Imhoff and Lobo (1984)[5] and is one of the tests examined by Collins and Dent.

For event periods of more than one return interval, say one day, average cumulative errors (ACE_D) are formed:

$$ACE_D = \sum_{t_1}^{t_2} AE_t, \tag{12}$$

where AE_t is the average error for event day t and D is the number of event days between and including t_1 and t_2. The test statistic for $ACED_D$ is the sum of the daily test statistics for t_1 to t_2, whichever method is used to calculate them, divided by $\sqrt{D}$.[6]

It may be asked in what circumstances each of these methods should be used. It is a fact that the variance of market model errors varies considerably from share to share, which is a good reason to use the share time series method and the evidence indicates that it provides a powerful significance test in most circumstances.

The portfolio time series method does not adjust for different error variances; each share's abnormal return is given the same weight in determining the portfolio's abnormal return. But it does allow for share returns and errors not being independently distributed because they are cross-correlated. This causes the standard deviation of average errors to be higher than if individual errors were independently distributed. Thus a test which assumes independence of individual

© Basil Blackwell 1995

errors will overestimate the significance of average errors by underestimating their standard deviation. The greater the cross-correlation of returns of shares in the portfolio, the higher the standard deviation of portfolio errors, so this method is preferable when cross-correlation exists. The cross-sectional method does not adjust either for share returns having different variances or being cross-correlated but does allow for the variance of returns being greater in the test period than in the estimation period.

Cross-correlation is possible when shares have the same event and estimation periods and likely when they are in the same industry. When errors are estimated over the same chronological period and shares are from the same industry, defined as a three digit Standard Industrial Classification (SIC) category, Collins and Dent (1984) report an average correlation coefficient of 0.18. They use weekly CRSP returns over a 100 week period from July 1975 to June 1977 and 'a value weighted portfolio'. The coefficient rises dramatically if portfolio rather than share errors are considered, to 0.49 between portfolios of five shares from the same industry and 0.66 between portfolios of ten. Between portfolios of 50 *randomly* selected shares with the same estimation period, the correlation coefficient is 0.85. A significance test which does not account for cross-correlation at any of these levels would be virtually useless. For example, Collins and Dent estimate that 'with a sample size of 40 firms exhibiting cross-correlation of 0.2 ... over repeated trials one would expect the true null of H_0: $\mu = 0$ to be rejected roughly 29% of the time, even though the nominal significance level is 0.05' (p. 54). The sample variance in this test is the simple average of the share residual variances, which ignores residual covariance (and is not one of the above methods).

However, in Brown and Warner's tests the performance of the share time series method is little different whether randomly selected shares in each portfolio of 50 have the same event dates or not. This method does not allow for cross-correlation but it is 'negligible for randomly selected securities' using an equally weighted index (Brown and Warner, 1980, p. 235). This is not so using a value weighted index, in which case 'simulations of clustering ... result in rejection rates under the null of about 15%' at the 0.05 level of significance (p. 235). The reason for this is that the sample portfolios, with randomly selected shares, are dominated by the returns of smaller company shares whereas a value weighted index reflects the returns on large companies. When small and large company returns differ, event dates are shared and a value weighted index is used, cross-correlation of errors can occur. Dyckman *et al.* report a small (10%) reduction in power of the time series method with clustering and note that this is inconsistent with Brown and Warner. But, surprisingly, they do not state whether they use a value or equally weighted index, so it is not possible to compare this result with Brown and Warner's.

Table 2 shows a comparison of the share and portfolio time series methods using results from both Brown and Warner's studies (with the equally weighted index). The share time series method is more powerful for daily data and of similar power for monthly. Though it assumes no cross-correlation, it weights the abnormal return of each share separately by its standard error whereas the

© Basil Blackwell 1995

portfolio time series method adjusts to cross-correlation but gives equal weight to each abnormal return. Apparently the weighting of abnormal returns yields better results even when event dates share calendar dates, for portfolios of shares from different industries. It can also be seen from Table 2 that, for a given level of abnormal return introduced on day 0 and month 0, the rejection frequencies using daily data are three to four times greater than those using monthly data.

The daily returns study found that the cross-sectional method is also less powerful than the share time series method unless error variance increases (specifically, is doubled) during the event period, in which case the share time series method rejects the null too often when it is true. This result is supported by Collins and Dent and Boehmer, Musumeci and Poulsen (1991); the latter find that error variance increases of only one-and-a-half times can still cause serious over-rejection of the null with the share time series method.

5.2. *Prediction error adjustment*

Equation (6) is a crude method of standardisation. It is more correct to treat event period errors explicitly as prediction errors, so long as the event period does not form part of the estimation period. Each share's prediction error is standardised to

Table 2. Summary of Brown and Warner's Results on Market Model Identification of Significant Abnormal Returns

Proportions of 250 portfolios, each of 50 shares, exhibiting significant abnormal returns at the 5% level (one tailed test).

