Measuring Security Price Performance Using Chilean Daily Stock Returns: The Event Study Method

Rodrigo Saens N.[1], y Eduardo Sandoval A.[2]

[1]Universidad de Talca, Chile. E-mail: rsaens@utalca.cl. [2]Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Aguas Calientes, México. E-mail: esandov@itesm.mx

ABSTRACT. Following the Brown-Warner simulation approach and using Chilean daily security return data we examine the specification and power of three parametric t-tests commonly employed in event studies: the standardized, the cross-sectional and the portfolio t-test. Our findings show that although individual security returns and security abnormal returns are evidently non-normal, the cross-sectional mean abnormal returns converge to normality as the sample size increases. Thus, in event study setting involving event period of one day, methods based on the use of parametric t-tests seem to be well specified, at least at the 5% significance level. In terms of power, our simulation results show the standardized t-test always more likely to detect the presence of an abnormal return that its two parametric competitors: the cross-sectional and the portfolio t-test. We also find, however, that the power of the three t-tests is very sensitive to either the sample size or the length of the event period.

Key words: event studies, specification and power of the test

JEL Classification: G14

Midiendo retornos diarios en el mercado accionario chileno: el método de estudio de eventos

RESUMEN. Siguiendo el enfoque de simulación de Brown y Warner y usando retornos diarios del mercado accionario chileno, examinamos la especificación y el poder de tres estadísticos comúnmente utilizados en estudios de evento: el test estandarizado, el de corte transversal y el de portfolio. Nuestros resultados muestran que aunque los retornos y excesos de retornos a nivel individual evidentemente no distribuyen normal, la media muestral converge hacia la normalidad en la medida que el número de acciones del portfolio muestral aumenta. Así, las pruebas estadísticas típicamente utilizadas en estudios de evento de un día estarían bien especificadas, al menos para un nivel de significancia del 5%. En términos del poder del test, el test estandarizado siempre se muestra más poderoso para capturar la presencia de un retorno anormal que sus dos competidores: el test de corte transversal y el de portfolio. También encontramos, sin embargo, que el poder de las tres pruebas estadísticas analizadas es muy sensible tanto al tamaño muestral como al número de días que involucre el evento.

Palabras clave: estudio de eventos, especificación y poder del test

Clasificación JEL: G14

(Recepción de originales: 31 de enero de 2005. Aceptado para publicación: 14 de marzo de 2005)

I. Introduction

Over the last twenty years, the performance of the events-study methodology has been the subject of a number of studies. The main concern of this research is to examine the power and the degree of specification of test statistics used in short-run and long-run event studies. Brown and Warner (1985), Dyckman, Philbrick and Stephan (1984), Campbell and Wasley (1993) and Cowan and Sergeant (1996) analyze how the particular properties of daily stock returns affect the performance of several test statistics employed in short-run event studies. On the other hand, Barber and Lyon (1997), Kothari and Warner (1997), Brav (2000) and Jegadeesh and Karceski (2004) examine the performance of alternative test statistics used in long-horizon event studies. All these studies, however, assess the specification and power of test statistics using daily stock return data drawn from developed equity markets.

As Campbell and Wasley (1993) point out, normality of abnormal returns is a key assumption underlying the use of parametric test statistics in the event-study method. Brown and Warner (1985) and Dyckman *et al.* (1984) study the effect of nonnormality in daily return data on tests performance using samples of randomly selected New York Stock Exchange (NYSE) and American Stock Exchange (AMEX) securities. They report that the nonnormality problem has not a substantive impact on the power of the short-run events study methodology and that the common parametric t-test used in these studies is well specified under the null hypothesis. They also indicate that the parametric t-test has a statistical power comparable to the theoretical power obtained under the normality assumption. However, Campbell and Wasley (1993) and Maynes and Rumsey (1993) find that with thinly traded samples the conventional standardized and portfolio t-tests are poorly specified. They report that these parametric tests reject a true null hypothesis too often with NASDAQ and Toronto Stock Exchange samples, respectively. Moreover, Cowan and Sergeant (1996) also report a similar misspecification of the portfolio t-test for thinly traded samples using NYSE-AMEX and NASDAQ daily stock returns files.

Extension of this research to Latin American security returns data is not a clear-cut. Given that the usefulness of the event study method is directly related to the market's ability to quickly reflect new information, the particular thin trading of the Latin American stock market can have a substantial impact on statistical properties of daily security returns. In fact, although empirical research using event study method is increasingly incorporating Latin American security daily returns data, some characteristics of these securities clearly differ from those traded in equity markets of developed countries. [1]

This paper examines statistical properties of daily stock returns and how the particular characteristics of these data affect the empirical performance of the short-run events-study methodology when security returns data are drawn from a Latin American equity market: the case of Chile. Thus, our study has two main objectives: First, we examine normality in the actual distribution of daily security returns and daily excess security returns drawn from the Chilean equity market. Given the more extensive thin trading we observe in Latin American equity markets, it is reasonable to expect a more severe degree of nonnormality in the distribution of these security excess returns than those found by previous authors in NYSE-AMEX daily excess returns. Second, we analyze the performance of the events-study method conducted in the Chilean equity market. We address three issues that determine the capacity of an event study to detect abnormal returns: The portfolio size, the magnitude of an eventual abnormal performance and the event date uncertainty. We examine the interaction over ranges of all these three variables simultaneously to determine their effect on the researcher's capacity to identify abnormal performance when event studies are conducted in thinly traded markets.

