EVENT STUDY TESTS: A BRIEF SURVEY

Ana Paula Serra
Faculdade de Economia do Porto/CEMPRE

ABSTRACT

In this paper, I describe some of the main parametric and non-parametric tests used in event studies to assess the significance of abnormal returns or changes in variance of returns.

Key-words: Event Studies

1. INTRODUCTION

In this paper, I describe some of the main event study tests used in empirical finance. This review is not intended to comprehend all the extensive literature on event studies. The parametric tests here described are Brown and Warner (1980, 1985), with and without crude independence adjustment, Patell’s (1976) standardised residual test and Boehmer, Musumeci and Poulsen’s (1991) standardised cross-sectional test. The non-parametric tests are the generalised sign test (Sanger and McConnell, 1986 and Cowen and Sergeant, 1996), the Wilcoxon signed rank test and Corrado’s (1989) rank test. In addition I outline briefly a bootstrapping procedure (please refer to Noreen, 1989). I also describe four tests to evaluate the significance of changes in variance. I present two parametric tests (an F test for the equality of two population variances and the Beavers’s/ May’s U test) and two non-parametric tests (the squared rank test and another rank test proposed by Rohrbach and Chandra, 1989). Serra (1999) illustrates these tests to examine the impact on returns of dual listing by emerging markets firms.

Event studies start with hypothesis about how a particular event affects the value of a firm. The hypothesis that the value of the company has changed will be translated in the stock showing an abnormal return. Coupled with the notion that the information is readily impounded into prices, the concept of abnormal returns (or performance) is the central key of event study methods. How does a particular event affect the value of a company? We must be careful because at any time we observe a mixture of market wide factors and a bunch of other firm events. To correctly measure the impact of a particular event we need to control for those unrelated factors. The selection of the benchmark to use or the model to measure normal returns is therefore central to conduct an event study.

The empirical model can be stated as follows: when an event occurs, market participants revise their beliefs causing a shift in the firm’s return generating process. For a given security, in non event periods,

\[ R_t = \beta_1 + e_t \]  

while in event periods

\[ R_t = \beta_1 + \beta_2 F + e_t \]

\( R_t \) is the return of the security in period \( t \); \( \beta_1 \) is a vector of independent variables (for example the return of the market portfolio) in period \( t \); \( B \) is a vector of parameters, such as the security beta; \( F \) is a row vector of firm characteristics influencing the impact of the event on the return process. \( G \) is a vector of parameters measuring the influence of \( F \) on the impact of the event and \( e_t \) is a mean zero disturbance term possibly differing in event and non event periods.

Hypotheses usually center on the parameters that measure the influence of the event \( (G) \) and most of the times \( F \) is set to unity. The null hypothesis is that such an event has no impact on the return generating process.

Event study methods are the econometric techniques used to estimate and draw inferences about the impact of an event in a particular period or over several periods.
The most common approach involves three steps: (1) Compute the parameters in the estimation period; (2) Compute the forecast errors (and obtain variance/covariance information) for a period or over an event window; aggregate across firms and infer about the average effect; (3) regress cross-sectionally abnormal returns on relevant features of the stock supposed to influence the impact of the event.

2. ABNORMAL RETURNS

Ex-post abnormal returns are obtained as the difference between observed returns of firm \( i \) at event week \( t \) and the expected return generated by a particular benchmark model:

\[
AR_i^t = R_i^t - E(R_i^t)
\]  

(3)

By averaging these residuals across firms in common event time, we obtain the average residuals \( (AR_0) \),

\[
AR_0 = \left( \frac{1}{N} \right) \sum_{i=1}^{N} AR_{i0}
\]  

(4)

where \( N \) is the number of firms in the sample and 0 refers to period 0 in event time. By cumulating the periodic average residuals over a particular time interval \( (L \) weeks around the listing date), we obtain the cumulative average residuals \( (CAR) \):

\[
CAR = \sum_{l=1}^{L} AR_l
\]  

(5)

3. THE SIGNIFICANCE OF ABNORMAL RETURNS

3.1. Parametric Tests

The parametric tests proposed in the literature rely on the important assumption that individual firm’s abnormal returns are normally distributed. The standard statistic is:

\[
t = \frac{AR_0}{S(AR_0)}
\]  

(6)

where \( AR_0 \) is defined as above and \( S(AR_0) \) is an estimate of standard deviation of the average abnormal returns \( \sigma(AR_0) \).

