The bootstrap in event study methodology

Jérémie Lefebvre

M.Sc. Thesis
Département de Sciences économiques
Faculté des Arts et des Sciences
Université de Montréal

Advisor: Pr. Sílvia Gonçalves

Abstract

This thesis makes a contribution to the event study methodology. We first provide a unified methodology to test the main hypotheses of interest when the events cluster in time. Second, we apply a bootstrap inference method the our event study approach. The bootstrap here has two functions: (1) it provides a method of testing hypotheses for which there is no parametric counterpart, and (2) it corrects for possible biases due to non-normalities and serial correlation in the data. Finally, Monte Carlo simulations show in which case the inference methods proposed in this thesis are well specified.

Contents

1 Introduction

Since the seminal paper by Fama, Fisher, Jensen, and Roll (1969), whose number of citations exceeds 500[1], more than 500 event studies have been published[2]. Fama, Fisher, Jensen, and Roll (1969) were the first to provide a methodology for testing the statistical significance of the effect an event of interest has on a firm or portfolio return. This methodology has proven to be very useful in a variety of finance related fields such as corporate finance, accounting, management, etc. Examples are studies of the impact of mergers and acquisitions, stock splits, new legislations, earning announcements, and other finance related events, on the profitability of firms.

A vast literature on the theory of event study methods also exists. The main references are Campbell, Lo, and MacKinlay (1997), Binder (1998) and Khotari and Warner (2007). But many papers extend the basic methodology in several directions. In this thesis, we are interested in the case where the event of interest occurs during the same calendar day for every firm under study. This has been called "clustering" in the literature. The first papers that studied this case are Schipper and Thompson (1983, 1985) and Binder (1985). They use an econometric method called MVRM (for Multivariate Regression Method). This method is a special case of the more general SURE methods in econometrics (Seemingly Unrelated Regression Equations).

The contribution of our study is first to provide a unified framework for estimating and testing abnormal returns. For this, we simplify greatly the MVRM approach by the use the Frisch-Waugh-Lowell theorem and also

[1]According to Binder (1998).

[2]According to the census made by Khotari and Warner (2007)

by the use of well known properties of the vec operator. Our methodology provides a unified means for testing different kinds of hypotheses.

Another extension of the event study methodology that is provided in the literature is for improving robustness of inference. The classical hypothesis tests are almost all based on the assumption of independent and normally distributed residuals. However, it is well known that financial data are characterized by non-normalities, expecially non-zero skewness and excess kurtosis, but also time series dependence (see e.g. Kramer, 2001).

A strand of literature has tried to provide robust inference methods in the case of non-normalities (Kramer, 2001), clustering events (Chou, 2004) and serial correlation of the residuals (Hein and Westfall, 2004). The technique used is the bootstrap. Our contribution is to go further in this direction. The bootstrap is an inference method that uses resamples of the data in order to study in a non parametric way the distribution of the test statistic. This method has proven to be very useful either when no parametric distribution exists for some test statistics, or when small samples and non-normalities alter the existing parametric methods. The bootstrap was introduced by Efron (1979) and has been studied extensively in the econometric literature (for references, see Horowitz, 2001).

Some characteristics of financial data contribute to bias existing parametric inference methods. The type of bias that we study in this thesis is when the true error type I is not equal to the specified level of the test. This means that the test will either over-reject the null hypothesis or under-reject (i.e. will be too conservative). We control the level of tests by the use of Monte Carlo simulations.

In the event study literature, two types of Monte Carlo simulations are used to assess event study tests. The first type is what we could call "pure" Monte Carlo. It generates artificial data from a pre-specified distribution (i.e. the data generating process). When using this approach, researchers usually use distributions that closely match characteristics of real financial data. The second type of Monte Carlo experiment is called "historical" Monte Carlo (see Binder, 1985; Butler and Frost, 1992). It uses a vast sample of historical financial data, and it uses random subsamples (instead of artificial data). The main advantage of this last method is that the researcher does not have to specify fully the data generating process. On the other hand, it has the drawback that if the test turns out to be biased, the researcher cannot determine exactly which characteristic of financial data is responsible for it. In this thesis, we use "pure" Monte Carlo methods in which we test separately the robustness of the inference methods for each aspect. This permits us to pin-point to the exact cause of mis-specification of some tests.

This thesis is organized as follows. The next section introduces the basic event study methodology as it is exposed, for instance, in Campbell, Lo, and MacKinlay (1997). Then, in Section 3 we expose in length our unified approach for event studies, and we precisely characterize three types of hypotheses that can be tested, as well as their test statistics. The following section reviews the existing bootstrap methods for event study analysis. Section 5 shows descriptive statistics from real financial data, in order to precisely specify the data generating process used in our Monte Carlo experiment. In Section 6, we present and analyze the results of our Monte Carlo experiments, in comparison with other existing inference methods. Finally, Section 7 concludes.

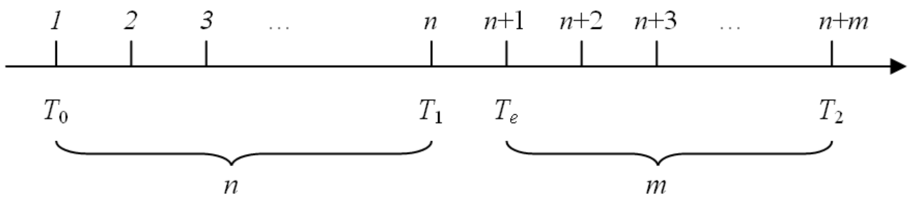
2 The classical event study methodology

The classical event study methodology is explained in Campbell, Lo, and MacKinlay (1997, Chap. 4). In this section we introduce the notation and summarize the method. An event study aims at measuring the effect a given event has on a security's return. In order to do this, the researcher uses a benchmark model for predicting the returns. The parameters of this model are estimated using data prior to the event, i.e. the estimation window. In the event window, where the actual event occurs, the predicted returns from the model are compared with the true returns. Then, an inference is made to determine if the difference between the true returns and the predicted returns is statistically different from zero.

The most common benchmark model is the market model, inspired by the CAPM (Capital Asset Pricing Model). The data used is the firm's returns and the market returns (as proxied by the returns on a relevant market index). The parameters are the firm's β and the intercept, α. The so-called normal returns are the firm's returns during the event window predicted by the market model. The normal returns minus the firm's actual returns are what we call the abnormal returns. If these abnormal returns are not zero, the researcher concludes that the event has affected the returns of the firm.

The estimation window has typically 250 daily observations, i.e. one year of returns. The event window can have different lengths, depending on the application. In this paper, we concentrate on short term effects of events, so

Figure 1: Time line of an event study



This figure represents the time line of a typical event study. The estimation window starts at time T_0, ends at T_1 and has n observations. Typically, the event of interest occurs at time T_e, i.e. at observation number $n+1$. The event window has m observations, ends at time T_2. In total, there are $T = n + m$ observations.

that the event window has ten observations. Long term event studies have their own issues, which we do not treat here. See reviews by Binder (1998) and Khotari and Warner (2007).

The parameters of the benchmark model are estimated in the estimation window with observations from T_0 to T_1, i.e. n observations. Then, the abnormal returns are computed for the m periods of the event window. Figure 1 illustrate the notation and timing of a typical event study.

The parameters are estimated using the following regression equation:

$$r_t = \alpha + \beta rm_t + \varepsilon_t \ , \ t \in [T_0, T_1] \ ,$$

where r_t is the return of the firm between at time t, rm_t is the return of the market index and ε_t is an error term. Then, using the estimated parameters and the explanatory variable (here: the market return), we obtain the normal returns for the event window. The abnormal returns, AR, are then obtained by subtracting the actual return by the normal return:

$$
\begin{aligned}
r_t^* &= \mathrm{E}\left[r_t | rm_t\right] \ , \ t \in [T_1+1, T_2] \\
&= \hat{\alpha} + \hat{\beta} rm_t \ ,
\end{aligned}
$$

where r_t^* is the predicted, or normal, returns.

$$
\begin{aligned}
AR_t &= r_t - \mathrm{E}\left[r_t | rm_t\right] \ t \in [T_1+1, T_2] \\
&= r_t - r_t^*
\end{aligned}
$$

The statistic of interest is the cumulative abnormal return over the event window, or *CAR* for short. The null hypothesis is that the event has no effect on the firm's return, i.e. $H_0 : CAR = 0$. To determine a test statistic and its distribution, we need some assumption on the *AR*s.

If we assume that the *AR*s are i.i.d. normally distributed, then the *CAR* is the sum of m normally distributed variables. The *AR*s have mean zero under the null hypothesis. Their variance has a component that is due to the variance of the residuals of the estimation window regression, σ_ε^2 and another component that is due to the estimation uncertainty.

$$
\mathrm{V}\left[AR | rm\right] = \sigma_\varepsilon^2 I_m + \sigma_\varepsilon^2 X^* \left(X'X\right)^{-1} X^{*\prime} \ , \qquad (1)
$$

where σ_ε^2 is the variance of the residuals from the estimation window regression, I_m is an identity matrix of size m, X^* is a $m \times 2$ matrix of the regressors in the event window and X is the matrix of the regressors in the estimation window. See Campbell, Lo, and MacKinlay (1997, p. 159) for a formal derivation of this result. We can now derive the distribution of the *CAR*:

$$\begin{aligned}\mathrm{E}\left[CAR|rm\right] &= \iota_m'\mathrm{E}\left[AR|rm\right] \\ \mathrm{V}\left[CAR|rm\right] &= \iota_m'\mathrm{V}\left[AR|rm\right]\iota_m \ ,\end{aligned}$$

where ι_x is a column vector of ones of size x. The test statistic is called J_1 and is computed as follows:

$$J_1 = \frac{\mathrm{E}\left[CAR|rm\right]}{\sqrt{\mathrm{V}\left[CAR|rm\right]}} \ . \tag{2}$$

The test statistic follows a Student t distribution with $n-2$ degrees of freedom (because the σ_ε^2 was estimated with n observations and there are 2 parameters: α and β).