The examination is conducted using a simulation approach analogous to that introduced by Brown and Warner (1980). Unlike a Monte Carlo simulation where the researcher samples artificially generated values from a specified theoretical probability distribution, the Brown-Warner approach randomly selects event dates and stocks to simulate event studies without assuming a particular distribution of stock returns. Our contribution attempts to help select statistic tests, reducing the probability of misspecification and increasing the power of tests when studies involve Latin American equity market securities. Although this

[1] For example, recently Bhattacharya *et al.* (2000) document a significant negative (positive) return bias on the good (bad) news announcement explained by a severe insider trading in the Mexican stock market. They conclude that this problem in event studies can bias the researcher toward falsely conclude that corporate news announcements are a non-event.

technique overcomes the theoretical question, it allows us to examine the statistical validation of different alternative methods. After all, as Brown and Warner (1980, p.210) indicate, "…the performance of alternative models (in event studies) is an empirical question."

II. The Issue

As McWilliams and Siegel (1997) point out, an important assumption underlying the use of parametric t-tests in the events study methodology is normality of excess returns. Fama (1976, p.21), conversely, documents evidence that the distributions of daily returns exhibit substantial departures from normality, suggesting that they are *fat-tailed* relative to a normal distribution. Brown and Warner (1985) support the same result for the case of NYSE-AMEX daily excess returns. They document that daily returns depart considerably from normality in term of skewness and kurtosis. Additionally, Cowan (1992), Campbell and Wasley (1993), and Cowan and Sergeant (1996) show that this is also the case for NASDAQ daily excess returns. Even though these findings are not consistent with the normality assumption in excess returns, Dyckman *et al.* (1984) and Brown and Warner (1985) report that the degree of nonnormality in daily NYSE security excess returns does not represent a serious problem for a correct test specification. They also show that the portfolio and the standardized t-tests have an empirical power comparable to the theoretical power obtained under the normality assumption. This result is based on the Central Limit Theorem that guarantees that if the excess returns in the cross-section of securities are independent and identically distributed the distributions of the sample mean excess return will asymptotically converge to a normal distribution.

Extension of these findings to a Latin American stock market, however, is not clear. As Urrutia (1995) and Rouwenhorst (1999) indicate, Latin American stock markets have higher average ex-post returns but, at the same time, their number of listed companies, market capitalization, amounts traded, and level of integration are relatively small. Thus, it is reasonable to expect a more severe degree of infrequent trading and nonnormality in the distribution of security excess returns. [2]

Campbell and Wasley (1993) and Cowan and Sergeant (1996) show that in markets with thinner trading there is a significant degree of nonnormality in the daily returns securities that persists even at the portfolio level. As a result, the t-statistics used in event studies depart from their theoretical unit normal distribution under the null hypothesis. This can be also the case of t-statistics used in the Chilean stock market.

III. Experimental Design

As Brown and Warner (1980) and Dyckman *et al.* (1984) argue, given the problems of using an analytical approach to compare different properties of alternative return-generating models (RGM), the simulation approach provides a useful method for dealing with conditions where either the analytical approach becomes extraordinarily difficult or the same approach yields results suggesting just directions but not magnitudes. In this paper we resemble that positive approach of Brown and Warner (1980, 1985) to analyze the specification and statistical power of three different RGM when event studies are conducted using samples from the Chilean stock market.

3.1 Abnormal Returns

An event study attempts to measure the effect of an observed event on the firm market value. In general, the main purpose of any event study is to find empirical evidence that shows whether a security performance is statistically different from what would be expected under the assumptions of one specific RGM. As MacKinlay (1997, p.13) indicates, "the usefulness of such a study comes from the fact that, assuming rationality in the market place, the effect of an event will be reflected immediately in assets

[2] Empirical research using event study method is increasingly incorporating Latin American security daily returns data which are now available from ECONOMATICA.™ Some examples of recent event studies conducted with Latin American stock returns are Castillo (2005), Morán (2003), Bhattacharya *et al*. (2000), Parisi and Pérez (2000), Saens (1999) and Celis and Maturana (1998).

prices." Thus, if the event conveys new –relevant– information to the market place, the mean or the variance of the security excess returns must reflect the new economic conditions.

For firm i and event date t the conditional *abnormal* return is given by:

$$AR_{it} = R_{it} - E(R_{it} / \Omega_{t-1}) \qquad (1)$$

Where AR_{it}, R_{it} and $E(R_{it}/\Omega_{t-1})$ are the abnormal, actual and normal (expected) return for time t, respectively. Notice that Ω_t is the conditional information set in period t and that the approach followed for the event study methodology assumes that securities returns are generated by some RGM.

Then, it is necessary to specify a model that generates *normal* returns before abnormal returns can be measured. This model can be based on simple statistical relationship as the market model or on more theoretical economic models as the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Model (APT).[3]

We report abnormal performance measures based on the three RGM

A. OLS Market Model

$$AR_{it} = R_{it} - \hat{\alpha} - \hat{\beta}_i R_{mt} \qquad (2)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are OLS values from the estimation period.

B. Market-Adjusted Returns Model

$$AR_{it} = R_{it} - R_{mt} \qquad (3)$$

where R_{mt} is the market index return for day t.

C. Mean-Adjusted Returns Model

[3] Unfortunately, obtaining a more accurate model of the process generating actual returns is not a sufficient condition for that model to generate a well specified and powerful test of abnormal return. First, as Brown and Warner (1980) indicate, there is a measurement error in each of the variables on which returns depend in the model. For example, in the case of the CAPM, as Roll (1977) argues, it is not possible to observe directly the market portfolio. Second, the efficiency of using either a statistical or an economic model depends seriously on the additional statistic assumptions about ε_t, the error term. If the assumed sampling distribution under the null hypothesis is incorrect we are exposed to obtain false inferences.

$$AR_{it} = R_{it} - \overline{R}_i \tag{4}$$

where $\overline{R}_i$ the simple average of security i`s daily returns in the estimation period.

These three RGM are discussed in Brown and Warner (1980) and MacKinlay (1997).

