A Simulation Investigation of Seemingly Unrelated Regression as Used in Accounting Information Event Studies

Walter Teets* Robert Parks†

August 21, 1992

Comments welcome.

The authors thank Cray Research, Inc., for the generous grant of time on a CRAY-2S/4-128 at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign. We thank Dick Dietrich, Nick Dopuch, John Fellingham, Grace Pownall, Barb Wheeling, and Dave Ziebart for helpful discussions on this topic.

Researchers studying stock price reactions to accounting information releases can choose among several statistical methods/models. Firm-specific equation methods appear to be particularly appropriate when the research hypotheses involve possible differences across firms. However, the firm-specific equation methods make more demands on the data, requiring estimation of firm-specific coefficients and (possibly) covariance parameters. Therefore, it is not clear whether the researcher realizes net gains by using firm-specific equation methods. In this paper, we examine the empirical behavior of test statistics arising from one firm-specific equation method, seemingly unrelated regression (SUR), and alternative test statistics based on the same set of equations, but not incorporating estimates of cross-sectional correlation.[1] Evidence on the empirical distributions of these statistics may guide researchers in designing and interpreting research.

Understanding the empirical behavior of SUR is essential before accounting researchers can correctly interpret results of existent research or take full advantage of its conceptually desirable features when testing hypotheses involving possible differences across firms. If characteristics of the data used by accountants lead to high Type I error rates or poor power, inference based on normal theory statistics derived from SUR may be faulty. However, use of critical values based on empirical distributions may overcome the difficulties and lead to appropriate inferences.

We provide evidence on the empirical distributions of several SUR statistics through a simulation of accounting information releases and associated stock price responses. Following Brown and Warner [1980, 1985], we start with actual stock returns, randomly select event dates and introduce abnormal performance on those dates. Using this data, we estimate a SUR model and calculate statistics. Multiple repetitions of this process generate empirical distributions that provide evidence on Type I error rates and power. We generate empirical distributions of the statistics for a number of different scenarios corresponding to different types of accounting information releases. We vary the number of firms used in the estimation procedure, the level of abnormal performance introduced on an event date, the number of event dates per firm in the estimation period, and the number of days included in the event window. This allows us to assess whether SUR should be expected to work better in some accounting research contexts than in others. We use NYSE, ASE, and NASDAQ firms. Briefly, we find (1) the null hypothesis that all coefficients relating stock returns to information events are simultaneously equal to zero is rejected far too often when no abnormal performance has been introduced into the returns; (2) after correcting for high Type I error rates, powers of statistics testing all coefficients simultaneously equal to zero are low, especially if there are few events per firm; (3) the hypothesis that the average of the coefficients relating stock returns to information events equals zero is sometimes rejected

[*] Assistant Professor, University of Illinois at Urbana–Champaign.

[†] Associate Professor, Washington University in St. Louis.

[1] In the interests of readability, we will use "SUR" to refer to both of the firm-specific equation models we discuss. It should be recognized that "SUR" properly refers only to the model that utilizes estimates of cross-sectional correlation. In the remainder of the text, when it is necessary to distinguish between the two methods, we will use the terms "consecutive equations" when the equations are estimated using OLS firm by firm, and "true SUR" when discussing simultaneous estimation incorporating estimates of cross-sectional correlation.

too often when no abnormal performance is introduced, but the over-rejection is not severe; and (4) event date uncertainty, as it is typically addressed in SUR, severely reduces the powers of the tests in most situations.

The paper is related to research that has used simulation techniques to examine the behavior of statistical methods used in event studies. Using monthly and/or daily returns data for American and New York stock exchange firms, Brown and Warner [1980, 1985] and Dyckman, Philbrick, and Stephan [1984] investigate how well various abnormal return metrics, pooled cross-sectionally, are able to identify significant average abnormal returns. Campbell and Wasley [1991] examine the same issue using daily returns for NASDAQ securities. None of these studies examine methods that utilize firm-specific equations relating information variables to firm returns. Hence, they do not examine specifically the situation where effects may differ across firms, nor do they address the incorporation of estimates of cross-sectional correlation into the estimation procedure.