In the case where multiple firms are affected by the event, the case that interests us, the test can be done by aggregating the *CAR*s of the firms. This is done by taking the average of the individual *CAR*s, or $\overline{CAR}$. Under the null hypothesis, no firms are affected by the event, implying that the expected value of $\overline{CAR}$ is zero. More formally, $H_0 : \overline{CAR} = \frac{1}{N}\sum_{i=1}^{N} CAR_i = 0$ and the alternative hypothesis is $H_1 : \overline{CAR} \neq 0$ (two-sided test). To compute the variance, we have to assume that the individual *CAR*s are independent (this is true when there is no clustering), and we sum the individual variances. The test statistic then becomes:

$$J_1 = \frac{\overline{CAR}}{\sqrt{\frac{1}{N^2}\sum_{i=1}^{N} \mathrm{V}\left[CAR_i|rm\right]}} = \frac{\frac{1}{N}\sum_{i=1}^{N} CAR_i}{\sqrt{\frac{1}{N^2}\sum_{i=1}^{N} \mathrm{V}\left[CAR_i|rm\right]}} = \frac{\sum_{i=1}^{N} CAR_i}{\sqrt{\sum_{i=1}^{N} \mathrm{V}\left[CAR_i|rm\right]}}\,.$$

Campbell, Lo, and MacKinlay (1997) provide no small sample distribution for this statistic, only an asymptotic distribution, which is the standard normal distribution as the number of firms, N, tends to infinity. This means that the test is best used when the number of firms under study is large.

Campbell, Lo, and MacKinlay (1997) provide an alternative test statistic called J_2, which is similar to the J_1 test. The difference is that the J_2 test is performed by scaling each *CAR* by its standard deviation (the standard deviation of the *CAR* is closely related to the standard deviation of the idiosyncratic risk, σ_ε). In other words, it is a weighted average of the *CAR*s instead of the average *CAR* like in the J_1. The rationale of this method is to give more weight to the firms with lower residual variance, in order to increase the power of the test (see Campbell, Lo, and MacKinlay, 1997, p. 162). See de Roon and Veld (1998) for a similar approach.

The J_2 statistic is constructed by weighting the *CAR* of each firm by its standard deviation, then by summing the standardized *CAR*s and finally by multiplying by an adjustment factor:

$$J_2 = \left[\frac{N(n-2)}{n-4}\right]^{-\frac{1}{2}} \sum_{i=1}^{N} SCAR_i \ ,$$

where $SCAR$ stands for standardized CAR and is defined as:

$$SCAR_i = \frac{CAR_i}{\sqrt{\mathrm{V}\left[CAR_i|rm\right]}} \ .$$

The J_2 statistic is the sum of N independent t statistics with $n-2$ degrees of freedom. Each standardized CAR has mean zero and variance $\frac{n-2}{n-4}$ under the null. Assuming the CARs to be independent, the sum of N standardized CARs has mean zero and variance $\frac{N(n-2)}{n-4}$. This means that multiplying by $\left[\frac{N(n-2)}{n-4}\right]^{-\frac{1}{2}}$, we have a statistic that follows the standard normal distribution asymptotically, as N goes to infinity (see Campbell, Lo, and MacKinlay, 1997, Chap. 4).

Although Campbell, Lo, and MacKinlay (1997) treat only one hypothesis to test in event studies, in the literature we see more hypotheses in which researchers are interested. Indeed, the hypothesis treated above (call it H_1) is typically used when the researcher knows that the event induces abnormal returns to firms in the same direction (either positive or negative) and he wants to test whether the average effect across firms is statistically significant or not. Alternatively, a researcher could want to do an event study when all firms are not thought to have abnormal returns in the same direction (some firms might profit from the event, others might be worse off). Then, a more adapted hypothesis (H_2) would be that all CARs are jointly equal to zero. A third possible hypothesis (H_3) would be to study whether all abnormal returns for all firms and all event period are jointly equal to zero.

This hypothesis is typically used when the number of firms under study is small.

In all hypotheses, the length of the event window can vary depending on the research question. But the third hypothesis uses usually a short event window. The second and third hypotheses are mostly used when there are reasons to think that returns react slowly to the event, or when the exact time at which the event occurred in not known precisely. In the next section, we will treat all three hypotheses using two approaches: in the spirit of the J_1 test (i.e. no weighting) and in the spirit of the J_2 test (i.e. by weighting the abnormal returns by their standard deviation).

As mentioned in the introduction, there are three problems with this methodology. First, most of the test statistics that have been developed by econometricians are based on the assumption of normality and are valid asymptotically. Moreover, some test statistics have distributions that are defined only approximately (see for instance Butler and Frost (1992)). This problem is especially severe when the sample size (either n or N) is small and is treated by the use of resampling methods like the bootstrap. Section 4 will look what has been suggested in the literature to treat this problem.

The second problem is the fact that residuals exhibit cross-sectional correlation when the event is clustering in time. This aspect is explicitly taken into account by the method called MVRM (for Multivariate Regression Model). The next section will develop in length this method. The third problem is related to serial correlation of the residuals. In Section 4, we will see how researchers in the literature have treated it.

The next section will study in length the estimation techniques and the test statistics for testing the three above mentioned hypothesis.

3 A unified methodology

In this section we explain the event study methodology and testing methods that have been introduced in the literature. The main estimation techniques are in Campbell, Lo, and MacKinlay (1997), but the approach does not include the dummy variable method from Karafiath (1988) and the MVRM framework from Schipper and Thompson (1983, 1985). The main contribution of our exposition is to merge these approaches and to provide a unified and consistent approach that embodies and simplifies the three approaches.

As for the inferences, we will distinguish between three types of hypotheses to test. All these hypotheses can be done in a J_1 spirit or in a J_2 spirit.

In the next subsection, we briefly discuss some properties of the Kronecker product and the vec operator that will be used in the following development. Subsection 3.2 explains how to estimate the abnormal returns in the MVRM framework. Subsection 3.3 explains a way to simplify the calculations both analytically and computationally using the Frisch-Waugh-Lowell (or FWL) theorem. Subsection 3.4 shows that estimating the abnormal returns equation by equation (i.e. firm by firm) is mathematically equivalent to estimating them in the MRVM framework, when one treats correctly the covariance matrix of the residuals. Subsection 3.6 explains what are the 2 types of tests that are explained in Campbell, Lo, and MacKinlay (1997): the J_1 and J_2 tests. The three most common hypothesis to test in event studies are explained in Subsection 3.5. Each of them will be developed separately in Subsections 3.7, 3.8 and 3.9, and each of them will treat both the J_1 and J_2 approaches.

3.1 Preleminary: the Kronecker product and vec operator

The Kronecker product uses the symbol $\otimes$ and means that 2 matrices are to be multiplied, but differently from the normal matrix multiplication. Let A be a matrix of size $m \times n$, B a matrix of size $o \times p$. Then $A \otimes B$ means that each element of A multiplies the whole matrix B and the result forms matrix C, which is of size $mo \times np$. Here is an example:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \otimes B = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix}$$

There is no conformity requirement for the Kronecker product, contrary to the normal matrix multiplication. Matrices A and B can be of any size, including vectors and scalars. If one of the matrices is actually a scalar, the Knonecker product collapses to a normal product of a scalar with a matrix. If both matrices A and B are scalars, the Kronecker product is simply the normal multiplication of two scalars.

Useful properties include:

1.
$$(A \otimes B)' = A' \otimes B'$$

2.
$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

   for conforming matrices A, C and B, D.

3.
$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

for invertible matrices A and B.

The vec operator transforms a matrix into a column vector by stacking all the columns of the argument. Here is an example:

$$\text{vec}(A) = \text{vec}\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix}$$

Useful properties include:

1. $$\big(B' \otimes A\big)\,\text{vec}(C) = \text{vec}(ACB)$$

   for conforming matrices A, B and C.

2. $$\text{vec}(C)'(B' \otimes A)\text{vec}(C) = \text{tr}(C'ACB)$$

   for conforming matrices A, B and C, and where tr is the trace operator (i.e. the sum of all the diagonal elements of the argument matrix).

3.2 Estimation

This subsection treats the estimation of abnormal returns. We will use the MVRM framework introduced first by Schipper and Thompson (1983, 1985) and the dummy variable technique as formalized in Karafiath (1988).

The MVRM framework is a special case of the SURE (Seemingly Unrelated Regression Equations) methodology that has the same regressor for every equation[3]. In event studies, the regressor is the market return and a set of dummy variables. The dummy variables *a la* Karafiath (1988) is a set of m dummy variables, where m is the length of the event window for which

[3]See for instance Greene (2003, Chap. 14) for an introduction to SURE methods.

we want to assess the abnormal returns of the firms. Each dummy variable is a column of zeros with a single 1 at the position of the corresponding event window observation.

The estimation window has n observations, and we have N firms that are clustering exactly in time (hence the MVRM framework for identical regressors). Let T be the total length of the observation and estimation window, $T = n + m$. Let r_i be the vector of returns of firm i, $i = 1, \ldots, N$. Let ε_i be the vector of regression errors for firm i that are i.i.d. with mean zero and variance σ_i^2. At a given time period, the errors of the firms are correlated (cross-sectional correlation):

$$\mathrm{cov}(\varepsilon_{is}, \varepsilon_{jt}) = \begin{cases} \sigma_i^2 \quad \text{if } i = j \text{ and } s = t \\ \sigma_{ij} \quad \text{if } i \neq j \text{ and } s = t \\ 0 \quad \text{if } i \neq j \text{ and } s \neq t \end{cases}$$

For now, we assume that there is no serial correlation in the error terms (hence the i.i.d. assumption above). We will relax this assumption when treating the bootstrap below.