3.2 Test Statistics

The test statistics for day 0 analyzes whether or not the portfolio mean excess return in day 0 is equal to zero. We study the specification and power of three parametric t-tests: The standardized, the cross-sectional and the portfolio t-test.

A. The Standardized t-test (θ_1)

Following Patell (1976) and Dodd and Warner (1983) many authors use a standardized abnormal return (SAR) where each abnormal security return is normalized by its estimation period standard deviation:

$$SAR_{it} = \frac{AR_{it}}{SD(AR_{it})} \tag{5}$$

The standard deviation $SD(AR_{it})$ of each abnormal return is given by:

$$SD(AR_{it}) = \sqrt{\frac{1}{T_0 - 1} \sum_{t=1}^{T_0} AR_{it}^2} \tag{6}$$

Where T_0 is the number of days in the estimation period. Thus, the day 0 of the standardized t-test is:

$$\theta_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} SAR_{i0} \tag{7}$$

The standardized t-test assumes that the individual abnormal returns are cross-sectionally independent and identically distributed. By the Central Limit Theorem, the standardized t-test converges to unit normal under the null hypothesis of no abnormal return. Brown and Warner (1985) report that this test is well specified under the null hypothesis for NYSE-AMEX daily security returns data. However, Campbell and Wasley (1993) and Cowan and Sergeant (1996) document that the standardized test is misspecified for NASDAQ samples.

B. *The Cross-Sectional t test (θ_2)*

As the standardized t-test, this method also assumes that the day 0 abnormal returns are independent and identically distributed. The t-statistic is estimated by dividing the average event-period abnormal return ($\overline{AR}_0$) by its contemporaneous cross-sectional standard deviation.

$$\theta_2 = \frac{\overline{AR}_0}{SD(\overline{AR}_0)} \tag{8}$$

The cross-sectional test ignores the estimation period variance and the standard deviation $SD(\overline{AR}_0)$ is given by:

$$SD(\overline{AR}_0) = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^{N} (AR_{i0} - \overline{AR}_0)^2} \tag{9}$$

This procedure, however, has some limitations. If the variance differs across sample securities or security abnormal performances are correlated across firms the test statistic is likely to be misspecified.

C. *The Portfolio t-test (θ_3)*

For each day *t*, the cross-sectional average excess return of *N* securities is computed. The portfolio *t-test* is the ratio of the mean excess return in $t=0$ to its estimated standard deviation:

$$\theta_3 = \frac{\overline{AR}_0}{SD(\overline{AR}_t)} \tag{10}$$

Where for each day the cross-section average excess return of *N* securities is obtained as:

$$\overline{AR}_t = \frac{1}{N} \sum_{i=1}^{N} AR_{it} \tag{11}$$

And the standard deviation is computed over an estimation period of T_0 days. Thus,

$$SD(\overline{AR}_t) = \sqrt{\frac{1}{T_0 - 1} \sum_{t=1}^{-T_0} (\overline{AR}_t - \overline{\overline{AR}})^2} \tag{12}$$

$$\overline{\overline{AR}} = \frac{1}{T_0} \sum_{t=1}^{-T_0} \overline{AR}_t \tag{13}$$

If $\overline{AR_t}$ are independent, identically, and normally distributed, the test statistic is distributed t-student with (T_0-1) degrees of freedom and is asymptotically unit normal under the null hypothesis. Brown and Warner (1980) call this method "Crude Dependence Adjustment" because, according to this test, the standard deviation of the day 0 average excess return is estimated from the values of the mean excess returns using for each security the estimation period. Thus, the portfolio t-test explicitly takes into account any potential cross-sectional dependence in the security specific excess returns. However, Campbell and Wasley (1993) find that, although less pervasive than in the standardized t-test, misspecification is also present in the portfolio t-test when event studies include thinly traded samples.

3.3 Data and Sampling Procedure

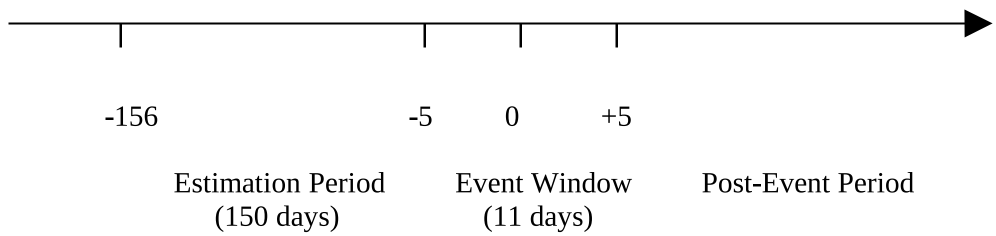
The data come from daily closing prices series for stocks traded in the Santiago Stock Exchange from January 1985 to July 2003. As market proxy we use the domestic index of stock prices IPSA, which includes the 40 most traded stocks in the Chilean market.

Series of 161 observations –trading and not trading days– are drawn randomly with replacement to conform portfolios of 10, 20, 30, 40 and 50 series each. Event dates are assumed to take place with equal probability on each trading day from 01/02/1985 to 07/30/2003. Each stock series is selected by generating two random numbers; the first number identifies a row –a date– and the second number a column –a stock– over the feasible database. With these directions we obtain the initial observation of a 161 daily observations series. This process is repeated over and over again until conform one portfolio.

The period -156 through -6 of each series (150 days) is the estimation period, in which the parameters of expected return models are estimated.[4] The period -5 through +5 (11 days) and the day 0 are

[4] As Kothari (2001) and Kothari and Warner (2005) indicate, the length of the estimation period is arbitrary. This has to be long enough to contain a "reasonable" number of observations to estimate the parameters of the model and short enough to avoid an eventual instability of these parameters. In general, the literature uses a length between 120 days (Dyckman *et al*, 1984) and 250 days (MacKinlay, 1997).

the event period and the event date. In order to include a security in a sample it must have at least 40 daily returns in the entire 161 days period, and no missing return data in the last 10 days.