Cross-Sectional Independence

Considering cross-sectional independence, i.e., that the residuals are not correlated across securities,

\[
\sigma^2(AR_0) = \sigma^2 \left( \frac{1}{N} \sum_{i=1}^{N} AR_{i0} / N \right) = \left( \frac{1}{N^2} \right) \sum_{i=1}^{N} \sigma^2(AR_{i0})
\]  

(7)

The standard deviation of the average abnormal return for each security, \( \sigma(AR_{i0}) \), is then estimated on the basis of the standard deviation of the time series of abnormal returns of each firm during the estimation period \( (T \) weeks), as follows:

\[
S(AR_t) = \left[ \frac{\sum_{t=1}^{T} \left( AR_{it} - \frac{\sum_{i=1}^{T} AR_{it}}{T} \right)^2}{T-d} \right]^{1/2}
\]  

(8)
Under the null hypothesis of no abnormal performance, the statistic above (6) is distributed as Student-t with $T-d$ degrees of freedom$^8$.

**Cross-Sectional Dependence (Crude Adjustment)**

To account for the dependence across firms’ average residuals, in event time, Brown and Warner (1980) suggest that the standard deviation of average residuals should be estimated from the time series of the average abnormal returns over the estimation period.

$$ S^* (AR_0) = \sqrt{\frac{\sum_{i=1}^{T} \left( \frac{\sum_{t=1}^{T} AR_i}{N} - \bar{AR}^* \right)^2}{T-d}} \quad (9) $$

where

$$ \bar{AR}^* = \frac{\sum_{i=1}^{N} \left( \sum_{t=1}^{T} AR_i \right)}{N} \quad \text{and} \quad \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} AR_i}{T} $$

$$ (10) $$

Under the null hypothesis of no abnormal performance, the statistic above (6) is distributed as Student-t with $T-d$ degrees of freedom.

**Standardised Abnormal Returns**

The purpose of standardization is to ensure that each abnormal return will have the same variance$^5$. By dividing each firm’s abnormal residual by its standard deviation (obtained over the estimation period), each residual has an estimated variance of 1$^6$. The standardised residuals are given by:

$$ AR'_{10} = \frac{AR_{10}}{S(AR_i)} \quad (11) $$

In a particular event week, the test statistic of the hypothesis that the average standardised residuals across firms is equal to zero, is computed as:

$$ z = \frac{\bar{AR}'_{01} - 0}{S(AR_0')} = \frac{(1/N) \sum_{i=1}^{N} AR'_{10}}{S(AR_0')} \quad (12) $$

Considering independence across firms and that $AR_i$ are i.i.d., $AR_i'$ are assumed distributed as unit normal and the standard error of the average standardised residuals is given by:$^7$

$$ S(AR_0') = \frac{1}{\sqrt{N}} \quad (13) $$

By the Central Limit Theorem, the statistic in (12) is distributed unit normal for large $N$.

**Changes in Variance**

If the variance of stock returns increases on the event date, the above tests, using the time series of non-event period data to estimate the variance of the average abnormal returns, reject the null hypothesis too often. Boehmer et al. (1991) propose that the variance of average abnormal returns is estimated from the cross-section of event date (instead of estimation period) prediction errors. For the case of standardised residuals:

$$ S^2 (AR'_{10}) = \frac{1}{N(N-d)} \sum_{i=1}^{N} \left( \frac{AR'_{10} \sum_{t=1}^{T} AR_i}{N} \right)^2 \quad (14) $$
The procedure requires the assumption that the event date variance is proportional to the estimated period variance and is similar across securities. Replacing (14) in (12), we have a new statistic that is assumed to be distributed unit normal.

This statistic is well specified even when there are no changes in variance, but if that is the case the test is less powerful. For lower tailed alternative hypotheses, this parametric test rejects too often. A non-parametric test (like the generalised sign test described below) is more powerful.