Abnormal return on day or month 0	0%	0.5%	1.0%	2.0%
A. Share time series method:				
test statistic = $\sqrt{N}(\text{ASE}_t)$				
(i) Calendar dates differ				
Daily data rejection rate	6.4%	53.2%	97.6%	
Monthly data rejection rate	4.4%	—	22.8%	—
(ii) Event dates share calendar dates				
Daily data rejection rate	8.0%	61.2%	96.0%	—
Monthly rejection rate	4.0%	—	23.2%	—
B. Portfolio time series method:				
test statistic = $\bar{e}_t/s(\bar{e})$				
(i) Calendar dates differ				
Daily data				
rejection rate of null hypothesis	4.4%	27.2%	80.4%	99.6%
(no abnormal return)				
Monthly rejection rate not given				
(ii) Event dates share calendar dates				
Daily data rejection rate	8.0%	39.2%	84.4%	—
Monthly rejection rate	5.6%	—	24.8%	—

Source: Brown and Warner (1980, 1985)

© Basil Blackwell 1995

produce a standardised prediction error (SPE) as follows:

$$\mathrm{SPE}_{it} = e_{it}/\mathrm{S}_{it}, \tag{13}$$

where

$$\mathrm{S}_{it} = \mathrm{s}_i\sqrt{\left(1 + \frac{1}{\mathrm{T}_i} + \frac{(R_{mt} - \bar{R}_m)^2}{\Sigma(R_{m\tau} - \bar{R}_m)^2}\right)}. \tag{14}$$

Here τ is used to distinguish an observation in the estimation period, $\bar{R}_m$ is the mean market return in the estimation period and T_i is the number of observations in the estimation period for share i (they may differ from share to share). The standardisation used in the share time series method is a simpler version of this. Equation (14) is derived from expressions for the two sources of inaccuracy in market model predictions. First, the true regression line for the relationship between returns on the share and returns on the market is unknown and only estimated from the estimation period data. Second, actual returns may vary from expected returns given the true regression line. This source of error can be due to any of the reasons why changes in the independent variable, R_{mt}, do not 'explain changes in the dependent variable, R_{it}. The formal derivation can be found, for example, in Kmenta (1971, ch. 7).

Under the formula for S_{it}, the market model will be less accurate, and S_{it} larger,

(1) the larger the standard error of the regression, s_i;
(2) the shorter the estimation period, i.e. the smaller is T_i
(3) the smaller the dispersion of the explanatory variable, $R_{m\tau}$, during the estimation period;
(4) the larger the difference between the observed R_{mt} and its estimation period mean.

Intuitively, it is clear why (1) and (2) should apply. (3) applies because the less the explanatory variable varies in the sample, the smaller the range over which any relationship can be estimated. (4) means that predictions from the model are more reliable the closer is the observed R_{mt} to its mean, $\bar{R}_m$, from the estimation period, because it is around the mean that there is most evidence for the relationship. Hence, the further away from the mean, the less reliable the prediction and the larger an error for it to be significant.

As with the share time series method, SPEs are comparable in terms of significance of deviation from zero and the test statistic is

$$\sqrt{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \mathrm{ASPE}_{it}.$$

The difference is that S_{it} is a more accurate estimator of the standard deviation of PE_{it} than is s_i. Mikkelson and Partch (1986) use this methodology to examine the response to announcements of security issues.[7] Call it the *prediction error* method. Though it is now widely used, simulation evidence shows that the impact

© Basil Blackwell 1995

of the additional terms to s_i in equation (14) have a negligible impact on results (Brown and Warner, 1980 and 1985; Corrado, 1989; Corrado and Zivney, 1992).

To cope better with event-induced increases in error variance, Boehmer *et al.* (1991) propose what they call a *standardised cross-sectional* method. First SPEs are calculated, as above (presumably SEs would do as well), then the average event day SPE, $ASPE_t$, is divided by its cross-sectional standard deviation, $s(SPE_{it})\sqrt{N}$, where $s(SPE_{it})$ is the standard deviation of the SPEs, which will be greater than its expected value of one if error variance is higher in the event period than in the estimation period. This cures the problem of rejecting the null too often when it is true, with little loss of sensitivity to the presence of abnormal returns, giving better results than the cross-sectional method.

5.3. *Generalised least squares estimation*

The most complete statistical procedure is to use generalised least squares (GLS), in which each share's prediction error is standardised by dividing it by a factor which reflects both the variance of its market model residual and the covariance of that residual with the residuals of other shares. Following Collins and Dent (1984), we can start with a similar expression for standardisation to equation (14). Let this be C_{it}.

$$C_{it} = S_{it}/s_i = \sqrt{\left(1 + \frac{1}{T_i} + \frac{(R_{mt} - \bar{R}_m)^2}{\Sigma(R_{m\tau} - \bar{R}_m)^2}\right)}. \tag{15}$$

Instead of C_{it} being multiplied by s_i, it is multiplied by $\Sigma_{i=1}^{N} C_{it}s_{ij}$, where

$$s_{ij} = (1/T - 2) \sum_{\tau=1}^{T} e_{i\tau}e_{j\tau}, \tag{16}$$

the estimated covariance between residuals of shares i and j (variance when i = j) during period T. A variance-covariance matrix is estimated which has the following elements:

$$\begin{matrix} C_{it}C_{it}s_i^2 & C_{jt}C_{it}s_{ji} & \cdots & C_{Nt}C_{it}s_{Ni} \\ C_{it}C_{jt}s_{ij} & C_{jt}C_{jt}s_j^2 & \cdots & C_{Nt}C_{jt}s_{Nj} \\ \vdots & \vdots & \ddots & \vdots \\ C_{it}C_{Nt}s_{iN} & C_{jt}C_{Nt}s_{jN} & \cdots & C_{Nt}C_{Nt}s_N^2 \end{matrix}$$