3.4 Simulating Abnormal Performance

In order to artificially introduce a given level of abnormal return we follow the Brown-Warner procedure. A constant sample-wide abnormal performance –from 0% to 2.5%– is added to the actual day 0 return for each security. For example, to simulate 1.5% abnormal return, 0.015 is added. This procedure allows us to analyze the power of test statistics for different abnormal returns level.

IV. Results

4.1. Time Series Properties for Individual Chilean Securities

Table 1 documents some statistical properties of daily returns and daily abnormal returns for individual securities selected with replacement from the Chilean stock market. Parameter estimates are computed based on 500 samples of 50 securities, randomly selected. Using the time series of estimation period data, we calculate the mean, standard deviation, skewness and kurtosis coefficients, and the studentized range. Each value on table 1 represents the mean of 25,000 estimates. In the case of daily abnormal returns, they are based on the three different RGM we examine in this paper.

Results on table 1 indicate that for the case of the Chilean stock market daily returns and daily abnormal returns of individual securities depart significantly from the theoretical normal distribution. For example, mean values for the skewness and kurtosis coefficients equal or exceed 0.30 and 6.38, respectively. Additionally, the studentized ranges are 6.93 or greater. All the skewness and kurtosis coefficients and the studentized range for the daily returns and daily abnormal returns exceed the 99th percentile of the respective distribution under the normality hypothesis. Departures from normality are comparable to those documented in Brown and Warner (1985, table 1) but less severe than those showed by Campbell and Wasley (1993, table 1) for individual NYSE and NASDAQ daily security returns, respectively.[5]

Table 1

Properties of daily returns and daily excess returns for individual Chilean securities when no abnormal performance is introduced. For each security parameter estimates are based on time-series data in the estimation period. Each number in the table shows the mean of 25.000 estimates. Securities and event dates are randomly selected (with replacement) from 02/01/1985 through 07/31/2003.

Performance Measure	Mean	Standard Deviation	Skewness	Kurtosis	Studentized Range
Returns	0.0014	0.0225	0.40	6.92	7.08
OLS Market Model	-0.0001	0.0206	0.32	6.55	6.98
Market-Adjusted	0.0003	0.0198	0.30	6.38	6.93
Mean- Adjusted	0.0000	0.0225	0.40	6.92	7.08

Table 1 also shows that the results are not sensitive to different measures of performances. For example, the mean abnormal return using the three RGM is 0.0 % with a very similar standard deviation around 2.1 %. These finding are also consistent with those in Brown and Warner (1980, 1985), which suggest that simple statistic models as the mean-adjusted returns model often produce comparable results to those of more sophisticated models.

4.2 Properties of Sample Mean Excess Returns

[5] As Brown and Warner (1984) and Campbell and Wasley (1993) indicate, for the cases of skewness and kurtosis coefficients, and the studentized range, the 95th and 99th percentiles for a normal population are:

Variable

0.950.99Skewness0.060.13Kurtosis3.523.87Studentized Range6.156.85

Table 2 shows cross-sectional properties of the sample mean of daily excess return at day zero. Similar to table 1, the different measures of abnormal performance are based on three different RGM. Parameter estimates are computed based on 500 samples of 50, 40, 30, 20 and 10 securities. For each sample, the mean sample estimate is the simple average of the abnormal performance measures for the individual securities in the sample. Mean, standard deviation, skewness and kurtosis coefficients, and the studentized range are computed based on 500 values of the sample mean estimate, one for each sample.

As we should expect under the Central Limit Theorem, results of table 2 show that departures from normality are less severe for cross-sectional mean excess returns than for individual excess returns. For samples of 50 securities and using the OLS market model, the cross-sectional distribution of the sample mean excess returns seems close to normal. Results in table 2 also indicate that departures from normality still persist for portfolios less or equal than 40 securities. For example, the cross-sectional distributions of day-0 mean abnormal performance for portfolios of 30 securities exhibit skewness above 0.23. These departures, however, are less pronounced than those documented in Campbell and Wasley (1993, table 1) using NASDAQ daily security returns.

4.3 Properties of the Test Statistics

For the OLS market model table 3 summarizes the empirical distributions of each test statistic based on 500 portfolios when no abnormal performance is introduced. Under the null hypothesis of no abnormal performance, the distribution of each test statistic should be unit normal. For a portfolio size equal to or above 30 securities the empirical distribution of the standardized statistic shows small departures from its theoretical distribution. However, table 3 also indicates that as the portfolio size decreases the degree of nonnormality of the standardized test increases severely. Results in table 3 also show that the cross-sectional test presents a highly negative skewness and that departures from normality for the portfolio test persist even for sample sizes of 40 securities.

Table 2

Cross-sectional properties of sample-wide mean abnormal performance measures on day 0 using three different return-generating models (RGM) when no abnormal performance is introduced. Each number in the table is based on 500 estimates of the mean, one for each sample. Securities and event dates are randomly selected (with replacement) from 02/01/1985 through 07/31/2003.