The regression model is the following.

$$r = G\gamma + \varepsilon \ , \tag{3}$$

where

$$r = \begin{bmatrix} r_1' & r_2' & \cdots & r_N' \end{bmatrix}', \text{ a } NT \times 1 \text{ vector,}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1' & \varepsilon_2' & \cdots & \varepsilon_N' \end{bmatrix}', \text{ a } NT \times 1 \text{ vector,}$$

$$G = \begin{bmatrix} \bar{X} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{X} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \bar{X} \end{bmatrix} = I_N \otimes \bar{X} \ ,$$

$$\bar{X} = [X \quad D] \ , \quad X = [\iota_T \quad rm] \ , \quad D = \begin{bmatrix} \mathbf{0}_{n\times m} \\ I_m \end{bmatrix} \ ,$$

where ι_T is a column vector of length T and rm is the vector of market returns of length T, and

$$\gamma = \begin{bmatrix}\alpha_1 & \beta_1 & \delta_1' & \alpha_2 & \beta_2 & \delta_2' & \cdots & \alpha_N & \beta_N & \delta_N'\end{bmatrix}'$$

where α_i is the intercept of each firm, β_i is the slope estimator (the firm's beta) and δ_i is the firm's vector of abnormal returns (of length m).

Since we allow for cross-sectional correlation between firms but no serial correlation (for the moment), the regression errors have the following covariance matrix:

$$\mathrm{V}\left[\varepsilon|X\right] = \Sigma_N \otimes I_T \ ,$$

where Σ_N is a positive definite covariance matrix (the index N is simply to emphasize that the matrix is of size $N \times N$).

The estimation is done by simple OLS.

$$\begin{aligned} \hat{\gamma} &= \left(G'G\right)^{-1} G'r \\ &= \left[\left(I_N \otimes \bar{X}'\right)\left(I_N \otimes \bar{X}\right)\right]^{-1}\left(I_N \otimes \bar{X}'\right) r \\ &= \left(I_N \otimes \bar{X}'\bar{X}\right)^{-1}\left(I_N \otimes \bar{X}'\right) r \\ &= \left[I_N \otimes \left(\bar{X}'\bar{X}\right)^{-1} \bar{X}'\right] r \end{aligned}$$

If we assume that the firms have different variance of their residuals (that is, they have a different variance of idiosyncratic risk), one could use GLS instead of OLS. But it is known in the literature that the GLS estimator is mathematically equivalent to the OLS estimator in the case of the MVRM framework. See for instance Greene (2003).

The expectation of the estimator is:

$$\mathrm{E}\left[\hat{\gamma}|\, X\right] = \gamma + \left[I_N \otimes \left(\bar{X}'\bar{X}\right)^{-1} \bar{X}'\right] \varepsilon\ .$$

Its variance is:

$$\begin{aligned}
\mathrm{V}\left[\hat{\gamma}|\, X\right] &= \left[I_N \otimes \left(\bar{X}'\bar{X}\right)^{-1} \bar{X}'\right] \mathrm{V}\left[\varepsilon|X\right] \left[I_N \otimes \bar{X}\left(\bar{X}'\bar{X}\right)^{-1}\right] \\
&= \left[I_N \otimes \left(\bar{X}'\bar{X}\right)^{-1} \bar{X}'\right] \left(\Sigma_N \otimes I_T\right) \left[I_N \otimes \bar{X}\left(\bar{X}'\bar{X}\right)^{-1}\right] \\
&= \Sigma_N \otimes \left(\bar{X}'\bar{X}\right)^{-1}
\end{aligned}$$

An estimator for Σ_N could be for instance

$$\hat{\Sigma}_N = \frac{E'E}{T-m-2} = \frac{E'E}{n-2}$$

where E is a $T \times N$ matrix of the residuals, i.e. $E = [\hat{\varepsilon}_1 \quad \hat{\varepsilon}_2 \quad \cdots \quad \hat{\varepsilon}_N]$. See Schipper and Thompson (1985).

3.3 Short-cut: using the FWL theorem

There is a way to simplify the computations by using the FWL (for Frisch-Waugh-Lowell) theorem (Davidson and MacKinnon, 2004, see e.g.). This theorem provides a method to estimate directly the abnormal returns, without having the αs and βs in the estimator vector.

Assume for now that we have $N = 1$ and so we leave the MVRM framework. We have $\bar{X} = [X \quad D]$ as defined above. Then

$$\begin{aligned} r &= \bar{X}\gamma + \varepsilon \\ &= X \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + D\delta + \varepsilon \end{aligned}$$

If we pre-multiply all the terms by the idempotent projection matrix $M = I_T - X\left(X'X\right)^{-1} X'$, we have

$$\begin{aligned} Mr &= MX \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + MD\delta + M\varepsilon \\ &= MD\delta + \xi \end{aligned}$$

because $MX = 0$. The estimator for δ becomes:

$$\begin{aligned} \hat{\delta} &= \left[(MD)'(MD)\right]^{-1}(MD)' r \\ &= \left(D'MD\right)^{-1} D'Mr \end{aligned}$$

We can simplify further by noticing that the matrices D and M can be written as block matrices:

$$D = \begin{bmatrix} \mathbf{0}_{n\times m} \\ I_m \end{bmatrix} \quad M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

where the four matrices M_{11}, M_{12}, M_{21} and M_{22} are of size $n\times n$, $m\times n$, $n\times m$ and $m \times m$, respectively. The matrix multiplication $D'MD$ can simplify to:

$$\begin{aligned} D'MD &= \left[\mathbf{0}_{m\times n} \quad I_m\right] \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \mathbf{0}_{n\times m} \\ I_m \end{bmatrix} \\ &= M_{22} \end{aligned}$$

The estimator for δ then becomes simply:

$$\hat{\delta} = M_{22}^{-1} D' M r \ .$$

We now come back to the MVRM framework where $N \geq 1$. The FWL theorem can be applied to the MVRM model (3) by pre-multiplying each term by $(I_N \otimes M)$. Let y be the vector of firm returns that has been pre-multiplied by the above matrix. Similary, let Z be the pre-multiplied matrix G, i.e.

$$\begin{aligned} Z &= (I_N \otimes M)\, G \\ &= (I_N \otimes M)\left(I_N \otimes \bar{X}\right) \\ &= \left(I_N \otimes M\bar{X}\right) \\ &= (I_N \otimes MD) \ . \end{aligned}$$

The model becomes

$$y = Z\delta + \xi \ ,$$

and the estimator the the abnormal returns is:

$$\begin{aligned} \hat{\delta} &= (Z'Z)^{-1} Z'y \\ &= \left[(I_N \otimes MD)'(I_N \otimes MD)\right]^{-1} (I_N \otimes MD)'(I_N \otimes M)r \\ &= (I_N \otimes D'MD)^{-1}(I_N \otimes D'M)(I_N \otimes M)r \\ &= (I_N \otimes M_{22}^{-1})(I_N \otimes D'M)r \\ &= (I_N \otimes M_{22}^{-1} D'M)r \ . \end{aligned}$$

And the variance of the estimator is

$$
\begin{aligned}
\mathrm{V}\left[\hat{\delta}\middle| X\right] &= (I_N \otimes M_{22}^{-1} D' M) \mathrm{V}\left[(I_N \otimes M)\varepsilon | X\right] (I_N \otimes M D M_{22}^{-1}) \\
&= (I_N \otimes M_{22}^{-1} D' M)(I_N \otimes M) \mathrm{V}\left[\varepsilon | X\right] (I_N \otimes M)(I_N \otimes M D M_{22}^{-1}) \\
&= (I_N \otimes M_{22}^{-1} D' M)(\Sigma_N \otimes I_T)(I_N \otimes M D M_{22}^{-1}) \\
&= \Sigma \otimes M_{22}^{-1} D' M D M_{22}^{-1} \\
&= \Sigma \otimes M_{22}^{-1}
\end{aligned}
$$

This simplification using the FWL theorem will prove to be useful, below, for imposing linear restrictions on the abnormal returns for the purpose of testing.

3.4 SURE or equation by equation?

In this section, we will show that actually, we do not need the Kronecker product for estimation. We show that the SURE framework can be avoided by estimating the abnormal returns equation be equation, and this leads to the same result.

This might be useful when, instead of stacking all the firms' returns in a single vector, one would have to keep the returns in a matrix of size $T \times N$.

This properties of the MVRM method is simply shown by using the vec operator (see Subsection 3.1). Start with the above equation for the OLS estimation of the abnormal returns. Then, define the matrix R to be the unstacked firms' returns, i.e. $R = [r_1 \quad r_2 \quad \cdots \quad r_N]$. With the vec operator, this means that $\mathrm{vec}(R) = r$. Then, using the properties of the vec operator and the Kronecker product, we have that:

$$
\begin{aligned}
\hat{\delta} &= (I_N \otimes M_{22}^{-1} D' M) r \\
&= (I_N \otimes M_{22}^{-1} D' M) \mathrm{vec}(R) \\
&= \mathrm{vec}(M_{22}^{-1} D' M R I_N) \\
&= \mathrm{vec}(M_{22}^{-1} D' M R) \ .
\end{aligned}
$$

We can then define $\hat{\Delta}$ to be the un-stacked matrix of abnormal returns, which is simply $\hat{\Delta} = M_{22}^{-1} D' M R$. This matrix is of size $m \times N$. In other words, $\hat{\Delta} = \begin{bmatrix} \hat{\delta}_1 & \hat{\delta}_2 & \cdots & \hat{\delta}_N \end{bmatrix}$ and $\mathrm{vec}(\hat{\Delta}) = \hat{\delta}$.

We have just shown that the estimation of the abnormal returns in the MVRM framework is mathematically equivalent to equation by equation estimation. Also, we have shown how to avoid completely the vec operator and the Kronecker product in the computations.

In the next subsections, we will speak about hypothesis testing.

3.5 Three hypotheses to test

The most common hypotheses to test from the literature (Chou, 2004, see e.g.) in event studies are:

1. The average cumulative abnormal return across firms is equal to zero.

$$
H_1 : \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{m} \delta_{ij} = 0 \quad \text{or} \quad \frac{1}{N} \iota_m' \Delta \iota_N = 0
$$

2. All cumulative abnormal returns during the event period are all jointly equal to zero.

$$
H_2 : \sum_{j=1}^{m} \delta_{i,j} = 0 \ \forall \ i \quad \text{or} \quad \iota_m' \Delta = \mathbf{0}_{1 \times N}
$$

3. All abnormal returns are jointly equal to zero for all firms and all event periods.

$$H_3 : \delta_{ij} = 0 \ \forall \ i, j \quad \text{or} \quad \Delta = \mathbf{0}_{m \times N}$$

3.6 Two approaches for testing

There seems to be two approaches for testing the above hypotheses. The first one is in the spirit of the J_1 test of Campbell, Lo, and MacKinlay (1997). The test statistic is either the t statistic, in the case of hypothesis 1, or the Wald statistic in the case of hypotheses 2 and 3.