**Multi-Week Tests**

All the tests described above can be easily changed to test for the significance of cumulative (standardised or not) abnormal returns. For example, for non-standardised residuals, the statistic is defined as:

\[
t = \frac{\sum_{l=1}^{L} \overline{AR}_l}{\sqrt{S^2\left(\frac{\sum_{l=1}^{L} \overline{AR}_l}{L}\right)}}
\]

where \(L\) denotes the weeks in the event window. If we assume independence over time, we get:

\[
S^2\left(\frac{\sum_{l=1}^{L} \overline{AR}_l}{L}\right) = \sum_{l=1}^{L} S^2(\overline{AR}_l)
\]

where \(S^2(\overline{AR}_l)\) is defined as in (8), (9) or (14). The statistic is distributed as Student-\(t\) with \(T-d\) degrees of freedom.

The standard deviations of multi-week periods, defined as in (16), fail to account for autocorrelation in average abnormal returns over the event window. Autocorrelation in event time may occur because of contemporaneous correlation across securities in calendar time. Failing to account for autocorrelation, usually leads to underestimation of the multi-week variance and therefore the tests above reject too often. When the diagnostic of autocorrelation is serious, one way to overcome the problem is to estimate the autocovariance matrix from the estimation period average abnormal returns and change the denominator in (15) with the sum of the elements of the estimated autocovariance matrix.

### 3.2. Non-Parametric Tests

Previous studies have shown that abnormal returns distributions show fat tails and are right skewed. Parametric tests reject too often when testing for positive abnormal performance and too seldom when testing for negative abnormal performance. When the assumption of normality of abnormal returns is violated, parametric tests are not well specified. Non-parametric tests are well-specified and more powerful at detecting a false null hypothesis of no abnormal returns.

#### Generalised Sign Test

The sign test is a simple binomial test of whether the frequency of positive abnormal residuals equals 50%. The generalised test is a refined version of this test by allowing the null hypothesis to be different from 0.5. To implement this test, we first need to determine the proportion of stocks in the sample that should have non-negative abnormal returns under the null hypothesis of no abnormal performance. The value for the null is estimated as the average fraction of stocks with non-negative abnormal returns in the estimation period. If abnormal returns are independent across securities, under the null hypothesis the number of non-negative values of abnormal returns has a binomial distribution with parameter \(p\). The alternative hypothesis, for any level of abnormal performance, is that the proportion is different than that prior. The advantage of the generalised sign test is that it takes into account the evidence of skewness in security returns.

The following statistic has an approximate unit normal distribution:

\[
GS = \frac{\left|p_0 - p\right|}{\sqrt{p(1-p)/N}}
\]

where \(p_0\) is the observed fraction of positive returns computed across stocks in one particular event week, or the average fraction of firms with non-negative abnormal returns for events occurring over multiple weeks.
Wilcoxon Signed-Ranks Test

This test considers that both the sign and the magnitude of abnormal returns are important. The statistic is given by:

\[ S_N = \sum_i r_i^+ \]  \hspace{1cm} (18)

where \( r_i^+ \) is the positive rank of the absolute value of abnormal returns. It is assumed that none of the absolute values are equal, and that each is different from zero. The sum is over the values of abnormal returns greater than zero. When \( N \) is large, the distribution of \( S_N \), under the null hypothesis of equally likely positive or negative abnormal returns, will be approximately a normal distribution with:

\[ E(S_N) = N(N+1)/4 \]
\[ \sigma^2(S_N) = N(N+1)(2N+1)/24. \]  \hspace{1cm} (19)

Corrado (1989) Rank Test

To implement the rank test, it is first necessary to transform each firm’s abnormal returns in ranks \( (K_i) \) over the combined period that includes the estimation and the event window \( (T_i) \):

\[ K_{it} = \text{rank}(AR_{it}) \]
\[ AR_{it} > AR_{is} \Leftrightarrow K_{it} > K_{is} \]  \hspace{1cm} (20)