Size	Performance Measure	Mean	Standard Deviation	Skewness	Kurtosis	Studentized Range	Jarque-Bera Test
50	Market Model	0,0002	0,0031	0,04	3,03	6,06	0,17
	Market-Adjusted	0,0005	0,0031	0,06	3,18	5,78	0,92
	Mean-Adjusted	0,0002	0,0033	0,09	3,05	6,17	0,72
40	Market Model	0,0000	0,0036	0,09	3,59	6,99	7,88
	Market-Adjusted	0,0004	0,0034	0,03	3,59	7,12	7,26
	Mean-Adjusted	0,0000	0,0038	0,17	3,93	7,72	20,69
30	Market Model	0,0001	0,0040	0,24	3,38	6,56	7,96
	Market-Adjusted	0,0005	0,0039	0,25	3,36	6,02	7,84
	Mean-Adjusted	0,0002	0,0043	0,23	3,76	6,68	16,41
20	Market Model	0,0000	0,0051	0,15	3,12	5,85	2,11
	Market-Adjusted	0,0005	0,0049	0,23	3,57	7,35	11,43
	Mean-Adjusted	0,0002	0,0054	0,36	3,27	6,18	12,42
10	Market Model	-0,0002	0,0065	0,18	3,70	7,21	12,96
	Market-Adjusted	0,0001	0,0063	0,20	3,93	7,52	21,20
	Mean-Adjusted	-0,0002	0,0074	0,05	3,97	6,40	19,86

4.4 Specification of the Tests

For a given sample, when no abnormal return is present we test whether the hypothesis of no abnormal return is accepted or rejected. Given a particular RGM, the null hypothesis should be true if the securities of the random sample do not, on average, evidence any abnormal return. Thus, rejecting the null hypothesis of no abnormal performance on day 0 when it is true constitutes a *Type I* error.

Table 3

Summary measures for the empirical distribution of each test statistic, one for each sample, with sample sizes from 50 to 10 securities. The procedure to detect abnormal performance is the OLS market model and no abnormal performance has been introduced. Each number in the table represents the simple average of 500 estimates.

	Mean	Standard Deviation	Skewness	Kurtosis	Studentized Range	Jarque-Bera Test
Standardized	0,0997	1,01	0,04	3,00	6,24	0,14
Cross-Sectional	0,0203	1,00	-0,31	2,70	5,47	9,73
Portfolio	0,0606	0,98	0,00	3,02	6,11	0,01
Standardized	-0,0057	1,03	0,14	3,33	6,45	4,00
Cross-Sectional	-0,0360	1,05	-0,16	3,05	6,03	2,06
Portfolio	-0,0117	1,00	0,07	3,63	7,05	8,69
Standardized	0,0156	1,02	0,08	3,11	7,12	0,82
Cross-Sectional	-0,0404	1,02	-0,20	2,78	6,10	4,44
Portfolio	0,0127	0,97	0,21	3,34	6,30	6,20
Standardized	0,0151	1,08	0,23	3,33	6,83	6,56
Cross-Sectional	-0,0499	1,09	-0,17	3,01	6,86	2,27
Portfolio	0,0074	1,00	0,26	3,21	5,86	6,71
Standardized	-0,0251	0,99	0,26	3,54	6,83	11,64
Cross-Sectional	-0,0848	1,07	-0,08	3,40	6,80	3,84
Portfolio	-0,0245	0,90	0,32	3,55	6,38	14,64

Table 4 shows the frequency of rejection using three different tests to detect abnormal performance. Notice that implicit in the three t-tests used to detect abnormal performance is the strong assumption that security returns have a normal distribution. If this assumption is not correct, then the sampling distribution of test statistics assumed for the hypothesis tests departs from the true distribution and false inference may result.[6] The numbers in table 4 indicate that when a test of size 5% is used, Type I error rates range from 3.2% to 6.4%. Thus, these results reveal that using one tail test at the 0.05 significance the three test

[6] As Brown and Warner (1980, pp.216) indicate, when the null hypothesis is true, even though the empirical distribution of a specific test statistic is consistent with the assumed theoretical distribution, the proportion of rejections will not be precisely equal to the test level. The reason is that the proportion is itself a random variable –a Bernoulli process– with a media equal to α and standard deviation equal to α *(1- α).* For a test of size α, if the proportion of rejection distributes normal and the test is properly specified, the empirical percentage of rejection for each of the *m* sample should be into the interval:

$$\alpha \pm \phi^{-1}(1-\frac{\alpha}{2})\left[\frac{\alpha(1-\alpha)}{m}\right]^{1/2}$$

For example, for a significance level of α = 5%, if the outcomes for each of the 500 samples are independent (*m*=500 trials), the rejection rates follow a Bernoulli process with mean 0.05 and standard deviation 0.0097. Then, if the test are properly specified the proportion of rejections should be between 2.7% and 7.3% approximately 99% of the time.

statistics are well specified under the null hypothesis of no abnormal performance. However, results in Table 4 also indicate that symptoms of misspecification arise using a significance level of 1% for both the standardized and the portfolio t-test. For example, the Type I error rate for the standardized test ranges from 2.4 to 3.6% and from 1.4 to 2.8% for the standardized and portfolio t-test, respectively.

4.5 Power of the Test

We also examine how the test statistics perform when the null hypothesis is false. To simulate an abnormal performance a particular abnormal return is introduced into the mean abnormal returns of the sample. Then, the hypothesis of no abnormal performance is tested again. Thus, failing to reject the null hypothesis of no abnormal return when it is false constitutes a *Type II* error.

Table 4 shows, for three tests and three RGM, the frequency with which the hypothesis of no abnormal performance in day 0 is rejected. For example, for a significance level of α=5% and using the OLS market model, when we introduce a 0.5% level of abnormal performance, the rejection rate for the standardized test is 59% compared to 44% and 38% for the cross-sectional and the portfolio test. Moreover, the higher power of the standardized test does not depend on the level of significance. For a test of α=1%, and also with 0.5% of abnormal performance, the rate of rejection of the standardized test ranges from 36% using the mean-adjusted model to 50% using the market-adjusted model. Thus, our findings indicate that using Chilean daily security return data the standardized t-test is more likely to detect abnormal return than both the cross-sectional and the portfolio t- test.