The GLS procedure for estimating the average standardised prediction error for time t ($ASPE_t$) involves weighting each error by the sum of the inverse of its variance plus N−1 covariances and dividing the sum of these weighted errors by the sum of their weights. So $ASPE_t$ is a weighted average in which the higher is the variance of a share's residuals plus their covariances with the residuals of other shares, the less weight in the average its prediction error has. In the time

© Basil Blackwell 1995

series and prediction error methods, each share's prediction error is divided by s_i and S_{it} respectively, both estimates of its standard deviation. In the GLS method, each share's prediction error is divided by its variance plus N−1 covariance terms, which therefore dominate when covariance is non-negligible. So the more cross-correlation there is, the larger (unweighted) prediction errors have to be for $ASPE_t$ to be significant.

Collins and Dent also adjust for any change in the variances of errors generally between estimation period and test period by multiplying the GLS estimate of the variance of $ASPE_t$ by a factor f_t^2 which is the cross-sectional variance of errors for time t divided by the GLS variance of estimation period residuals. A value of one for f_t^2 indicates no difference in variance; a value greater than one indicates higher test period variance. Call the last two paragraphs the *GLS* method.[8]

We have now considered six methods of testing the significance of abnormal returns and are in a position to summarise, in Table 3, how each is affected by the variances of errors differing from share to share, cross-correlation of errors and shift in variances between estimation and event periods. Only the performance of the GLS method can be expected to be sustained in the presence of any or all these characteristics; the other five cope with one or two each. But estimation of the variance-covariance matrix in the GLS method requires that there be more observations for each share than there are shares in the sample. This can be a problem in some studies, and if n explanatory variables are used rather than the

Table 3. Significance Testing Methods in Relation to Characteristics of Market Model Errors

The table shows which characteristic each method can accommodate. 'No' means that performance may be diminished should it obtain. Either the null hypothesis is rejected too often for a given significance level (type 1 error) or the test is less powerful than the best alternative (type 2 error).

	Characteristic		
Method	Variances of errors differ	Errors are cross-correlated	Shift in variance
Share time series	Yes	No (type 1)	No (type 1)
Portfolio time series	No (type 2)	Yes	No (type 1)
Cross-sectional	No (type 2)	No (type 1)	Yes
Prediction error	Yes	No (type 1)	No (type 1)
Standardised cross-sectional	Yes	No (type 1)	Yes
GLS	Yes	Yes	Yes

© Basil Blackwell 1995

one in the market model, observations for each share numbering at least n times the number of shares are required. Furthermore, to obtain accurate estimates which are not biased downwards, many times more than these minima are required (see Bernard, 1987). The author is not aware of a study which uses this method.

Collins and Dent's results confirm what would be expected a priori from the design of each method, as shown in Table 3.

(1) In the absence of cross-correlation and with equal and constant error variances between estimation and test periods, all the tests perform similarly.
(2) With cross-correlation of 0.2 or more only the GLS and portfolio time series methods, which account for it, are satisfactory. The prediction error and cross-sectional methods underestimate the variance of average prediction errors, causing the null hypothesis to be rejected too often.
(3) With an increase in the variance of errors between estimation and test periods, only the GLS method does not reject the null too often, and the cross-sectional method if there is no cross-correlation.
(4) With unequal error variances from share to share and no cross-correlation, the GLS and prediction error methods are of equal power and superior to the cross-sectional and portfolio time series methods. With cross-correlation as well, only the GLS and portfolio time series methods do not reject the null too often, and the latter is less powerful.

It should be remembered that the focus of this study is on the impact of cross-correlation. Brown and Warner's evidence is that if shares are randomly chosen across industries, the impact is very small even when event dates are shared. Collins and Dent do not test the share time series or standardised cross-sectional methods and Brown and Warner do not test the GLS method.

5.4. *Non-parametric tests*

All the methods we have considered so far apply the *t*-test to some measure of abnormal return. Brown and Warner (1980 and 1985) report that the sign and Wilcoxon tests are not as well specified as the *t*-test, but many event studies use such non-parametric tests, or at least note the proportion of positive errors, in addition to whatever parametric testing procedure is used. Recently researchers have proposed refined non-parametric tests and provided evidence that they perform better than *t*-tests.

The conventional sign and Wilcoxon tests assume that market model errors in the absence of abnormal returns are distributed symmetrically around a mean of zero, so that they also have a median of zero. In fact market model errors are skewed slightly to the right, so the median is less than zero, meaning that the non-parametric test statistics will tend to be negative even when the mean error is zero. This causes under-rejection of the null hypothesis (Brown and Warner, 1980

© Basil Blackwell 1995

and 1985; Berry, Gallinger and Henderson, 1990). Corrado and Zivney (1992) present a version of the sign test in which the sign allocated to an abnormal return is determined by its difference from the share's time series median abnormal return, rather than from zero. Simulations suggest that this revised sign test is correctly specified and of similar power to the *t*-test using the share time series method.