Collins and Dent [1984] propose and examine via simulations a technique that incorporates cross-sectional correlation in the case where all events affect all firms on the same day(s), using NYSE and ASE firms. Malatesta [1986] and McDonald [1987] both examine SUR in event study frameworks, but both use only one event per firm, and use only samples of 30 firms. In Malatesta, the independent variables are the same across all firms—the so-called multivariate analysis. We extend previous research on SUR on several dimensions. First, we simulate 2, 5, and 20 events per firm, instead of only 1. We also examine how the number of firms in the model affects the statistics by using samples of 25, 50, and 75 firms, rather than only 1 sample size of 30 firms. Finally, we investigate the effect of event date uncertainty on test statistics by examining 1, 2, and 5 day event windows.

The remainder of the paper is organized as follows. Section 1 discusses the use of SUR in accounting and finance event studies, and reasons for concern about the test statistics generated. Section 2 presents the approach used in the simulations, including the scenarios simulated and how those scenarios relate to accounting research situations. Section 3 covers the hypotheses tested in event studies, the statistics collected from each simulation, and the theoretical distributions of these statistics. Sections 4 and 5 present the empirical results of the simulations and implications for design and interpretation of accounting research. Finally, section 6 gives directions for future work.

1 SUR as used in accounting event studies

Accounting information event studies assess stock price reactions to releases of accounting information. The researcher attempts to estimate the relation between new information in an accounting number and the change in stock price (the stock return). Early event studies examined (cumulative) abnormal returns from the event period to see whether there were significant abnormal returns on average across firms during the event period. To assess differential returns across different groups of firms, firms were assigned to portfolios based on some firm characteristic, and the abnormal returns for different portfolios were compared. A problem with this approach to assessing differential returns in some research contexts is that either (1) the levels of information in the accounting numbers must be assumed to be the same for all firms, so a firm's abnormal return captures the (cross-sectionally varying) stock price effect of the "unit" of information, or (2) the abnormal return for each firm must be assumed to be a composite, due partly to the level of new information in the firm's accounting number, which could vary cross-sectionally, and partly to the coefficient mapping a unit of information into the stock return, which could also vary cross-sectionally. In other words, in cases where the level of information in an accounting number may vary across firms, and the stock price response to a unit of information may vary across firms, abnormal returns give composite measures that are not easily separated into their constituent parts. Researchers must ignore part of the information available in the accounting release if portfolio abnormal return analysis is used. Alternatively, abnormal returns can be regressed on measures of information in cross-sectional pooled regressions. In this case, the researcher implicitly assumes the coefficient relating information to abnormal returns is cross-sectionally constant.

To incorporate both different levels of information, and different mappings across firms of stock price response to a "unit" of information, an extended market model can be used. The extended market model is

$$r_{it} = \alpha_i + \beta_i \, r_{mt} + \gamma_i \, A_{it} + \varepsilon_{it} \tag{1}$$

where r_{it} and r_{mt} are daily returns to firm i's stock and the market portfolio for day t, ε_{it} is a random error term, and α_i and β_i are market model parameters. The terms α_i and $\beta_i \, r_{mt}$ are included to control for market wide movements in stock prices unrelated to the accounting information releases of interest. The term ε_{it} is included to allow for other events affecting stock price that are not related to the accounting information release of interest. A_i is a vector of accounting information releases whose elements take on nonzero values only for those days on which there is a release of interest. γ_i is a firm-specific coefficient relating accounting information to stock returns.

If there is only one nonzero element per firm, equal to unity, in each A_i vector, the γ_i coefficients are equivalent to market model abnormal returns. If there are multiple nonzero elements in the A_i vector, each equal to unity, the γ_i coefficient is equal to the average of all event period abnormal returns for firm i. Finally, if the nonzero elements of the A_i vectors vary cross-sectionally and through time, the γ_i coefficients are firm-specific response coefficients that map different levels of information into stock returns. It is this last situation that we simulate in this paper.

Equation (1) can be estimated using ordinary least squares for each firm. However, if there is contemporaneous correlation across firms among the errors ε_i, stated significance levels of statistical tests may be incorrect, potentially leading to incorrect inference. SUR may be used to incorporate estimates of cross-sectional correlation in the estimation process and in statistical tests.

The general SUR model is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_n \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2}$$

where

n	=	number of firms in the sample
Y_i	=	$t \times 1$ vector of t time series observations on the dependent variable for the ith firm; this corresponds to the r_i vectors in equation (1)
X_i	=	$t \times k$ matrix of explanatory variables for firm i; from (1), this matrix consists of the constant, r_{mt}, and A_i vectors
Γ_i	=	$k \times 1$ vector of firm-specific coefficients (α_i, β_i, and γ_i from (1)) relating the dependent variable to the explanatory variables, and
ε_i	=	$t \times 1$ vector of errors.