4.6 Sensitivity Analysis

A. Comparing Alternative RGM

Table 4 also compares the power of detecting abnormal performances among three different methods to detect abnormal performance. In general, the rejection frequencies indicate that both the market-adjusted returns model and the OLS market model show somewhat better performance than the mean-adjusted return method. For example, using a standardized t-test of size 5% and 0.5% of abnormal performance, the mean-adjusted returns model rejects the null hypothesis 49% of the times while the OLS market model and the market-adjusted method register rejection rates of 59% and 64%, respectively. These findings also seem to be robust with respect to changes in the significance level.[7]

Thus, our results suggest that in terms of procedure to measure abnormal performance there is some evidence indicating a better performance of those methods that consider the systematic risk of each security. However, the improvement power of the tests using these two methods over the simpler mean-adjusted model is limited. When an abnormal performance of 2.5% is introduced, the three RGM allow us to identify this abnormal return all of the times, regardless of the test size.

B. Different Sample Sizes

Results in Table 5 show that the specification of the tests is not particularly sensitive to the number of securities in the sample. When a test of size 5% is used no special misspecification of either test statistic is found in samples from size 50 to 10 securities. Some symptoms of misspecification arise only at a 1% level of significance. For example, for a portfolio size of 20 securities the standardized and cross-sectional tests seems to be misspecified with error rates of 3.6% and 2.6%, respectively.

As we should expect, the power of the tests also falls strongly when the sample size decreases. Using the standardized test of size 5% and 1% level of abnormal performance, decreasing the sample size from 50 to 10 securities reduces the rejection frequency from 99% to 42%.

[7] However, as Brown and Warner (1980, 1985) point out, it is possible that these results dependent significantly on the fact that in this simulation work the precise time at which the abnormal return occurs is known with certainty.

Table 4

A comparison of three alternative RGM and three test statistics for detecting abnormal excess return. Values in the table indicate the percentage of 500 samples where the null hypothesis of no abnormal performance on day 0 is rejected. Sample size is equal to 50 securities. Chilean stock securities and event dates are randomly selected (with replacement) from 01/02/85 through 07/31/03.

Panel A

Two tailed test, α=0.05

		Artificial level of abnormal performance (%) introduced at day 0					
Performance Measure	Test Statistic	0.0	0.5	1.0	1.5	2.0	2.5
OLS Market Model	Standardized	5.0	58.6	98.6	100	100	100
	Cross-Sectional	4.8	44.0	90.4	99.2	99.8	100
	Portfolio	4.0	38.4	89.6	99.8	100	100
Market-Adjusted	Standardized	5.6	64.4	98.8	100	100	100
	Cross-Sectional	6.4	49.4	92.6	99.6	99.8	100
	Portfolio	6.0	42.6	92.2	100	100	100
Mean-Adjusted	Standardized	5.0	49.4	94.2	100	100	100
	Cross-Sectional	4.2	37.0	85.8	99.0	99.8	100
	Portfolio	3.2	32.0	83.8	99.6	100	100

Panel B

One tailed test, α=0.01

		Artificial level of abnormal performance (%) introduced at day 0					
Performance Measure	Test Statistic	0.0	0.5	1.0	1.5	2.0	2.5
OLS Market Model	Standardized	2.4	44.2	96.0	100	100	100
	Cross-Sectional	1.4	30.0	82.6	98.2	99.6	100
	Portfolio	2.4	23.0	79.8	99.0	100	100
Market-Adjusted	Standardized	3.6	50.0	98.0	100	100	100
	Cross-Sectional	2.0	35.0	86.2	98.8	99.8	100
	Portfolio	2.8	29.8	85.6	99.4	100	100
Mean-Adjusted	Standardized	2.6	35.8	90.8	100	100	100
	Cross-Sectional	1.4	23.4	75.4	98.4	99.6	100
	Portfolio	1.4	18.8	73.6	98.6	100	100

A test of significance is well-specified under the null hypothesis at the 5% (1%) level if the percent of rejections falls between 2.7% and 7.3% (0.0 and 2.2%).

Table 5 also indicates that the relative power of different test statistics also seems to be independent of the sample size. In terms of power, dominance of the standardized test over the cross-sectional and the portfolio test does not change.

C. Multiday-Event Periods

The simulations we have performed at this time make the strong assumption that the date at which abnormal performance takes place is known with entire certainty. However, given that most of the times the calendar date of the event cannot be identified exactly, most event study settings involve multiday event periods where the date itself becomes a random variable. To analyze this, we also examine how uncertainty about the precise date of the abnormal performance affects the power of the event study technique.

Using the OLS market model as a RGM for each security in the 500 samples we select one day of the event period at random and add a particular level of abnormal performance in one specific day in windows of 3, 5, and 11 days.[8] For example, for a window of 11 days we add a particular level of abnormal performance in one specific day (randomly selected) in the interval from day -5 through +5. Thus, this experiment simulates a situation where the abnormal performance occurs at some –unknown– date in the event period including the event day.

[8] For each security, the event day is a drawing from a uniform distribution.

Table 5

The effect of different sample sizes for detecting abnormal excess return using the OLS market model. Values in the table indicate the percentage of 500 samples where the null hypothesis of no abnormal performance on day 0 is rejected. Sample sizes are equal to 50, 40, 30, 20 and 10 securities. Chilean securities and event dates are randomly selected (with replacement) from 01/02/85 through 07/31/03.