However, it is Corrado's (1989) rank test which may prove to be a significant advance in event study methodology since simulation evidence so far shows it to be more powerful than any alternative. The test is simple and does not assume that abnormal returns are distributed symmetrically around the mean. Estimation and event period errors are ranked for each share and the average rank of all errors is subtracted from the rank of the event day error. So positive abnormal return on the event day tends to be reflected in a higher than average rank for that day's error, producing a positive average difference across all shares for that day. The test statistic is formed by dividing this average difference by the standard deviation of average differences over the estimation and event periods.

Let the average difference for day t be AD_t.

$$AD_t = 1/N \sum_{i=1}^{N} (K_{it} - [\{T+1\}/2]), \tag{17}$$

where K_{it} is the rank of the market model error for share i on day t, N is the number of shares in the sample and T is the number of days in the estimation and event periods.

The test statistic is

$$AD_t/s(AD),$$

where

$$s(AD) = \sqrt{\left([1/T] \sum_{t=1}^{T} [AD_t]^2\right)}. \tag{18}$$

If some of the shares have missing returns, this can be allowed for by dividing K_{it} by one plus the number of non-missing returns for share i; the average rank for each share's errors is then one half instead of $(T+1)/2$. Corrado (1989), Corrado and Zivney (1992) and Maynes and Rumsey (1993) find by simulation under a variety of circumstances that the rank test is more powerful than the (adjusted) sign test and the *t*-test using the share or portfolio time series methods, though the improvements are typically marginal (a few percentage points).[9] Corrado argues that the superior specification of the rank test is a consequence of non-normality in the distribution of daily share returns and market model errors. It will be interesting to see if further research confirms the superiority of the rank test and if it starts to be used in real event studies.

© Basil Blackwell 1995

6. Further comment and related issues

6.1. *Cross-correlation*

The problem of cross-correlation is extensively discussed by Bernard (1987), who analyses the underestimation of residual variance by the cross-sectional method for the market model and for 'cross-sectional returns studies' in which the model is of the form

$$e_{it} = \alpha_t + \beta_1 X_{it} + \beta_2 Y_{it} + \cdots + \varepsilon_{it}, \tag{19}$$

where X_{it}, Y_{it} are firm-specific explanatory variables. He confirms the context in which cross-correlation is likely to be sufficient to cause serious underestimation, i.e. when event dates are shared and shares are drawn from the same industry. He also adds a further factor: the observation interval of the data (i.e. days, months etc). As the interval increases, so does cross-correlation. For a large sample of individual shares, 'the mean intra-industry cross-sectional correlation is only 0.04 when daily data are used; the mean rises to 0.09 in weekly data, 0.18 in monthly data, 0.24 in quarterly data and 0.30 in annual data' (p. 10). The sample size ranges from 1,080 for daily data to 274 for annual and an industry is again defined as a three digit SIC category. Bernard believes that small departures from weak-form efficiency are responsible for this phenomenon. However, 'the degree of inter-industry cross-sectional correlations is small relative to intraindustry correlations', ranging from 0.01 for daily data to 0.06 for annual (p. 11).

6.2. *Beta estimates*

6.2.1. *Thin trading*

Researchers have been concerned about the effect on estimates of beta of thin trading — shares trading infrequently during the day or not at all on some days, so that share and market return intervals are non-synchronous. Brown and Warner (1980, and 1985) test both the Scholes and Williams (1977) and Dimson (1979) adjustments, but they 'convey no clear-cut benefit in an event study' (1985, p. 18), even for a sample of American Stock Exchange (AMEX) (smaller company) shares. This result is supported by Dyckman *et al.* (1984) and Thompson (1988). However, Brown and Warner (1985) report that it is around twice as difficult to identify abnormal returns of AMEX shares because residual standard deviations are higher.

McInish and Wood (1986) directly test the effect on beta estimates of five adjustments for thin trading by constructing portfolios with the same return, and therefore the same expected beta, but with shares grouped according to trading frequency, from least to most. Six months of daily data for NYSE shares are used to calculate mean OLS and adjusted betas of the shares in the portfolios. The OLS mean betas range from 0.744 for the most thinly traded to 1.494 for the most frequently, a difference of 0.75. The most effective adjustment, Dimson's, reduces this difference to 0.54.

© Basil Blackwell 1995

6.2.2. *Beta stability*

Event period betas may differ from estimation period betas and various adjustments have been proposed. For example, Kalay and Loewenstein (1985) present evidence that betas and the variance of share prices are both higher on the day of a dividend announcement and the day following than on other days. They argue that, as dividend announcements are predictable events, the higher event period betas should be used in estimating abnormal returns.

It is known that betas tend to regress towards their mean of one. Making adjustments for regression towards their mean reduces the dispersion of betas and can significantly alter results compared with non-adjustment. For example, Krueger and Johnson (1991), using 60 months' data to estimate historic betas and a simple adjustment for their regression tendency, state that their 'results indicate that anomaly studies are sensitive to the characterisation of beta' (p. 578). The error in estimating betas due to this phenomenon can be minimised by using a split estimation period before and after the event period and is also smaller the more shares in the portfolio and the longer the return interval (Foster, Hansen and Vickrey, 1988).