The second approach is used when there is heterogeneity across firms, i.e. when the variance of the residuals is different from one firm to another. In other words, when the idiosyncratic variance differs across firms. This approach gives a different weight to the abnormal returns depending on the residual variance. It is in the spirit of the J_2 test in Campbell, Lo, and MacKinlay (1997). It can be shown that hypotheses 1 and 2 are tested with an increased power when this approach is used. But the third hypothesis turns out not to be affected when using this approach (see below).

In each hypothesis, a test statistic is computed similarly to the Wald statistic, see Schipper and Thompson (1985). For this, we need a restriction matrix A' that will pre-multiply the estimator of the abnormal returns. For the first approach, in the spirit of the J_1 test, the Wald statistic becomes:

$$\begin{aligned} \tilde{W}^1 &= \left(A'\hat{\delta}\right)' \mathrm{V}\left[A'\hat{\delta}\middle| X\right]^{-1} \left(A'\hat{\delta}\right) \\ &= \hat{\delta}' A \left\{A' \mathrm{V}\left[\hat{\delta}\middle| X\right] A\right\}^{-1} A'\hat{\delta} \end{aligned}$$

For the first hypothesis, there is a single linear restriction, so that A is a vector, and thus a t statistic can be used instead of the Wald statistic:

$$\tilde{t}^1 = \frac{A'\hat{\delta}}{\sqrt{\mathrm{V}\left[A'\hat{\delta} \middle| X \right]}} = \frac{A'\hat{\delta}}{\sqrt{A'\mathrm{V}\left[\hat{\delta} \middle| X \right] A}}$$

In the second approach, the J_2-type test, we need a weighting matrix for the abnormal returns. This matrix is obtained from the estimated variance of the abnormal returns. Define ψ such that $\psi^2 = \psi\psi = \Sigma_N^{-1}$ and similarly, define θ such that $\theta^2 = \theta\theta = M_{22}$. We then see that:

$$\begin{aligned}
\mathrm{V}\left[\hat{\delta} \middle| X \right]^{-1} &= \left(\Sigma_N \otimes M_{22}^{-1} \right)^{-1} \\
&= \Sigma_N^{-1} \otimes M_{22} \\
&= \psi^2 \otimes \theta^2 \\
&= (\psi \otimes \theta)(\psi \otimes \theta)
\end{aligned}$$

The idea behind the J_2 test is to pre-multiply the parameter vector $\hat{\delta}$ by $(\psi \otimes \theta)$. In this case, the Wald statistic becomes:

$$\begin{aligned}
\tilde{W}^2 &= \left[A'(\psi \otimes \theta)\hat{\delta} \right]' \mathrm{V}\left[A'(\psi \otimes \theta)\hat{\delta} \middle| X \right]^{-1} \left[A'(\psi \otimes \theta)\hat{\delta} \right] \\
&= \hat{\delta}'(\psi \otimes \theta)A \left\{ A'(\psi \otimes \theta)\mathrm{V}\left[\hat{\delta} \middle| X \right] (\psi \otimes \theta)A \right\}^{-1} A'(\psi \otimes \theta)\hat{\delta}
\end{aligned}$$

For the first hypothesis, the J_2 approach gives:

$$\tilde{t}^2 = \frac{A'(\psi \otimes \theta)\hat{\delta}}{\sqrt{\mathrm{V}\left[A'(\psi \otimes \theta)\hat{\delta}\middle| X\right]}} = \frac{A'(\psi \otimes \theta)\hat{\delta}}{\sqrt{A'(\psi \otimes \theta)\mathrm{V}\left[\hat{\delta}\middle| X\right](\psi \otimes \theta)A}}$$

The difference between, on one side, the J_1 and J_2 statistics as they are presented in Campbell, Lo, and MacKinlay (1997), and on the other side the $\tilde{t}^1$ and $\tilde{t}^2$ is that here we have only firms that cluster in time, so that the regressors are exactly the same (the market return is the same for all firms). In the book, they explain the methodology when firms do not cluster, and thus they have different values for the corresponding market return. But in this case, there is no problem of cross-sectional correlation.

In the next subsections, we will derive with more details the exact test statistics to compute.

3.7 Hypothesis 1

As shown above, hypothesis 1 means that the average cumulative abnormal return is zero. The t statistic is used here. Econometric theory tells us that multiplying by a constant does not change the value of the test statistic. This means that the test is the same as testing whether the sum (instead of the average) of the cumulative abnormal returns is zero. We will thus test the following linear restriction:

$$A'\hat{\delta} = (\iota_N' \otimes \iota_m')\hat{\delta} = \iota_m'\hat{\Delta}\iota_N$$

The variance is:

$$\begin{aligned}
\mathrm{V}\left[A'\hat{\delta}\middle|X\right] &= A'\mathrm{V}\left[\hat{\delta}\middle|X\right]A \\
&= (\iota_N' \otimes \iota_m')\left(\Sigma_N \otimes M_{22}^{-1}\right)(\iota_N \otimes \iota_m) \\
&= \iota_N'\Sigma_N\iota_N \otimes \iota_m' M_{22}^{-1}\iota_m \ ,
\end{aligned}$$

which is a scalar. The t statistic is:

$$\tilde{t}_1^1 = \frac{\iota_m'\hat{\Delta}\iota_N}{\sqrt{(\iota_N'\Sigma_N\iota_N) \times (\iota_m' M_{22}^{-1}\iota_m)}} \ .$$

For the case where $\Sigma_N = \sigma^2 I_N$ and $\mathrm{V}[\varepsilon|X] = \sigma^2 I_{NT}$, the test statistic is distributed exactly as Student t distribution with $N(n-2)$ degrees of freedom (following standard econometric theory). But when there is heteroscedasticity across firms and structure is imposed on the covariance matrix of the residuals, we are in a GLS framework. According to Greene (2003), the distribution of the test statistic is the same but is valid asymptotically (as n tends to infinity). For this result to hold, we need only consistency of the estimates of the parameters in the covariance matrix (which is the case for the estimator above, proposed by Schipper and Thompson (1983, 1985)).

Now for the J_2-type approach, the restriction matrix is the same, but the abnormal return vector is pre-multiplied by a weighting matrix $\psi \otimes \theta$:

$$\begin{aligned}
A'(\psi \otimes \theta)\hat{\delta} &= (\iota_N' \otimes \iota_m')(\psi \otimes \theta)\hat{\delta} \\
&= (\iota_N'\psi \otimes \iota_m'\theta)\hat{\delta} \\
&= \mathrm{vec}\left(\iota_m'\theta\hat{\Delta}\psi\iota_N\right) \\
&= \iota_m'\theta\hat{\Delta}\psi\iota_N \ .
\end{aligned}$$

The variance is:

$$
\begin{aligned}
\mathrm{V}\left[A'(\psi\otimes\theta)\hat{\delta}\,\middle|\,X\right] &= (\iota_N'\otimes\iota_m')(\psi\otimes\theta)\mathrm{V}\left[\hat{\delta}\,\middle|\,X\right](\psi\otimes\theta)(\iota_N\otimes\iota_m) \\
&= (\iota_N'\otimes\iota_m')(\psi\otimes\theta)\left(\Sigma_N\otimes M_{22}^{-1}\right)(\psi\otimes\theta)(\iota_N\otimes\iota_m) \\
&= \iota_N'\psi\Sigma_N\psi\iota_N\otimes\iota_m'\theta M_{22}^{-1}\theta\iota_m \\
&= \iota_N'\iota_N\otimes\iota_m'\iota_m \\
&= N\otimes m \\
&= Nm\ .
\end{aligned}
$$

The t statistic is:

$$
\tilde{t}_1^2 = \frac{\iota_m'\theta\hat{\Delta}\psi\iota_N}{\sqrt{Nm}}\ .
$$

The distribution of this test statistic is not known. In the case of non clustering, i.e. $\mathrm{V}[\varepsilon|X]$ is a diagonal matrix, Campbell, Lo, and MacKinlay (1997) argue that each standardized *CAR* follows a Student t distribution, with $n-2$ degrees of freedom, thus having a variance of $\frac{n-2}{n-4}$. Summing the *CAR*s gives a statistic that has a variance of $\frac{N(n-2)}{n-4}$. The authors propose to standardize the statistic, $\sum_{j=1}^{N} CAR_j$, by its standard deviation. This would give a statistic that is approximately normally distributed. This approach, on the contrary to ours, does not take into consideration the cross-sectional correlation between firms, although the heteroskedasticity across firm is taken into account. The problem with clustering is that the standardized *CAR*s are not independent, and thus one cannot simply sum their variance.

The two statistics proposed here, $\tilde{t}_1^1$ and $\tilde{t}_1^2$, are standardized with the correct standard deviation (assuming clustering and no serial correlation). The degrees of freedom to apply is not known precisely, but the approximation to normality is probably better than in the case of the J_1 and J_2 statistics in Campbell, Lo, and MacKinlay (1997).