The test then compares the ranks in the event period for each firm, with the expected average rank under the null hypothesis of no abnormal returns \( (\bar{K}_i = 0.5 + T_i/2) \). The test statistic for the null hypothesis is:

\[ R = \frac{1}{N} \sum_{i=1}^{N} (K_{it} - \bar{K}_i) \]
\[ S(\bar{K}) \]  \hspace{1cm} (21)

where

\[ S(\bar{K}) = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \frac{1}{N^2} \sum_{i=1}^{N} (K_{it} - \bar{K}_i)^2} \]  \hspace{1cm} (22)

This statistic is distributed asymptotically as unit normal. With multi-week event periods, the rank statistic is:

\[ R = \frac{1}{N} \sum_{i=1}^{N} (K_{it} - \bar{K}_i) \]
\[ S(\bar{K}) \sqrt{L} \]  \hspace{1cm} (23)

Under the assumption of no cross-sectional correlation, the statistic is distributed unit normal.

Cowen and Sergeant (1996) show that if the return variance is unlikely to increase, then Corrado’s rank test provides better specification and power than parametric tests. With variance increases this test is, however, misspecified. The Boehmer et al. (1991) standardised cross-sectional test is properly specified for upper tailed tests. Lower tailed alternatives can be best evaluated using the generalised sign test (please refer to Corrado and Zivney, 1992).

Bootstrapping procedure
This test consists of the following steps:
- first, each observation of the sample - consisting of abnormal returns for the cross-section of firms in event time - is indexed by a number and \( N \) randomly numbers are generated (where \( N \) is as above, the number of firms in the sample).
- second, generate a large number of bootstrapping samples (\( NS \)) from the sample above, for each event date. Thus, for each simulation, one computes the average abnormal return. The expected value and standard deviation of the bootstrapping distribution are drawn from the simulation averages. For the purpose of testing multi-week significance, we need to repeat the procedure for each event date (\( L \) times) and compute the cumulative abnormal returns for each bootstrapping.
- third, use the shift method (for example) to infer. For the shift model, reject the null hypothesis if:
  \[
  \frac{NGE + 1}{NS + 1} < \alpha
  \]  
  (24)

where \( NGE \) is the number of times the bootstrapping simulation generates average abnormal returns less than a criterion value. The criterion value is defined as the actual sample value (average abnormal return or cumulative average abnormal return) minus the population expected value (by assumption, zero) plus the expected value of the bootstrapping samples. \( \alpha \) is the size of the test.

4. VARIANCE CHANGES

The previous section highlighted the importance of assessing for changes in variance of abnormal returns following the event when choosing a well-specified and powerful test. Moreover, the issue of the impact of the event on total risk is interesting by itself.

4.1 Parametric Tests

**F Test for Equality of Two Population Variances**

This test assumes that abnormal returns from the estimation and event periods are independent and normally distributed. Under the null hypothesis, the variance of the event period is equal to the variance of the prior listing period. For each firm, the test statistic is:

\[
F = \frac{S^{2}_{l,b}(AR_{b})}{S^{2}_{b}(AR_{b})} 
\]  
(25)

where the subscript \( b \) and \( l \) label the variance estimates for the estimation and event periods. Under the null, the statistic is distributed \( F(T-d, L-d) \), where \( T \) and \( L \) are, as before, respectively, the number of weeks in the estimation and event and \( d \) depends on the degrees of freedom lost to estimate the variance.

To test variance changes across firms, the variance ratio is computed for each firm as before, and the null hypothesis that the average ratio across firms is equal to 1 is tested using a standard parametric Student t-test.

**Beavers’s/May’s U**

These tests involve the \( N \) event firms. Given

\[
U_{i} = \left( \frac{AR_{0,i}}{S(AR_{i})} \right)^{2} \]  
(26)

where \( S(AR_{i}) \) is defined as in (8), and if abnormal returns are distributed normal and are independent across firms, the squared standardised residual \( U_{i} \) is asymptotically distributed as \( F(1,T-d) \). Applying the normal approximation for sums of random variables the statistic
Event Study Tests: a brief survey

\[
Z = \frac{\sum_{i=1}^{N} U_i - N \frac{T - d}{T - d - 2}}{\sqrt{\frac{2(T - d)^2(T - d - 1)}{(T - d - 2)^2(T - d - 4)}}}
\]

(27)

is distributed unit normal. As before, \( N \) refers to the number of event firms and \( T \) are the number of weeks in the estimation period. A comparable test (May's U) uses the absolute value of standardised residuals instead of squared residuals. These two statistics are severely biased against the null hypothesis for non-normal abnormal returns and this bias does not vanish asymptotically.