Brown, Lockwood and Lummer (1985) present a statistic for testing for a significant change in market model error variances and coefficients between estimation period and event period. They note that the proportion of a sample of shares exhibiting a significant change depends positively on the length of event period chosen, and develop a switching regression technique to select optimal event periods on a share-by-share basis. This technique assumes that the estimation and event periods constitute two different 'regression regimes' and selects each share's event period to maximise the likelihood of observations being allocated to the appropriate regime. Examining abnormal returns over many months around a sample of 373 stock splits, in the manner of Fama *et al.* (1969), they conclude that optimal event periods 'are dispersed to the extent that the imposition of a universal interval would indeed be inappropriate' (pp. 325–6). Whether it is worth applying this technique clearly depends on how long an event period is deemed necessary and how much the variance of errors increases around the event; stock splits may be an extreme case. (Fama *et al.* note that the cross-sectional mean deviation of errors is much higher than usual around the month of the split and is 'well over twice as large as the corresponding average residuals'.)

6.3. *Bid–ask spread*

Prices used to calculate share returns are closing transactions prices which may be at market makers' bid or ask quotations or somewhere in the middle (see Lease, Masulis and Page, 1991, p. 1530). For a sample of shares, the average closing prices should be approximately the average of the midpoints between the bid-ask spreads. However, some events may cause a predominance of closing prices at the bid or ask which will affect event returns but does not represent a genuine price response. For example, Lease *et al.* document a negative average return on

© Basil Blackwell 1995

offering dates of seasoned share issues in the USA, as do others. They argue that this is a result of temporary order flow disruptions because 'for a short period of time both a primary market and a secondary market are open concurrently. As a result, many purchase orders that normally would be routed to the secondary market are channelled to the primary market where offer price discounts and exemption from brokerage fees prevail. In contrast, sell orders continue to be directed exclusively to the secondary market...' (p. 1524). They suggest using midpoint bid–ask prices and a day longer event window to remove the effect of this temporary disruption.

A different type of problem arises from many closing prices being at the bid or the ask and hence not the 'true' price. This causes noise in the series of prices which biases single period returns upwards (see Blume and Stambaugh, 1983). The bid–ask spread, and so the upward bias, are larger relative to price for low-priced shares, and this can significantly affect the results of certain abnormal return studies. For example, Conrad and Kaul (1993) present evidence that much of the abnormal returns from DeBondt and Thaler's (1985) strategy of buying shares which have performed badly ('losers') and selling those which have performed well ('winners') is due to estimating three year returns by adding 36 single period monthly returns, because the 'losers' have lower prices than the 'winners', so that the three year return incorporates 36 upward biases of losers relative to winners. Using a holding period return produces different and more accurate results.

7. Conclusions and recommendations

Event study methods are worth reviewing because of their many variations and their very wide application in empirical research. They continue to be refined, and testing via simulation experiments has become a minor industry since 1980. It may be helpful in conclusion to offer recommendations on the main choices facing the researcher who wishes to carry out an event study.

7.1. *Choice of model*

The market model is now much the most commonly used model of expected returns in event studies and is the best supported by the evidence. The CAPM is also used in tests of investment rules and fund performance, which are not strictly event studies. As the CAPM and market model can produce different results, both should be used in such tests.

Points to note:
The shorter the interval over which each return observation is measured (daily, weekly etc.), the easier it is to identify abnormal returns. So the event date should be identified as precisely as possible and daily data used if possible. Estimation periods of 100 days or more are sufficient for accurate estimation of α and β, though 200–300 days is more common.

© Basil Blackwell 1995

Thin trading does cause bias in beta estimates but adjustments for thin trading make little difference to event study results. Other adjustments to betas have been proposed, but not generally adopted.
Shares with contaminating news in the event period, distinct from news of the event itself, are normally removed from samples.
Researchers should be aware of the possible impact on results of the size effect and other 'efficient markets anomalies', and also of the bid–ask spread.

7.2. *Choice of significance test*

This depends on the characteristics of market model errors. We can consider three scenarios.

(1) Market model errors are not cross-correlated and there is little or no event period increase in error variance.
Use the share time series method with a *t*-test, or use Corrado's rank test. There is no evidence that more complicated prediction error or GLS estimation increases the power of tests, though the prediction error method is quite widely used.

(2) Errors are not cross-correlated but there is an appreciable increase in event period variance (one-and-a-half times or more).
Use the cross-sectional method or, better, Boehmer *et al.*'s standardised cross-sectional method or the rank test. The unadjusted share time series method will be mis-specified (because it finds abnormal returns too often where none exist).

(3) Errors are cross-correlated.
This will not occur unless events share the same calendar date and shares are from the same industry. Use the portfolio time series method.

Acknowledgements

I wish to thank Andrew McCosh, Peter Moles, Journal editors Stuart Sayer and Les Oxley and three anonymous referees for their helpful comments on earlier drafts of this paper and Jake Ansell and, especially, John Banasik for advice on statistical points.

Notes

1. The reason why Dimson and Marsh reject the market model is interesting, though not relevant to most other studies. 'For stocks which performed well (poorly) during the estimation period, a_j will be projected to be positive (negative) and since favourable (unfavourable) performance does not persist in weak-form efficient markets ..., the e_{jt} can be predicted to be negative (positive). The market model would thus provide a quite inappropriate benchmark, since analysts who simply predict trend reversals will always appear to have positive forecasting skills' (p. 1264). Jensen (1969) and Brown and Warner (1985) do indeed find small negative first order autocorrelation of market model residuals of −0.063 (annual data) and −0.071 (daily data). In a comprehensive

© Basil Blackwell 1995

test of efficient markets hypothesis anomalies using monthly data, Jacobs and Levy (1988) found that 'residual reversal' was statistically very significant and 'by far the most powerful effect' (p. 30).