3.8 Hypothesis 2

Hypothesis 2 is that all *CAR*s are jointly equal to zero. In this case, the restriction matrix is $A' = I_N \otimes \iota_m'$. We then have:

$$
\begin{aligned}
A'\hat{\delta} &= \text{vec}\left(\iota_m'\hat{\Delta}I_N\right) \\
&= \text{vec}\left(\iota_m'\hat{\Delta}\right) \\
&= \hat{\Delta}'\iota_m \ .
\end{aligned}
$$

The variance of $A'\hat{\delta}$ is:

$$
\begin{aligned}
\text{V}\left[A'\hat{\delta}\middle| X\right] &= A'\text{V}\left[\hat{\delta}\middle| X\right]A \\
&= \left(I_N \otimes \iota_m'\right)\left(\Sigma_N \otimes M_{22}^{-1}\right)\left(I_N \otimes \iota_m\right) \\
&= \Sigma_N \otimes \iota_m' M_{22}^{-1}\iota_m \\
&= c\Sigma_N \ ,
\end{aligned}
$$

where $c = \iota_m' M_{22}^{-1}\iota_m$ (a scalar). The test statistic is the Wald statistic:

$$\begin{aligned}
\tilde{W}_2^1 &= \left(A'\hat{\delta}\right)' \left\{ \mathrm{V}\left[A'\hat{\delta} \middle| X \right] \right\}^{-1} \left(A'\hat{\delta}\right) \\
&= \iota_m' \hat{\Delta} \left(c\Sigma_N\right)^{-1} \hat{\Delta}' \iota_m \\
&= \frac{1}{c} \iota_m' \hat{\Delta} \Sigma_N^{-1} \hat{\Delta}' \iota_m \ .
\end{aligned}$$

Now for the J_2-type of test:

$$\begin{aligned}
A'(\psi \otimes \theta)\hat{\delta} &= \left(I_N \otimes \iota_m'\right) (\psi \otimes \theta)\hat{\delta} \\
&= (\psi \otimes \iota_m'\theta)\hat{\delta} \\
&= \mathrm{vec}\left(\iota_m'\theta\hat{\Delta}\psi\right) \\
&= \psi\hat{\Delta}'\theta\iota_m \ ,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{V}\left[A'(\psi \otimes \theta)\hat{\delta} \middle| X \right] &= \left(I_N \otimes \iota_m'\right) (\psi \otimes \theta) \mathrm{V}\left[\hat{\delta} \middle| X \right] (\psi \otimes \theta) \left(I_N \otimes \iota_m\right) \\
&= \left(I_N \otimes \iota_m'\right) (\psi \otimes \theta) \left(\Sigma_N \otimes M_{22}^{-1}\right) (\psi \otimes \theta) \left(I_N \otimes \iota_m\right) \\
&= \psi\Sigma_N\psi \otimes \iota_m'\theta M_{22}^{-1}\theta\iota_m \\
&= I_N \otimes \iota_m' I_m \iota_m \\
&= m I_N \ .
\end{aligned}$$

The Wald statistic is:

$$\begin{aligned}
\tilde{W}_2^2 &= \left[A'(\psi \otimes \theta)\hat{\delta}\right]' \left\{ \mathrm{V}\left[A'(\psi \otimes \theta)\hat{\delta} \middle| X \right] \right\}^{-1} A'(\psi \otimes \theta)\hat{\delta} \\
&= \left(\psi\hat{\Delta}'\theta\iota_m\right)' (mI_N)^{-1} \psi\hat{\Delta}'\theta\iota_m \\
&= \frac{1}{m} \iota_m'\theta\hat{\Delta}\psi\psi\hat{\Delta}'\theta\iota_m \\
&= \frac{1}{m} \iota_m'\theta\hat{\Delta}\Sigma_N^{-1}\hat{\Delta}'\theta\iota_m \ .
\end{aligned}$$

The test statistics derived in this subsection have no known distribution.

3.9 Hypothesis 3

The Wald statistic for this test uses a restriction matrix A that is simply an identity matrix of size Nm (note that A can be written as the Kronecker product $I_N \otimes I_m$).

The J_1-type of test for this hypothesis is:

$$A'\hat{\delta} = \hat{\delta} = \text{vec}\left(\hat{\Delta}\right) .$$

The variance of $A'\hat{\delta}$ is simply $\left(\Sigma_N \otimes M_{22}^{-1}\right)$. The Wald statistic is:

$$\begin{aligned} \tilde{W}_3^1 &= \hat{\delta}' \left(\Sigma_N \otimes M_{22}^{-1}\right)^{-1} \hat{\delta} \\ &= \hat{\delta}' \left(\Sigma_N^{-1} \otimes M_{22}\right) \hat{\delta} \\ &= \text{tr}\left(\hat{\Delta}' M_{22} \hat{\Delta} \Sigma_N^{-1}\right) . \end{aligned}$$

Finally, for the J_2-type test for hypothesis 3:

$$A'(\psi \otimes \theta)\hat{\delta} = (\psi \otimes \theta)\hat{\delta} ,$$

$$\begin{aligned} \text{V}\left[A'(\psi \otimes \theta)\hat{\delta}\middle| X\right] &= (\psi \otimes \theta)\text{V}\left[\hat{\delta}\middle| X\right](\psi \otimes \theta) \\ &= (\psi \otimes \theta)\left(\Sigma_N \otimes M_{22}^{-1}\right)(\psi \otimes \theta) \\ &= \psi \Sigma_N \psi \otimes \theta M_{22}^{-1} \theta \\ &= I_N \otimes I_m \\ &= I_{Nm} . \end{aligned}$$

And the Wald statistic becomes:

$$
\begin{aligned}
\tilde{W}_3^2 &= \left[A'(\psi \otimes \theta)\hat{\delta}\right]' \left\{\mathrm{V}\left[A'(\psi \otimes \theta)\hat{\delta}\middle| X\right]\right\}^{-1} A'(\psi \otimes \theta)\hat{\delta} \\
&= \hat{\delta}'(\psi \otimes \theta)(\psi \otimes \theta)\hat{\delta} \\
&= \hat{\delta}'\left(\Sigma_N^{-1} \otimes M_{22}\right)\hat{\delta} \\
&= \mathrm{tr}\left(\hat{\Delta}' M_{22}\hat{\Delta}\Sigma_N^{-1}\right) \ .
\end{aligned}
$$

We see that the Wald statistic for the J_1-type of test and the J_2-type of test are exactly equivalent. The weighting of the estimator by its standard deviation has no effect for this hypothesis.

The test statistic $\tilde{W}_3$ follows no known distribution, but a likelihood ratio test statistic for a this hypothesis is studied in Rao (1973) and is known to follow asymptotically an F distribution if the rank of $A'\hat{\delta}$ is less than equal to 2. This means that either the number of event period m or the number of firms N is equal or less than 2. See e.g. Chou (2004); Butler and Frost (1992).

To summarize all the results of this section, Table 1 shows the null hypothesis and the test statistics that we derived.

4 The bootstrap in event studies

In this section, we look at the bootstrap and its use in the event study literature. To our knowledge, the first application of the bootstrap method to event studies is in Marais (1984). The purpose of this application was to cope with non-normalities when there is limited observations in the estimation period (or using monthly data), but treats the case of single firm event studies. The methodology resembles another one by Chou (2004), which

Table 1: The three hypotheses and their test statistics

Hypothesis		**J_1- and J_2-type and Test statistic**
H_1:	The average *CAR* across firms is equal to zero	J_1: $\tilde{t}_1^1 = \dfrac{\iota_m'\hat{\Delta}\iota_N}{\sqrt{(\iota_N'\Sigma_N\iota_N)\times(\iota_m'M_{22}^{-1}\iota_m)}}$ J_2: $\tilde{t}_1^2 = \dfrac{\iota_m'\theta\hat{\Delta}\psi\iota_N}{\sqrt{Nm}}$
H_2:	All *CAR*s during the event period are jointly equal to zero	J_1: $\tilde{W}_2^1 = \dfrac{1}{c}\iota_m'\hat{\Delta}\Sigma_N^{-1}\hat{\Delta}'\iota_m$ J_2: $\tilde{W}_2^2 = \dfrac{1}{m}\iota_m'\theta\hat{\Delta}\Sigma_N^{-1}\hat{\Delta}'\theta\iota_m$
H_3:	All *AR*s are jointly equal to zero for all firms and all event periods	J_1 and J_2: $\tilde{W}_3 = \mathrm{tr}\left(\hat{\Delta}'M_{22}\hat{\Delta}\Sigma_N^{-1}\right)$

has the same motivation: to be robust in the case of non-normalities of return data but extends the method to the MVRM dummy variable method. Another extension is proposed by Hein and Westfall (2004). The authors claim that in addition to provide more precise tests in small sample of firms, their method is robust to time series correlation in returns. A third type of extension is provided by Kramer (2001), which treats the problem of small sample of firms (not observation period).

4.1 Marais, 1984

Marais (1984) studies the impact of small sample size and non-normality on a test similar to the Campbell, Lo, and MacKinlay (1997) J_2 test, but in the quadratic form (i.e. a Wald test statistic based on the standardized abnormal returns). The author recognizes that this test statistic is very sensitive to non-normalities of residuals, and he proposes a bootstrap method to remedy to this problem. His bootstrap method is a residual based bootstrap.

The steps for his bootstrap method are:

1. Estimate by OLS the regression equation $r_t = \alpha + \beta rm_t + \varepsilon_t$ for $t = 1, \ldots, n$. Obtain the parameter estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_\varepsilon^2$ and the residuals $\hat{\varepsilon}_t$;

2. Compute the test statistic;

3. Repeat a large number of times the following steps:

   (a) Select randomly with replacement $n + m$ observations from the couple $(\hat{\varepsilon}_t, rm_t)$, $t = 1, \ldots, n$;

   (b) Treat the OLS parameters $\hat{\alpha}$, $\hat{\beta}$ as if they were the true parameters and compute $n+m$ artificial returns from the residuals randomly chosen in step (a);

   (c) Estimate by OLS the market model parameters α^*, β^* and the abnormal returns using the artificial returns. Compute the artificial test statistic J^* or W^*;

4. Use the empirical distribution function of the bootstrap test statistics to determine whether the true test statistic is to be rejected or not.

The author studies the properties of his bootstrap method by a Monte Carlo experiment. The results are that the bootstrap helps to correct for the size of the test when data exhibits excess kurtosis. The rejection rate is closer to the specified size of the test when using the bootstrap than when using asymptotic distribution of the test statistic.

Note that this method assumes no time series dependences in the returns or the residuals because the resampling is done independently. This is called an i.i.d. bootstrap. Also, this method is used for a single firm. There is no indication how it can be extended to multiple firms with or without cross-sectional correlation. This last issue is the extension presented by Chou (2004).