This set of equations may be written

$$Y = X\Gamma + \varepsilon.$$

The most efficient estimator of Γ is

$$\hat{\Gamma} = (X'(\hat{\Sigma}^{-1} \otimes I)X)^{-1}(X'(\hat{\Sigma}^{-1} \otimes I)Y), \tag{3}$$

where $\hat{\Sigma}$ is an $n \times n$ matrix of pairwise covariances among the n firms. The elements of $\hat{\Sigma}$ are obtained by estimating the firm specific equations in (2) equation by equation, obtaining the n error vectors ε_i, and calculating covariances between all pairs of error vectors.

Hypotheses about the relations between returns and accounting information events can be tested using linear combinations of the γ_i coefficients, either the OLS estimates from the individual firm-specific equations in (1), or the EGLS estimates from equations (2) and (3). If the researcher believes *ex ante* that cross-sectional correlation is negligible, s/he may choose to use the OLS estimates, hence avoiding the possible introduction of estimation error in $\hat{\Sigma}$. If the probability of significant cross-sectional correlation is high, however, the researcher may choose to use the EGLS estimates, even though estimation error may be present in $\hat{\Sigma}$. We present evidence on several OLS and EGLS statistics that may be used in hypothesis testing.

Although the capability to estimate firm-specific coefficients and incorporate estimates of cross-sectional correlation is desirable in many accounting contexts, these capabilities do not come costlessly. There are four potential problems. First, SUR assumes normally distributed error terms, but it is well known that market model daily abnormal returns (essentially the ε_i in equation (1)) are leptokurtic. It is not known how sensitive the SUR statistics are to departures from normality. Second, the test statistics frequently used in SUR are only asymptotically correct. While a typical accounting event study using SUR uses a time series of approximately five years of daily data per firm, there is no theory specifying how many observations are needed before the asymptotic statistics are relevant. Third, although the researcher may use five years of daily returns, the number of non-zero observations per firm in the event vector itself is generally much smaller—perhaps as few as 1 or 2 in the case of management forecasts of earnings. While the total number of observations per firm is large, the number of meaningful observations used to estimate the main parameters of interest is very small. But the degrees of freedom used to establish rejection regions are based on the total number of observations, not the meaningful observations for the parameters of interest. Fourth, the cross-sectional correlation matrix is not known, but must be estimated from the data. Another source of estimation error is therefore introduced into the model.

In this paper, we provide evidence on the joint effect on the test statistics of these four potential problems. Determining the contribution of each effect individually is beyond the scope of this paper. Parks and Teets [1993] provide additional evidence on the contribution of the individual elements.

2 Simulation overview, rationale, and implementation details

The objective of this study is to furnish evidence on the empirical behavior of SUR by simulating SUR in contexts representative of those accounting researchers might encounter. This section gives a broad overview of the simulation process, followed by a more detailed examination of the rationale for and the implementation details of each step of the process.

2.1 Overview of simulation procedures

All of the simulations reported in this paper started with actual firm and market returns. There were four different factors that we manipulated across the simulations: (1) the number of events per firm, (2) the magnitude of the abnormal performance added to the actual return on an event date, (3) the number of firms in the model, and (4) the length of the event window around each event date. All combinations of these four factors (detailed in the following sections) gave us 108 different scenarios:

	Ranges of abnormal performance introduced	4
	0, [.00125,.00375], [.00375,.00625], [.0025,.0075]	
$\times$	Numbers of events per firm	3
	2, 5, or 20	
$\times$	Numbers of firms in model	3
	25, 50, or 75	
$\times$	Length of event window	3
	1, 2, or 5 trading days	
	Total number of scenarios	108

2.2 Returns

The underlying returns used in these simulations are actual firm returns. Starting with actual returns is important, as it is well-known that daily returns are leptokurtic. The departure from normality may affect the empirical distributions of the SUR statistics used in hypothesis tests. We gathered returns data for 30 different sets of 75 firms, and corresponding market returns. For each firm, 1,280 returns were used. This represents approximately five years of daily returns.[2]