2. Jensen (1968) measures the performance of 115 managed funds. The following regression model is estimated for each fund j using as many years t of annual data as are available for each fund from 1945 to 1964:

$$R_{jt} - R_{ft} = \alpha_j + \beta_j(R_{mt} - R_{ft}) + e_{jt}. \tag{20}$$

Abnormal performance is measured by α_j and is tested for significance by the t test; α_j is Jensen's performance index.

3. The explanation may be that Brenner uses a different version of the Fama–MacBeth model.

4. Jaffe (1974) measures the abnormal returns earned by company insiders on their share trades. Normal returns are estimated using the Fama–MacBeth model, equation (5). The α values are taken from Fama and MacBeth (1973). They divide shares on the New York Stock Exchange into 20 portfolios, ranked by beta estimated over a four year period using monthly returns data. Portfolio betas are calculated by taking the simple average of the betas of constituent shares, estimated over the next five years. This procedure is to eliminate the following bias: 'In a cross section of $\hat{\beta}_i$, high observed $\hat{\beta}_i$ tend to be above the corresponding true β_i and low observed $\hat{\beta}_i$ tend to be below the true β_i. Forming portfolios on the basis of ranked $\hat{\beta}_i$ thus causes bunching of positive and negative sampling errors within portfolios' (p. 615). Since the rankings are determined by the first estimation but the betas used by the second, the sampling errors in the second do not affect the rankings. For each month after the second, five year estimation period, a cross-sectional regression is carried out:

$$R_{pt} = \hat{\alpha}_{0t} + \hat{\alpha}_{1t}\hat{\beta}_{p,t-1} + \varepsilon_{pt}, \tag{21}$$

where R_{pt} is there turn on portfolio p in month t and $\hat{\beta}_{p,t-1}$ is the portfolio β for the previous month which may change due to delistings. Also βs for individual shares are 'updated' each year after the end of the estimation period by adding that year's data to the data used to estimate βs. Each of the testing periods lasts four years. So the α values in Jaffe's model can vary month by month. He constructs portfolios of shares which have been traded by insiders during a certain period (e.g. May–June 1962) and tests whether observed R_{pt} are significantly different from expected R_{pt} from the model for months during and surrounding this event period. The prediction errors (e_p) of each insider trade portfolio are standardised by dividing them by the standard deviation (SD_p) of the monthly errors for that portfolio over 50 months up to and including the last month of the event period:

$$SPE_{pt} = e_{pt}/SD_p \tag{22}$$

where SPE_{pt} is the standardised prediction error for portfolio p at month t. The test statistic for the average standardised prediction error is

$$ASPE_t/(s/\sqrt{n}) = \sqrt{n}(ASPE_t) \tag{23}$$

because s equals one; n is the number of portfolios in the sample.

5. Imhoff and Lobo (1984) investigate whether abnormal returns on shares occur during months when analysts revise their forecasts pertaining to those shares. Market model coefficients are estimated using monthly returns for 84 months preceding each forecast revision. For each month t, average abnormal returns ($\bar{w}_t$) are subtracted from abnormal returns on each share (w_{it}) to ensure zero average cross-sectional abnormal returns for the sample in that month; $w_{it} = R_{it} - (\alpha_i + \beta_i R_{mt})$. Each share's measure of abnormal returns, $w_{it} - \bar{w}_t$, is standardised by dividing it by its

© Basil Blackwell 1995

standard deviation,

$$s_t(w) = \sqrt{\left\{[(1/n_t - 1)] \sum_{i=1}^{n_t} (w_{it} - \bar{w}_t)^2\right\}}. \tag{24}$$

Shares are grouped into upward and downward revision samples, the mean standardised abnormal returns for which are tested for significance.

6. Salinger (1992) notes that serial correlation of estimated errors can cause serious underestimation of the variance of cumulative errors if the event period approaches the length of the estimation period, resulting in overstatement of the significance of cumulative errors. He presents an alternative formula for estimating their variance. Brown and Warner (1985) report that mean market model errors exhibit negative serial correlation, but that 'the benefits from autocorrelation adjustments appear to be limited' (p. 20) for an event period of eleven days.
7. The first use of the prediction error method in finance research is credited to Patell (1976). Mikkelson and Partch's estimation period of 140 trading days begins 21 trading days after the announcement date 'because many types of security offerings follow a period of statistically significant abnormal returns' (p. 40). Data from this period may therefore produce biased estimates of α and β. For each share they calculate daily prediction errors over an event period that begins 60 trading days before the announcement and ends 20 trading days after issuance or cancellation. Each prediction error is standardised by dividing it by the expression in equation (14) to produce an SPE. The test statistic for each event day t is $\sqrt{N}(\text{ASPE}_t)$, as in equations (8) and (23).
8. Thompson (1985) argues that the natural way of measuring event period abnormal returns for a sample of shares from the perspective of 'classical econometric theory' is to estimate the set of equations of the following type using joint generalised least squares:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \delta_{it}\gamma_i + e_{it}, \tag{25}$$