4.2 Chou, 2004

The contribution of Chou (2004) is that the bootstrap method is adapted to the MVRM framework, and that a wider class of tests can be performed in an unbiased way. He studies three different null hypotheses:

1. The *AR*s are jointly equal to zero for all firms for event date j. This hypothesis is equivalent to our H_3 but with $m = 1$, i.e. only one event period.
2. The *AR*s are zero for all firms and all event periods, for $N = 1$ and $N = 2$. This hypothesis is equivalent to our H_3.
3. The mean *AR* for event period j is zero. This is equivalent to both our hypothesis H_1 and H_2, but with $m = 1$.

His bootstrap method is similar to Marais (1984) except that a covariance matrix of the residuals is computed (Σ_N) in each bootstrap iteration. The resampling is not done with the couple $(r_{i,t}, rm_t)$ but now with the $N + 1$-tuple $(r_{1,t}, \ldots, r_{N,t}, rm_t)$.

The author uses historical simulation following the method of Butler and Frost (1992). For hypothesis H_3 of abnormal return jointly zero for all firms and for a given event period, the existing asymptotic test is mis-specified (over-rejects). The bootstrap test slightly over-rejects for small N, but largely over-rejects for $N = 25$.

The second hypothesis is where the author tests whether all abnormal returns for all firms are jointly zero for $m = 2$, and for various number of firms. All tests, including bootstrap, over-reject the null hypothesis for a specified size and a specified N, although the bootstrap method performs slightly better.

All the tests for hypotheses H_1 and H_2 over-reject of null hypothesis, even for the bootstrap tests (in a lesser manner when N is small). We do not know whether this over-rejection is due to mis-specification of the tests in the framework of MVRM, or if it is due to non-normality and/or serial correlation of the data.

For hypothesis H_1 (similar to the J_1 test) and for $m = 1$, the specified size matches the rejection rate, both for the conventional tests and the bootstrap test. This suggests that non-normality is not an issue with a estimation period of size 200 ($n = 200$). It means also that the cross-sectional correlation issue is robustly taken into account in the MVRM framework. The fact that an event window of only one period is used rule out the mis-specification due to serial correlation. Here also, we do not know if serial correlation significantly bias the test statistics when more than one event period is studied.

In the next subsection, we will see a different approach of the bootstrap method for event study, namely the test statistic-based bootstrap (in the terminology of Hein and Westfall, 2004).

4.3 Kramer, 2001

Kramer (2001) provides a bootstrap method that is useful in the case when non-normalities affect the specification of the test statistic. The hypothesis under study is the sum of the N t statistics based on the single event period market model ($m = 1$). In the notation of Section 3, this the hypothesis H_1 for the J_2-type test. Asymptotic theory stipulates that a sum of N Student t distributed variables divided by the square root of N, what Kramer calls the Z statistic, is asymptotically distributed as standard normal:

$$Z = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N}} \sim N(0, 1) \ .$$

But, as Kramer argues, in samples where N is small and where returns are not normally distributed, this approximation might be weak and the test statistic might be biased.

Her solution to the problem is to take the N t statistics, to subtract the mean, and then to bootstrap them in order to obtain a relatively precise empirical distribution function. The rationale is that under the null, the t statistics have mean zero. This is why the method uses de-meaned t statistics in the bootstrap. The true distribution of the t statistic differs from the null distribution only from a shift parameter.

More formally, here are the steps of the bootstrap method:

1. Estimate the market model and obtain the abnormal returns.
2. Compute the t statistics t_i, $i = 1, \ldots, N$.
3. Compute the $\tilde{Z}$ statistics that are the standardized Z statistics:

   $$\tilde{Z} = \frac{Z}{\hat{\sigma}_N}$$

   where

   $$\hat{\sigma}_N = \sqrt{\frac{\sum_{i=1}^{N} (t_i - \bar{t})^2}{N - 1}}$$

   and $\bar{t}$ is the simple average of the t_i statistics.

4. Subtract the mean $\bar{t}$ from the t_i statistics:

$$t_i^* = t_i - \bar{t} \ , i = 1, \ldots, N \ .$$

5. Repeat a large number of times the following steps:

   (a) Draw a sample of size N from the t_i^* statistics. Call this sample $t_i^{*b}, i = 1, \ldots, N$.

   (b) Compute a pseudo $\tilde{Z}$ statistic with the t_i^{*b} statistics (just like in step 3). Call it $\tilde{Z}^b$.

6. Then, compute the empirical distribution function of $\tilde{Z}^b$ and decide to reject or not the statistic $\tilde{Z}$.

The results suggest that the method provides a good improvement relatively to the asymptotic distribution for small sample of firms (that is, relatively small; the algorithm does not work for $N = 2$ or 3). However, this bootstrap method is not robust to cross-sectional correlation between firms, as noted in Hein and Westfall (2004), because in this case the statistic Z or $\tilde{Z}$ is the sum of dependent t statistics, which does not converge to the normal distribution (see Hein and Westfall, 2004, Section 3 for a theoretical analysis of the cross-section correlation case).

4.4 Hein and Westfall, 2004

Hein and Westfall (2004) provide a comparison of several methods of testing hypothesis H_3 in the MVRM framework when the number of event period is 1 (m=1). They compare the classical parametric F test as developed in Binder (1985) and Schipper and Thompson (1985) to four types of bootstrap algorithm. The first algorithm is similar to Chou (2004) and tests the hypothesis H_3. The second is an algorithm called BT that is similar to the preceding one, but tests the hypothesis H_1 with the J_2 approach. The third is called BK and is like the Kramer (2001) bootstrap but extended for the MVRM framework. The last type of bootstrap is the one proposed by Kramer (2001) without adjustment for cross-sectional correlation.

To give more details, the second bootstrap algorithm is simply like the Chou (2004) bootstrap but the test statistic used is the (un-standardized) sum of the firms' t statistics. In other words, it is similar to the J_2 test for hypothesis H_1 except that the sum of the CARs is not multiplied by the scaling factor $\frac{N(n-2)}{n-4}$. The third bootstrap algorithm under study is very similar to this one except that the test statistic is the Z statistic of Kramer (2001) (this means that it is not a t statistic-based bootstrap, but rather a data-based bootstrap of the Z statistic).

The results are that all the inference methods have the right size in presence of even extreme cross-sectional correlation, except the method of Kramer (2001). They note that the first bootstrap method, similar to Chou (2004), is too conservative. In the presence of serial correlation, all testing methods have the right size, except in extreme cases of high AR(1) correlation, which is not likely with real data. Unfortunately, they only study the case of a single period event window, which is the case that is least sensitive to serial correlation. We do not know the effect of serial correlation when the event window expands to several days.

The authors provide also results for the study of power of the proposed tests, using historical simulations. Among the tests that have the right size, namely the Chou (2004) bootstrap, the BT and the BK bootstraps, the one with highest power is BT.

A drawback of the analysis of Hein and Westfall (2004) is that they compare inference methods that test different hypothesis. The hypothesis of the F test of Binder (1985) (hypothesis H_3 in our notation) is quite different than the one of Kramer, which is similar to the J_2 test of Campbell, Lo, and MacKinlay (1997), which we called J_2-H_1. These differences are most notable in the study of power, because their data generating process might simulate a case more easily identified by one of the tests and less by the other.

Before to study our test statistics with the bootstrap, we will have an empirical look on the distribution of residuals from the market model.

5 Descriptive statistics

It is well known in the finance literature that stock market returns are not normally distributed. But for event study analysis, the distribution that counts is not the distribution of the returns, but rather of the residuals. In this section, we give some descriptive statistics about the market model residuals.

We use 15 years of daily data (from 1990 to 2004) from 30 companies included in the Dow Jones Industrial Average index in the end of 2004. For this time period, we estimate the market model using the Dow Jones Industrial Average as market proxy. First, a test of normality using the Jarque-Bera test shows that residuals are far from normal. For all 30 firms, the normality of residuals can be rejected. Table 2, Panel A, gives an overview of the distribution of the Jarque-Bera statistics across firms.

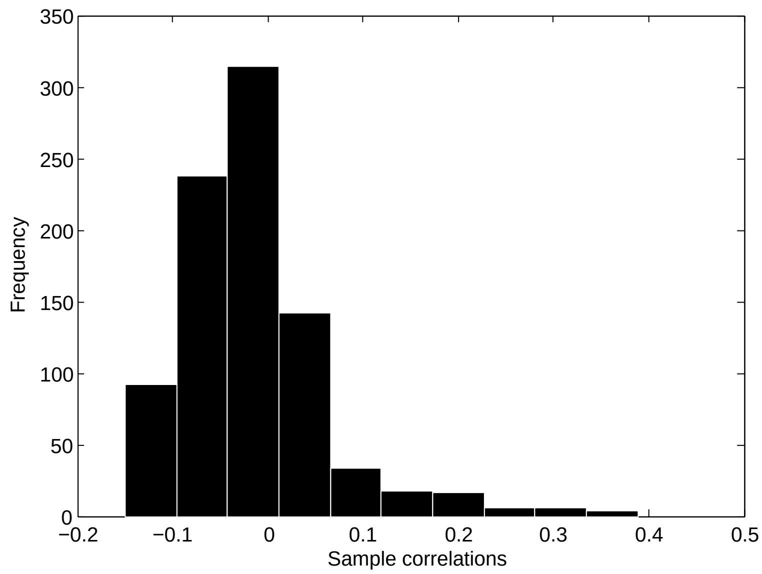
From the sample of 30 companies, we computed the skewness and kurtosis of each stock. In Table 2, we show the minimum, median, mean and maximum value of the statistics in the sample. We see that both positive and negative skewness are possible, but negative skewness is most likely. Concerning kurtosis, we see that all firms have residuals that exhibit fat tails. The mean kurtosis is already extremely high with a value of above 15.

As mentioned earlier, it is well known that the cross-sectional correlation of stock returns can be high, but it is less clear how large is the cross-sectional correlation of residuals. With a sample of $N = 30$ companies, the number of cross-sectional correlation coefficients is $\frac{N(N-1)}{2} = 435$. In the Figure 2, we show the histogram of these 435 correlation coefficients.

The time series correlation can also be quite varied across stocks. Testing for the statistical significance of the autocorrelations using the Ljung-Box test with five lags, we can reject at the 5% level that 25 out of 30 companies have regression residuals uncorrelated. For the remaining five companies, we cannot reject time series independence even at the 10% level. If we use the same test for ten lags, the number of companies for which we can reject the times series independence is 24 out of 30. Panel C of Table 2 shows the distribution of the Ljung-Box Q statistic.