4.2 Non-Parametric Rank Tests

Squared Rank Test

This non-parametric test compares, as in the \( F \)-test, the variance of the event period with the variance in the estimation period.

The first step involves computing the absolute value of abnormal returns for the estimation and the event weeks, for each firm. Then the two periods are combined and the absolute errors are ranked. Finally, the test statistic for the event period is defined as:

\[
Q_i = \sum_{l=1}^{I_i} (r_{il})^2
\]

(28)

where \( I \) refers to event weeks and \( r_{il} \) is the rank of the absolute value of abnormal returns. Under the null hypothesis of no variance shift, this statistic is tabulated (see Conover, 1984). If the number of weeks in estimation is greater than 10, this statistic can be approximated to the unit normal.\(^9\)

Rohrbach and Chandra (1989)

This non-parametric test evaluates whether variance shifts following listing are significant across firms. Squared abnormal returns are ranked for each firm for the combined sample of prior listing and event periods. Then within firm ranks are summed across securities. The test statistic is:

\[
C_0 = \sum_{i=1}^{N} \text{rank}(AR_{i0})^2
\]

(29)

The time series of the sum of ranks over the pre-listing period is the empirical distribution. For a right tail test, i.e., if the alternative hypothesis is an increase in variance, the empirical significance of the test is \( y/T \) where \( y \) is the number of observations in the empirical null distribution that exceed the test observation \( C_0 \). The null of no changes in variance is rejected if:

\[
y/T \leq \alpha
\]

(30)

where \( \alpha \) is the size of the test. The test is easily extended to a multi-week setting.

5. REFERENCES


Event Study Tests: A Brief Survey

Ana Paula Serra
Doutora em Finanças pela London Business School (Reino Unido).
Professora Auxiliar do Centro de Estudos Macroeconômicos e Previsão (CEMPRE) da universidade do Porto.
E-mail: aserra@fep.up.pt
Rua Dr. Roberto Frias
4200-456 - Porto – Portugal.

---

1 This study was previously disseminated as a WP (Working Papers da Faculdade de Economia da Universidade do Porto, nº117, May 2002).
2 CEMPRE - Centro de Estudos Macroeconômicos e Previsão - is supported by the Fundação para a Ciência e a Tecnologia, Portugal, through the Programa Operacional Ciência, Tecnologia e Inovação (POCTI) of the Quadro Comunitário de Apoio III, which is financed by FEDER and Portuguese funds.
4 The degrees of freedom depend on how we estimate the standard deviation of abnormal returns. For example, in the case of prediction errors from the one-factor market model, the degrees of freedom are T-2.
5 See, for example, Patell (1976).
6 It is common to adjust the standard error by the prediction error. For example, in the one-factor market model, the adjustment is:

\[ \sqrt{1 + \frac{1}{T} + \frac{(R_{m,t} - \bar{R}_m)^2}{\sum_i (R_{m,i} - \bar{R}_m)^2}} \]

where \( \bar{R}_m \) is the average market return in the estimation period and \( r \) is the prediction week. This term may be ignored if the estimation period is long.
7 Yet the standardised residual is not unit normal distributed, but follows a Student-\( t \) distribution. The variance of the normalised mean is \( d/d-2 \), where \( d \) are the degrees of freedom of the Student \( t \)-statistic. This simple adjustment works when the length of the estimation period is common to all firms.
8 Equivalently, one may look at the mean average abnormal returns over the event window (MCAR).
9 When comparing prior and event period variances, the approximation is

\[ q_p = \frac{T(T+1)(2T+1)}{6} + x_p\sqrt{\frac{LT(T+1)(2T+1)(8T+11)}{180}} \]

where \( q_p \) is the critical value for \( Q \) and \( x_p \) is the critical value of the unit normal distribution. \( T \) and \( L \) are, respectively, the number of weeks in the prior and event periods.