where δ_{it} is a dummy variable which equals 1 if time t is an event day or month for share i, and 0 otherwise; γ_i is the estimated abnormal return for each event day/month. This is a set of seemingly unrelated regressions, one for each share in the sample. Thompson comments that 'parameter estimation security by security is equivalent to OLS estimation of the entire system. In comparison to GLS with Σ [the variance–covariance matrix] known, OLS is less efficient. But when Σ is estimated, the improvement of estimated GLS is less obvious and depends on the tradeoff between estimation errors and the true departure from the OLS assumption' (p. 161). In fact Malatesta (1986), using simulation, finds GLS to be no more powerful than the share time series method or OLS estimation of γ_i, for samples of randomly selected shares and event dates.
9. For the case of an event period increase in variance, Corrado and Zivney (1992) compare the share time series method, Boehmer *et al.*'s standardised cross-sectional method and the rank test with and without a cross-sectional variance adjustment. The rank test with a cross-sectional variance adjustment is derived as follows. For the estimation period, market model errors for each share are standardised, as in equation (6), by their time series standard error, giving a series of SE_{it}s. For the event day, the SE_{it} is divided by the cross-sectional standard deviation of SE_{it} for that day, $s(SE_{it})$. The ranks used in the rank test are then derived from the estimation period series SE_{it}s together with $SE_{it}/s(SE_{it})$ for the event day. Corrado and Zivney find that 'when an event date variance increase is likely, correct specification of the *t*-test requires that a cross-sectional variance adjustment be implemented. For the rank test, in contrast, a variance adjustment appears to be unimportant in tests for positive

© Basil Blackwell 1995

abnormal performance, but necessary in tests for negative abnormal performance' (p. 477).

References

Ball, R. and Brown, P. (1968) An empirical evaluation of accounting income numbers, *Journal of Accounting Research*, 6, 159–178.

Banz, R. W. (1981) The relationship between return and market value of common stocks, *Journal of Financial Economics*, 9, 3–18.

Beaver, W. H. (1981) Econometric properties of alternative security return methods, *Journal of Accounting Research*, 19, 163–184.

—— (1982) Discussion of market-based empirical research in accounting: a review, interpretation, and extension, *Journal of Accounting Research* (Supplement), 20, 323–331.

Bernard, V. L. (1987) Cross-sectional dependence and problems in inference in market-based accounting research, *Journal of Accounting Research*, 25, 1–48.

Berry, M. A., Gallinger, G. W. and Henderson, G. V. (1990) Using daily stock returns in event studies and the choice of parametric versus nonparametric test statistics. *Quarterly Journal of Business and Economics*, 29, 70–85.

Blume, M. and Stambaugh, R. F. (1983) Biases in computed returns: an application to the size effect, *Journal of Financial Economics*, 12, 387–404.

Boehmer, E., Musumeci, J. and Poulsen, A. B. (1991) Event study methodology under conditions of event induced variance, *Journal of Financial Economics*, 30, 253–272.

Brenner, M. (1979) The sensitivity of the efficient market hypothesis to alternative specifications of the market model, *Journal of Finance*, 34, 915–929.

Brick, I. E., Statman, M. and Weaver, D. G. (1989) Event studies and model misspecification: another look at the benefits to outsiders from public information about insider trading, *Journal of Business Finance and Accounting*, 16, 399–424.

Brown, K. C., Lockwood, L. J. and Lummer, S. L. (1985) An examination of event dependency and structural change in security pricing models, *Journal of Financial and Quantitative Analysis*, 20, 315–334.

Brown, S. J. and Warner, J. B. (1980) Measuring security price performance, *Journal of Financial Economics*, 8, 205–258.

—— (1985) Using daily stock returns: the case of event studies, *Journal of Financial Economics*, 14, 3–31.

Chopra, N., Lakonishok, J. and Ritter, J. R. (1992) Measuring abnormal performance: do stocks overreact? *Journal of Financial Economics*, 31, 235–268.

Collins, D. W. and Dent, W. T. (1984) A comparison of alternative testing methodologies used in capital market research, *Journal of Accounting Research*, 22, 48–84.

Conrad, J. and Kaul, G. (1993) Long term market overreaction or biases in computed returns? *Journal of Finance*, 48, 39–63.

Corrado, C. J. (1989) A nonparametric test for abnormal security-price performance in event studies, *Journal of Financial Economics*, 23, 385–395.

—— and Zivney, T. L. (1992) The specification and power of the sign test in event study hypothesis tests using daily stock returns, *Journal of Financial and Quantitative Analysis*, 27, 465–478.

DeBondt, W. F. M. and Thaler, R. M. (1985) Does the stock market overreact? *Journal of Finance*, 40, 793–805.

Dimson, E. (1979) Risk measurement when shares are subject to thin trading, *Journal of Financial Economics*, 7, 197–226.

—— and Marsh, P. (1984) An analysis of brokers' and analysts' unpublished forecasts of UK stock returns, *Journal of Finance*, 39, 1257–1292.