Table 2, Panel C, gives an overview of the autocorrelations of the residuals by giving, for each lag, the minimum, the mean, the median and the maximum of the autocorrelation among the 30 companies. We see that the autocorrelations can be either positive or negative but are not high, ranging from -0.0974 (lag 2) to 0.0607 (lag 1). It is worthwhile noting that this

Figure 2: Histogram the the sample residual correlations



Note: This figure shows the histogram of all the values the the correlations of the 30 companies' residuals from the market model, for a total of 435 observations.

sample of firms is quite homogeneous, being all big companies from the Dow Jones Industrial Average index. Kramer (2001) argues that small companies and especially technological companies from the Nasdaq stock market usually have higher serial correlation.

Since event studies are performed with a selection of firms, depending on the event under study, the some firms under study can exhibit large cross-sectional correlation and departure from normal distribution that can be severe. But it is not likely that the residuals would show a high amount of autocorrelation.

6 Monte Carlo experiment

In this section, we test our five test statistics for the three hypotheses developed in Section 3. We use the Monte Carlo simulation technique. The aim of this experiment is to study whether the test statistics control the level of the test, and if they are robust to non-normalities and serial correlation. The test statistics are developed especially to take into account cross-sectional correlation of residuals in the case of clustering.

Test statistics for hypotheses J_1-H_1 and J_2-H_1 are studied with their parametric distribution as well as with the bootstrap method. For the two other hypotheses, the test statistics are studied here only with the bootstrap method because no parametric distribution is well specified in the case of many firms $N > 1$, multiple event periods $m > 1$ and cross-sectional correlation.

We study the case where residuals are normally distributed and where

Table 2: Descriptive statistics of the market model residuals

Panel A: Normality test: distribution of Jarque-Bera statistic

Statistic	minimum	mediam	mean	maximum
Jarque-Bera	621.4879	6535.9	67389	1203500
P-Value	0	0	0	0

Panel B: Distribution of third and fourth moment

Moment	minimum	mediam	mean	maximum
Skewness	-3.6782	-0.0672	-0.4038	0.3449
Kurtosis	4.9545	9.2943	15.2897	88.6344

Panel C: Time series independence test: distribution of Ljung-Box statistic

Statistic	minimum	mediam	mean	maximum
Ljung-Box (5 lags)	3.4865	15.8544	19.767	56.7087
P-Value	0.6254	0.0073	0.0014	0
Ljung-Box (10 lags)	9.9321	27.8047	28.8638	63.1941
P-Value	0.0772	0	0	0

Panel D: Distribution of autocorrelation values

Lag number	minimum	median	mean	maximum
0	1	1	1	1
1	-0.072	-0.0051	-0.01	0.0607
2	-0.0974	-0.0333	-0.0329	0.0153
3	-0.0638	-0.0204	-0.0195	0.0302
4	-0.0707	-0.0216	-0.0196	0.0369
5	-0.0356	-0.0043	-0.0027	0.0449

Note: The data used is the market model residuals for 30 stocks of the Dow Jones Industrials index from year 1990 to 2004. Panel A shows the distribution of the Jarque-Bera statistic for the 30 sets of residuals. "P-Value" is the probability value associated with the statistic. Panel B shows the distribution of skewness and kurtosis of the sets of residuals. Panel C shows the Ljung-Box Q statistics for time series independence. The first line shows the distribution of the statistics corresponding to the 30 sets of residuals for 5 lags. The third line shows the distribution of statistics for 10 lags. P-Value is the probability value associated to the above Q statistic. Panel D shows the distribution of the value of autocorrelation as a function of the lag.

they depart from normality by (1) non-zero skewness, (2) excess kurtosis, and (3) for both skewness and excess kurtosis. The data generating process uses the standard methodology for generating random normal distribution (i.e. generate first a uniform variable and use the inverse normal CDF). For the non-normal data, we use the same algorithm as Kramer (2001), namely the Lamda distribution.

For every test statistics, the bootstrap is done the same way: residual based bootstrapping and generating artificial data under the hull hypothesis. This method is proposed by Marais (1984) and adapted to the MVRM framework by Hein and Westfall (2004) and Chou (2004). Here are the steps:

1. Estimate the abnormal returns $\hat{\delta}$ and obtain the residuals $\hat{\varepsilon}$ for all firms and for time $t = 1, \ldots, n$ (the residuals for time $t = n+1, \ldots, T$ are automatically set to zero because of the dummy variables).
2. Estimate the correlation matrix $\hat{\Sigma}_N$ and compute the test statistic, say $\tilde{W}$.
3. Repeat a large number of times the following steps:
   (a) Draw with replacement a sample of T observations from the vector $(rm_t, \hat{E}_t)$, where $\hat{E}_t$ is the residuals of the N firms for time t, $t = 1, \ldots, n$. This represents an artificial sample of data from the null hypothesis of no event.
   (b) Estimate a bootstrap equivalent of the abnormal returns:
   $$\hat{\Delta}^* = M_{22}^{*-1} D' M^* \hat{E}^* ,$$
   as well as the covariance matrix $\hat{\Sigma}_N^* = \hat{E}^{*\prime} \hat{E}^* / (n-2)$ and the relevant test statistic, $\tilde{W}^*$.
4. With the empirical distribution function of the bootstrap test statistics, decide whether the null hypothesis is rejected of not.

The Monte Carlo simulations are all done with the same amount or cross-sectional correlation: 0.2. We used $N = 5$ firms and an event window of $m = 10$. The number of Monte Carlo repetitions is 1000 and the number of bootstrap repetitions is 200. The data generating process uses the following steps:

1. Generate a matrix $\tilde{E}$ of i.i.d. error terms of size $T \times N$ from the specified distribution (either normal or the Lamda distribution).
2. Post-multiply the matrix of error terms by a rotating matrix that confers cross-sectional correlation:

$$E = \tilde{E}\ \Sigma_N^{\frac{1}{2}} \ ,$$

   where $\Sigma_N^{\frac{1}{2}}$ is computed by the Cholesky decomposition of the theoretical covariance matrix, Σ, of the specified structure of cross-sectional correlation.
3. For the serial correlation case: add serial correlation to every column of E, by applying an autoregressive filter of order 1, with parameter ρ.
4. Generate T observations of the market return rm_t, from the standard normal distribution.
5. Form the matrix of firms' returns[4]:

$$R = \iota_T \iota_N' + rm\ \iota_N' + E \ .$$

We performed a series of four Monte Carlo simulations (called Panel A to Panel D). In every panel, we studied the rejection rate of the test statistic with nominal size of 5%. The tests that are studied are: (1) both parametric tests for hypothesis 1 (similar to Campbell, Lo, and MacKinlay (1997) statistic J_1 and J_2, what we have called $\tilde{t}_1^1$ and $\tilde{t}_1^2$ above). These

[4]Note that this is equivalent to forming firms' returns that all have unit variance, a β of 1 and an α of 1. As noted in Hein and Westfall (2004), the distribution of the test statistic depends on the distribution of the error terms, not on the distribution of the market returns or on the coefficients.

test statistics follow asymptotically the Student t distribution with $n - 2$ degrees of freedom and the standard normal distribution, respectively. (2) The bootstrap counterpart to the same hypothesis H_1 and the same test statistics $\tilde{t}_1^1$ and $\tilde{t}_1^2$. (3) The bootstrap test for hypothesis H_2 using the test statistics $\tilde{W}_2^1$ and $\tilde{W}_2^2$. (4) The bootstrap test for hypothesis H_3 using $\tilde{W}_3$ as test statistic.

Tables 3 and 4 show the results of the Monte Carlo simulations. For normally distributed residuals and no serial correlation (Panel A), most tests are well specified, including the parametric tests, except the bootstrap version of hypothesis 1, which seems to under reject the null hypothesis (this is consistent with the results in Hein and Westfall (2004)).

For non normal residuals (Panel B), we see that the parametric tests of hypothesis H_1 are still well specified. This suggests that a sample of $n = 200$ observations in the estimation window is sufficient for these tests to converge to their specified distributions. The bootstrap tests have rejection rates similar to the case of normal residuals, but hypothesis 1, the test still under rejects the null. For the other tests, the bootstrap performs quite well with non-normalities.

For the case of serial correlation of residuals (Panel C), two values of the autocorrelation parameters are studied: $\rho = 0.1$ and $\rho = 0.2$. Both parametric tests for hypothesis 1 over reject the null proportionally to ρ. For the bootstrap tests, hypothesis 1 and 3 slightly over reject with with small autocorrelation of 0.1. Tests for hypothesis 2 are biased by far. For larger autocorrelations, we see that the bootstrap version of H_1 have the correct level. Tests for hypothesis H_2 clearly over reject, while tests for hypothesis H_3 slightly over rejects.

Finally, Panel D shows results for a DGP that feature both non-normalities and serial correlation. The parametric tests over reject to a small extent, but are still of interest for all practical purposes. The bootstrap tests for hypothesis 1 are just well specified. But this might be misleading since Panel A showed that the test is biased even for i.i.d normal residuals. Just like in Panel C, tests for hypothesis 2 over reject to a large extent and test for hypothesis 3 slightly over reject.

Interpretation

For hypothesis 1, the best test is the parametric test, either J_1 or J_2. With an estimation window of 200 observations, the test is correctly specified event when the number of firms is relatively small (here it is 5). Campbell, Lo, and MacKinlay (1997) suggested that the distribution of the J_2 statistic was asymptotic in N. In the context of MVRM, this does not seems to be the case, at least in our Monte Carlo experiment.