		Two tailed test, α=0.05						Two tailed test, α=0.01					
		Artificial level of abnormal performance (%) introduced at day 0						Artificial level of abnormal performance (%) introduced at day 0					
Portfolio Size	Test Statistic	0.0	0.5	1.0	1.5	2.0	2.5	0.0	0.5	1.0	1.5	2.0	2.5
50	Standardized	5.0	58.6	98.6	100	100	100	2.4	44.2	96.0	100	100	100
	Cross-Sectional	4.8	44.0	90.4	99.2	99.8	100	1.4	30.0	82.6	98.2	99.6	100
	Portfolio	4.0	38.4	89.6	99.8	100	100	2.4	23.0	79.8	99.0	100	100
40	Standardized	5.2	44.4	94.0	100	100	100	3.0	31.6	89.2	99.8	100	100
	Cross-Sectional	7.2	35.6	83.0	98.0	99.4	99.6	2.6	24.8	74.2	95.6	99.4	99.4
	Portfolio	4.4	28.4	80.0	98.4	99.8	100	2.6	16.2	67.2	97.4	99.8	100
30	Standardized	5.2	36.0	86.4	99.6	100	100	1.8	22.2	77.2	99.6	100	100
	Cross-Sectional	4.6	27.0	75.6	94.6	99.8	100	1.4	17.0	62.6	91.6	98.8	99.8
	Portfolio	4.0	23.8	67.8	96.0	100	100	2.2	11.6	53.8	90.8	99.0	100
20	Standardized	6.2	26.6	70.0	96.0	100	100						
	Cross-Sectional	7.0	23.6	61.2	86.2	95.6	99.0	3.6	17.4	56.6	90.8	99.0	100
	Portfolio	4.8	17.8	48.8	86.6	96.8	99.8	2.6	15.8	47.2	78.2	93.0	97.8
								1.8	9.0	36.0	73.4	93.6	99.0
10	Standardized	5.2	14.2	42.0	76.2	92.8	98.6	2.0	7.2	27.4	64.4	87.2	97.0
	Cross-Sectional	6.6	14.4	40.6	68.4	84.6	93.4	3.2	8.2	28.6	57.6	78.6	88.8
	Portfolio	3.2	8.4	24.0	58.2	79.2	92.8	1.6	4.2	15.2	40.0	68.8	87.2

A test of significance is well-specified under the null hypothesis if the percentage of rejections falls between 2.7 % and 7.3% for a test size of 5% level and between 0.0% and 2.0% for a test size of 1% level.

Table 6 reports results in the multi-day setting for abnormal performance levels ranging from 0 to 2.5%. Similar to our findings involving a one-day setting, numbers in table 6 indicate that symptoms of misspecification arise for the three tests when a test of size 1% is used. However, results in table 6 also indicate that misspecification is more severe for the standardized t-test when event periods are longer than one day. For example, when a standardized t-test of size 5% is used, Type I error rates range from 7.8 to 8.8%. As we should expect, the power of the three tests are much lower than those where the precise date of abnormal performance was known with entire certainty. For example, for a window of 11 days and a 5% alpha level, the portfolio t-test is able to detect the presence of 1.5% abnormal performance only 27% of the times, compared to the 100% of the earlier table 4.

Table 6

A comparison of alternative test statistics when the precise date of abnormal performance is unknown. Using the OLS market model as RGM, abnormal performance for each security is introduced at random for one day during the event window. The numbers in the table show the percentage of 500 samples of 50 securities each where the null hypothesis of cumulative mean abnormal performance in the event period is rejected.

Panel A

Two tailed test, α=0.05

		Artificial level of abnormal performance (%) introduced at day 0					
Window	Test Statistic	0.0	0.5	1.0	1.5	2.0	2.5
3 days	Standardized	8.8	22.0	58.0	90.0	99.4	100
	Cross-Sectional	5.4	14.2	43.8	76.2	93.4	99.0
	Portfolio	6.4	15.2	42.0	73.0	93.2	99.2
5 days	Standardized	8.4	16.8	41.0	72.8	92.6	98.2
	Cross-Sectional	5.4	8.6	27.2	51.4	76.8	90.2
	Portfolio	7.2	11.2	28.0	50.6	78.2	91.2
11 days	Standardized	7.8	10.8	28.0	45.0	64.8	82.2
	Cross-Sectional	4.2	5.8	13.0	26.6	42.6	61.2
	Portfolio	5.0	8.4	17.0	27.4	44.0	63.2

A test of significance is well-specified under the null hypothesis at the 5% (1%) level if the percent of rejections falls between 2.7% and 7.3% (0.0 and 2.2%).

Table 6

A comparison of alternative test statistics when the precise date of abnormal performance is unknown. Using the OLS market model as RGM, abnormal performance for each security is introduced at random for one day during the event window. The numbers in the table show the percentage of 500 samples of 50 securities each where the null hypothesis of cumulative mean abnormal performance in the event period is rejected.

Panel B

Two tailed test, α=0.01

		Artificial level of abnormal performance (%) introduced at day 0					
Window	Test Statistic	0.0	0.5	1.0	1.5	2.0	2.5
3 days	Standardized	4.4	13.4	46.8	81.6	97.4	99.8
	Cross-Sectional	2.8	7.8	29.4	63.0	87.8	97.2
	Portfolio	3.0	7.4	28.6	61.8	87.0	97.0
5 days	Standardized	4.4	8.6	29.2	60.4	87.6	96.4
	Cross-Sectional	2.6	4.8	16.8	39.4	66.0	84.2
	Portfolio	3.8	5.8	18.2	38.4	64.8	84.6
11 days	Standardized	3.8	5.0	17.0	33.0	51.6	70.8
	Cross-Sectional	1.4	1.8	5.8	14.8	30.0	46.0
	Portfolio	2.2	3.4	9.4	18.6	31.6	48.0

A test of significance is well-specified under the null hypothesis at the 5% (1%) level if the percent of rejections falls between 2.7% and 7.3% (0.0 and 2.2%).