© Basil Blackwell 1995

—— and Marsh, P. (1986) Event study methodologies and the size effect: the case of UK press recommendations, *Journal of Financial Economics*, 17, 113–143.
Dyckman, T., Philbrick, D. and Stephan J. (1984) A comparison of event study methodologies using daily stock returns: a simulation approach, *Journal of Accounting Research* (Supplement), 22, 1–33.
Fama, E. F. (1991) Efficient capital markets: II, *Journal of Finance*, 46, 1574–1614.
——, Fisher, L., Jensen, M. C. and Roll, R. (1969) The adjustment of stock prices to new information, *International Economic Review*, 10, 1–21.
—— and French, K. R. (1991) The cross-section of expected stock returns, *Journal of Finance*, 47, 427–465.
—— and MacBeth, J. D. (1973) Risk, return, and equilibrium: some empirical tests, *Journal of Political Economy*, 81, 607–636.
Foster, T., Hansen, D. and Vickrey, D. (1988) Additional evidence on the abatement of errors in predicting betas through increases in portfolio size and on the regression tendency, *Journal of Business Finance & Accounting*, 15, 185–199.
Glascock, J. L., Henderson, G. V., Officer, D. T. and Shah, V. (1991) Examining the sensitivity of the standardised cumulative prediction error as an event study test statistic, *Journal of Economics and Business*, 43, 49–57.
Imhoff, E. A. and Lobo, G. J. (1984) Information content of analysts' composite forecast revisions, *Journal of Accounting Research*, 22, 541–554.
Jacobs, B. I. and Levy, K. N. (1988) Disentangling equity return regularities: new insights and investment opportunities, *Financial Analysts Journal*, 44, 18–43.
Jaffe, J. F. (1974) Special information and insider trading, *Journal of Business*, 47, 410–428.
Jensen, M. C. (1968) The performance of mutual funds in the period 1945–64, *Journal of Finance*, 23, 389–416.
—— (1969) Risk, the pricing of capital assets and the evaluation of investment portfolios, *Journal of Business*, 42, 167–247.
Kalay, A. and Loewenstein, U. (1985) Predictable events and excess returns: the case of dividend announcements, *Journal of Financial Economics*, 14, 423–449.
Klein, A. and Rosenfeld, J. (1987) The influence of market conditions on event-study residuals, *Journal of Financial and Quantitative Analysis*, 22, 345–351.
Kmenta, J. (1971) *Elements of Econometrics*, New York: Macmillan.
Krueger, T. M. and Johnson, K. H. (1991) Parameter specifications that make little difference in anomaly studies, *Journal of Business Finance and Accounting*, 18, 567–582.
Lakonishok, J. and Vermaelen, T. (1990) Anomalous price behaviour around repurchase tender offers, *Journal of Finance*, 45, 455–477.
Lease, R. C., Masulis, R. W. and Page, J. R. (1991) An investigation of market microstructure impacts on event study returns, *Journal of Finance*, 46, 1523–1536.
Malatesta, P. H. (1986) Measuring abnormal performance: the event parameter approach using joint generalised least squares, *Journal of Financial and Quantitative Analysis*, 21, 27–38.
Masulis, R. W. (1980) The effects of capital structure change on security prices, *Journal of Financial Economics*, 8, 139–177.
Maynes, E. and Rumsey, J. (1993) Conducting event studies with thinly traded stocks, *Journal of Banking and Finance*, 17, 145–157.
McInish, T. H. and Wood, R. A. (1986) Adjusting for beta bias: an assessment of alternative techniques: a note, *Journal of Finance*, 41, 277–286.
Mikkelson, W. H. and Partch, M. M. (1986) Valuation effects of security offerings and the issuance process, *Journal of Financial Economics*, 15, 31–60.
Patell, J. M. (1976) Corporate forecasts of earnings per share and stock price behaviour: empirical tests, *Journal of Accounting Research*, 14, 246–276.

© Basil Blackwell 1995

Peterson, P. P. (1989) Event studies: a review of issues and methodology, *Quarterly Journal of Business and Economics*, 28, 36–66.

Reinganum, M. R. (1981) Misspecification of capital asset pricing: empirical anomalies based on earnings' yields and market values, *Journal of Financial Economics*, 9, 19–46.

Salinger, M. (1992) Standard errors in event studies, *Journal of Financial and Quantitative Analysis*, 27, 39–53.

Scholes, M. and Williams, J. (1977) Estimating betas from non-synchronous data, *Journal of Financial Economics*, 5, 309–328.

Seyhun, H. N. (1986) Insiders' profits, costs of trading, and market efficiency, *Journal of Financial Economics*, 16, 189–212.

Sharpe, W. F. and Alexander, G. J. (1990) *Investments* (4th edn.), Englewood Cliffs: Prentice-Hall.

Thompson, J. E. (1988) More methods that make little difference in event studies, *Journal of Business Finance and Accounting*, 15, 77–86.

Thompson, R. (1985) Conditioning the return-generating process on firm- specific events: a discussion of event study methods, *Journal of Financial and Quantitative Analysis*, 20, 151–168.

Thompson, R. B., Olsen, C. and Dietrich, J. R. (1987) Attributes of news about firms: an analysis of firm-specific news reported in the Wall Street Journal Index, *Journal of Accounting Research*, 25, 245–273.

Wright, C. J. and Groff, J. E. (1986) Uses of indexes and data bases for information release analysis, *The Accounting Review*, 41, 91–100.

© Basil Blackwell 1995