For hypothesis 2, our Monte Carlo results tell us that the test is well specified only when there is no serial correlation in the residuals. For relatively large serial correlation of 0.2, the rejection rate more than four times the theoretical level of the test. In practice, one should verify empirically the presence of serial correlation before to test this hypothesis with the bootstrap procedure that we proposed. If there is indeed presence of autocorrelation, a possible alternative would be to model explicitely this autocorrelation (by ARMA models, for instance), or to use a benchmark model that contains more factors (i.e. a better suited asset pricing model). This could take away the serial correlation. In any case, researchers should test this hypothesis

Table 3: Monte Carlo rejection rates: no serial correlation

Panel A: Normal distribution, no serial correlation

DGP	Parametric tests		Bootstrap tests				
	H_1		H_1		H_2		H_3
	J_1	J_2	J_1	J_2	J_1	J_2	--
N(0,1)	0.053	0.059	0.027	0.027	0.058	0.059	0.062
	0.336	*0.1135*	*0*	*0*	*0.1396*	*0.1135*	*0.0578*

Panel B: Non-normal distribution, no serial correlation

DGP	Parametric tests		Bootstrap tests				
	H_1		H_1		H_2		H_3
	J_1	J_2	J_1	J_2	J_1	J_2	--
S	0.055	0.059	0.035	0.037	0.06	0.058	0.072
	0.244	*0.1135*	*0.0049*	*0.0147*	*0.0915*	*0.1396*	*0.0036*
K	0.065	0.073	0.021	0.025	0.061	0.059	0.072
	0.0272	*0.0026*	*0*	*0*	*0.0731*	*0.1135*	*0.0036*
S + K	0.044	0.052	0.027	0.026	0.06	0.058	0.084
	0.1775	*0.3879*	*0*	*0*	*0.0915*	*0.1396*	*0.0001*

Note: This table shows the rejection rates obtained from Monte Carlo experiments produced with 1000 repetitions. Event study parameters are: 200 observations in the estimation window; 10 periods for the event window; 5 firms that have equal variance of the residuals and cross-sectional correlation of 0.2. Two tests are done with the parametric distribution of the test statistic. Five tests are done using the bootstrap method. H_x means that hypothesis x is tested. J_1 and J_2 mean that the test statistic has the form corresponding to the J_1 of J_2 test in Campbell, Lo, and MacKinlay (1997). "DGP" is for Data Generating Process. The p-values of the rejection rate being different from the nominal level of 5% are shown in italic. The variance of the Monte Carlo estimates of the rejection rates is $\alpha(1-\alpha)/MC$ where α is the estimated rejection rate and MC is the number of Monte Carlo repetitions. "S" means that the data depart from normality by a skewness of 0.75. "K" means that the distribution of the residuals have a kurtosis of 8.

Table 4: Monte Carlo rejection rates: with serial correlation

Panel C: Normal distribution, with serial correlation

DGP	Parametric tests		Bootstrap tests				
	H_1		H_1		H_2		H_3
	J_1	J_2	J_1	J_2	J_1	J_2	--
ρ = 0.1	0.097	0.096	0.06	0.062	0.174	0.175	0.067
	0	*0*	*0.0915*	*0.0578*	*0*	*0*	*0.0158*
ρ = 0.2	0.102	0.109	0.056	0.054	0.206	0.204	0.084
	0	*0*	*0.2046*	*0.2879*	*0*	*0*	*0.0001*

Panel D: Non-normal distribution, with serial correlation

DGP	Parametric tests		Bootstrap tests				
	H_1		H_1		H_2		H_3
	J_1	J_2	J_1	J_2	J_1	J_2	--
S+K+ρ=0.2	0.084	0.084	0.05	0.05	0.18	0.181	0.088
	0.0001	*0.0001*	*0.5*	*0.5*	*0*	*0*	*0*

Note: This table shows the rejection rates obtained from Monte Carlo experiments. See note from Table 3 for details. "$\rho = x$" means that the residuals are generated from an autoregressive model of order 1 with coefficient x.

with great care.

The test for hypothesis 3 seems to react quite well to non normalities or to serial correlation or both. Our Monte Carlo experiments showed that the test slightly over reject the null, but to a small extent. In practical situations, this test could be used, although care should be taken if the null hypothesis is close to be rejected.

Comparison of these results with previous ones in the literature teaches us that the exact null hypothesis is important when performing event studies. The bootstrap methods that are studied in Hein and Westfall (2004) have different results (basically, the Kramer bootstrap over-rejects the null in the presence of cross-sectional correlation, while their HWZ bootstrap is well specified). The reason might be that they test different hypotheses more than because of intrinsic characteristics of the bootstrap algorithm. Namely, the Kramer bootstrap is equivalent to our $H_1 - J_2$ hypothesis while the HWZ bootstrap tests what we call the H_3 hypothesis.

On one hand, having a unique event period makes the test statistic very sensitive to non normalities, but it makes it robust to serial correlation. The contrary is also true: a large event window makes the test robust to non normalities but it causes bias in the case of serial correlation.

7 Conclusion

The contribution of this thesis is two-fold. First we provide a unified methodology for estimating and testing the abnormal returns in event studies when (1) many firms cluster in time and (2) the event window spans many days.

Our methodology also unifies the different hypotheses that are usually tested in the literature with the two approaches that are exposed in Campbell, Lo, and MacKinlay (1997): the J_1 approach that tests whether the sum of abnormal returns is equal to zero, and the J_2 approach that takes into account the heterogeneity between firms and tests whether a weighted average of the abnormal returns is equal to zero.

We separate the relevant hypotheses into three groups. H_1 tests whether the sum (or the weighted sum) of all abnormal returns across firms and event periods is equal to zero. The second hypothesis H_2 tests whether the cumulative abnormal returns of the firms are jointly equal to zero. The third hypothesis H_3 tests whether all abnormal returns in all event periods are jointly equal to zero for all firms.

We also propose a bootstrap algorithm that is not new, but that unifies the existing bootstrap methods proposed in the literature. We chose to use this bootstrapping technique to verify the claims in the literature that it is robust to serial correlation as well as to non-normalities of the data. We explain how to use this bootstrap procedure for inference in the case where many firms cluster in time and when the event window has more than one observation.

Monte Carlo simulations are used to assess the error type I of the tests that we propose. We used a data generating process that can incorporates non-normalities and serial correlation in a manner that is similar to real financial data.

Our results show that for practical purposes, the parametric J_1 and J_2 tests of Campbell, Lo, and MacKinlay (1997)(used to test hypothesis 1), when adapted for clustering, are well specified even in the presence of non-normalities or serial-correlation, even when the number of firms is relatively small (5).

Hypothesis H_2 is unbiased in the case of non-normalities, but not in the presence of serial correlation of the residuals. In order to test this hypothesis, the researcher has to make sure that the residuals in his sample do not suffer from serial correlation. For the third hypothesis, our results show that the test is slightly biased, but to a small extent.

This unified approach to event studies opens the way to wider research on robust inference methods. For instance, one could use more sophisticated bootstrap techniques that are more suited for time series data (e.g. the block bootstrap). Alternatively, one could use a covariance matrix that is robust to heteroskedasticity and autocorrelation (HAC covariance matrix), but probably to the price of lower power. Another alternative would be to study whether using a more complex benchmark model could improve inference in event studies.

Future research could also focus on the analysis of the power of the proposed inference methods. Since the three different hypotheses answer different research questions, the analyses should be specific to each hypothesis.

References

Binder, John J., 1985, On the use of the multivariate regression model in event studies, *Journal of Accounting Research* 23, 370–383.

——— , 1998, The event study methodology since 1969, *Review of Quantitative Finance and Accounting* 11, 111–137.

Butler, Ronald W., and Carol A. Frost, 1992, An analysis of "exactly distributed" test statistics for multivariate event tests, *Review of Quantitative Finance and Accounting* 2, 327–341.

Campbell, John Y., Andrew W. Lo, and Craig A. MacKinlay, 1997, *The econometrics of financial markets* (Princeton University Press: Princeton, N.J.).

Chou, Pin-Huang, 2004, Bootstrap tests for multivariate event studies, *Review of Quantitative Finance and Accounting* 23, 275–290.

Davidson, Russell, and James G. MacKinnon, 2004, *Econometric Theory and Methods* (Oxford University Press: New-York, N.Y.).

de Roon, Frans, and Chris Veld, 1998, Announcement effects of convertible bond loans and warrant-bond loans: An empirical analysis for the dutch market, *Journal of Banking and Finance* 22, 1481–1506.

Efron, B., 1979, Bootstrap methods: Another look at the jackknife, *The Annals of Statistics* 7, 1–26.

Fama, Eugene F., Lawrence Fisher, Michael C. Jensen, and Richard Roll, 1969, The adjustment of stock prices to new information, *International Economic Review* 10, 1–21.

Greene, William H., 2003, *Econometric Analysis* (Prentice Hall: Upper Saddle River, N.J.) 5 edn.

Hein, Scott E., and Peter Westfall, 2004, Improving tests of abnormal returns by bootstrapping the multivariate regression model with event parameters, *Journal of Financial Econometrics* 2, 451–471.

Horowitz, Joel L., 2001, The bootstrap, in James J. Heckman, and Edward E. Leamer, ed.: *Handbook of Econometrics*, vol. 5, pp. 3159–3228 (Elsevier: North Holland).

Karafiath, Imre, 1988, Unsing dummy variables in the event methodology, *Financial Review* 23, 351–357.

Khotari, S. P., and Jerold B. Warner, 2007, Econometrics of event studies, in B. Espen Eckbo, ed.: *Handbook of Corporate Finance*, vol. 1: Empirical Corporate Finance (Elsevier: North-Holland).

Kramer, Lisa A., 2001, Alternative methods for robust analysis in event study applications, in Cheng-Few Lee, ed.: *Advances in Investment Analysis and Portfolio Management*, vol. 8, pp. 109–132 (Elsevier: North-Holland).

Marais, M. Laurentius, 1984, An application of the bootstrap method to the analysis of squared, standardized market model prediction errors, *Journal of Accounting Research* 22 Special issue: Studies on Current Econometric Issues in Accounting Research, 34–54.

Rao, C. Radhakrishna, 1973, *Linear Statistical Inference and its Applications* (Wiley and Sons: New-York, N.Y.) 2 edn.

Schipper, Katherine, and Rex Thompson, 1983, The impact of merger-related regulations on the shareholders of acquiring firms, *Journal of Accounting Research* 21, 184–221.

——— , 1985, The impact of merger-related regulations using exact distributions of test statistics, *Journal of Accounting Research* 23, 408–415.