V. Summary and Conclusions

This paper examines the specification and power of the event studies technique using daily Chilean security return data. We test several procedures with which this methodology measures security price performance. Our findings indicate that although measures for individual security returns and security excess returns are evidently non-normal, the cross-sectional mean abnormal returns converge to normality as the number of securities in the sample increases. Methods based on the use of parametric tests for samples of 10 or more securities seem to be well specified at least at the 5% significance level. However, for a 1% significance level the three parametric tests we study reject the null hypothesis too often when this hypothesis is true.

Comparison across test statistics indicates important power differences among the three tests. For Chilean daily security return data, the standardized t-test shows to be more powerful than both the cross-sectional and the portfolio t-tests. Moreover, these results show to be robust to changes in the portfolio size, the RGM or the length of the event period. As we should expect, our results also indicate that the power of the three parametric tests we study falls strongly when the sample size decreases. For example, using the portfolio test with α=5% and 1% level of abnormal performance, diminishing the sample size from 50 to 10 securities reduces the rejection frequency between three and four times.

One additional and interesting result is that in terms of procedure to measure abnormal performance there is some evidence indicating a better performance of those methods that consider the systematic risk of each security. However, the improvement in specification and power of the tests using these methods over the simpler mean-adjusted model is only limited.

Finally, the specification and power of the tests also depend on the long of the event period. We find a more severe misspecification for the standardized t-test when the event period is longer than one day. As expected, the power of the three tests also decreases strongly as the longer is the event window. Results for 11-day event periods reveal that in samples of 50 securities, when we introduce a 1% of abnormal performance using a test of size 5%, the null hypothesis of no abnormal performance is rejected only 17 to 28% of the time. Thus, as Brown and Warner (1980) point out, even if the researcher conducting an event study today can take advantage of a more sophisticated pool of computation techniques, a good use of his time is still in determining more accurately event dates.

VI. References

Barber, B. and Lyon, J. (1997), "Detecting Long-Run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics," *Journal of Financial Economics 43, pp. 341-72.*

Bhattacharya, U., Daouk, H., Jorgenson, B. and Kehr, C. (2000), "When an Event is not an Event: The Curious Case an Emerging Market," *Journal of Financial Economics 55, pp. 69-101*

Brav, A. (2000), "Inference in Long Horizon Event Studies: A Bayesian Approach with Application to Initial Public Offerings." *The Journal of Finance 55, pp. 1979-2016.*

Brown, S. and Warner, J. (1980), "Measuring Security Price Performance," *Journal of Financial Economics 8, pp. 205-58.*

_______(1985), "Using Daily Stock Returns: The Case of Events Studies," *Journal of Financial Economics 14, pp. 3-31.*

Campbell, J., Lo A., and MacKinlay, C. (1997), "The Econometrics of Financial Markets," Princeton University Press.

Campbell, C. and Wasley C. (1993), "Measuring Security Price Performance Using Daily NASDAQ Returns," *Journal of Financial Economics 33, pp. 73-92.*

Castillo, A. (2005), "The Announcement Effect of Bond and Equity Issues: Evidence from Chile." *Estudios de Administración*. Forthcoming.

Celis, C., and G. Maturana (1998), "Initial Public Offerings in Chile," *Abante 1, pp. 7-31.*

Cowan, A. and Sergeant, A. (1996), "Trading Frequency and Event Study Test Specification," *Journal of Banking and Finance 20, pp. 1731-57.*

Dyckman T. Philbrick, D. and Stephan, J. (1984), "A Comparison of Event Study Methodologies Using Daily Stock Returns: A Simulation Approach" *Journal of Accounting Research 22, pp. 1-33.*

Fama, E. (1976), "Foundations of Finances," New York: Basic Books.

Kothari, S. (2001), "Capital Markets Research in Accounting," *Journal of Accounting and Economics 31, pp. 105-231.*

Kothari, S. and Warner, J. (1997) "Measuring Long-Horizon Security Price Performance," *Journal of Financial Economic 43, pp. 301-39.*

_____________(2005). "Econometrics of Event Studies." *Working Paper, Center for Corporate Governance, Tuck School of Business at Dartmouth.*

MacKinlay, C. (1997), "Event Studies in Economics and Finance." *Journal of Economics Literature 35, pp. 13-39.*

McWilliams, A. and Siegel, D. (1997), "Event Studies in Management Research: Theoretical and Empirical Issues," *Academy of Management Journal 40, pp. 626-57.*

Maynes, E. and Rumsey, J. (1993), "Conducting Event Studies with Thinly Traded Stocks," *Journal of Banking and Finance 17, pp. 145-157.*

Morán, P. (2003), "Looking Back at the Controversy: Unexpected Wealth Effects of a Transitory Cause," *Abante 6, pp. 117-47*.

Parisi, F., and D. Pérez (2000), "Cambios en el Rating de Bonos y su Efecto en los Precios Accionarios: El Caso Chileno,"*Abante 3, pp. 249-73.*

Patell, J. (1976), "Corporate Forecasts Earnings per Share and Stock Price Behavior: Empirical Tests*, Journal of Accounting Research 14, pp. 246-76.*

Peterson, P. (1989), "Event Studies: A Review of Issues and Methodology," *Quarterly Journal of Business and Economics 28 (3), pp. 36-66.*

Roll, R. (1977), "A Critique of the Asset Pricing Theory's Tests," *Journal of Financial Economics 2*, *pp. 129-176*.

Rouwenhorst, K.G. (1999), "Local Return Factors and Turnover in Emerging Stock Markets," *The Journal of Finance 54, pp. 1439-64.*

Saens, R. (1999), "Premia in Emerging Market ADR Prices: Evidence from Chile," *Abante 2, pp. 51-70.*

Urrutia, J. (1995), "Test of Random Walk and Market Efficiency for Latin American Emerging Equity Markets." *The Journal of Financial Research, 18, 299-309.*