One potential benefit of SUR is that it incorporates estimates of cross-sectional correlation of residuals in coefficient estimates and statistical tests. A reason for this correlation may be that firms in an industry are affected similarly by events for which the market model provides insufficient control. Therefore, we selected firms from related industries for each set of 75 firms. First, listings of NYSE/ASE and NASDAQ firms in each two digit SIC code were generated from the CRSP daily stock return files. Separate lists were generated for 1975–79, 1980–84, and 1985–89. To be included in a list, a firm could have no missing data during the respective time period. Next, we combined lists of firms from consecutive two digit SIC codes until we had lists containing 75 firms. For example, one set contained returns from 1975–79 for NYSE/ASE firms in SIC codes 34–35. The 30 sets of firms included 10 sets from each of the three time periods. Twenty sets were composed of NYSE/ASE firms and ten were from NASDAQ. Each set of returns was used as the basis for 55 sets of simulations, resulting in 1,649 simulations for each of the 108 different scenarios.[3]

2.3 Event dates

Two broad classes of events are of interest to accounting researchers. The first comprises those accounting-related announcements that occur irregularly. The second comprises events that occur regularly. SUR may not be equally effective in these different situations, because they give rise to different numbers of non-zero elements in the information event vectors. Management forecasts may give rise to only one or two announcements in a five year period. There will be five annual reports issued in a five year period, and 20 quarterly earnings announcements. To investigate whether the number of events per firm affects the distributions of test statistics, we ran simulations where each firm experienced 2, 5, or 20 events.

Two sets of event dates were generated for each set of simulations. The first set was used for the 5 and 20 event simulations. Twenty dates, corresponding to quarterly announcements, were generated for each firm. The first date for each firm fell within the first 77 observations. The next 19 dates were obtained by adding 63 days to the previous date (there are approximately 63 trading days per quarter). For the five events per firm simulations (yearly events), the first, fifth, etc., dates were used. Having the five events be a subset of the 20 events allows us to assess the effect of adding events to an existing set of events. This corresponds to the research situation where the researcher must decide whether to use quarterly announcements as well as annual announcements.

For the two events per firm simulations, different dates were generated. The only requirement for these dates was that they not be within thirty days of each other for the

[2]The number 1,280 was chosen due to hardware considerations on a Cray supercomputer used in the simulations.

[3]We generated 1650 simulations for each scenario, but one set of results was inadvertently erased.

same firm.

The maximum number of events per firm in a given simulation was 20. Over the 55 sets of simulations based on a single set of firms, this implies that 1,100 dates per firm were used as event dates. No attempt was made to insure that the dates for a given firm were different across simulations.

2.4 Levels of abnormal performance

In order to determine the Type I error rates of the various statistics, simulations were run where no abnormal performance was introduced into the returns vectors. Simulated values used in the A_i vectors on event dates as measures of (false) information were uniformly distributed [.00125,.00375] (mean of .0025).

To assess the powers of different statistics under the alternative hypothesis, three ranges of abnormal performance and associated information events were used: [.00125,.00375] (mean of .0025), [.00375,.00625] (mean of .005), and [.0025,.0075] (mean of .005). Three different ranges were used to assess how different levels of abnormal performance affect the ability of SUR to detect stock price responses to information events. In these simulations, the simulated values were used in the A_i vectors on event dates and were added to the respective firms' returns vectors on corresponding days. This implies that the true coefficient relating the information variables to the abnormal returns was unity for all firms.

2.5 Numbers of firms

Accounting researchers are faced with questions regarding adequacy of sample size. In SUR, there are two sample size issues. The first one, how many events per firm are needed, has been discussed in a previous section. The second one involves the number of firms that are needed for effective use of SUR. We did simulations with 3 different sample sizes, in terms of numbers of firms. Each of the 30 sets of 75 firms was broken down into subsets of 25, 50, and 75 firms. The models using 50 firms added 25 firms to the models using 25 firms, and the models using all 75 firms added 25 firms to the 50 firm models. When 25 additional firms were added to a model, the information event dates and values for the firms already in the model remained the same. This allows us to assess the effect of adding additional firms to an existing data set, rather than confounding the addition of firms with the effects of new event dates and abnormal returns for existing firms.

2.6 Event windows

Assessing the stock price change associated with an accounting information release is made more difficult because the researcher may not know exactly when a piece of accounting information became known. This difficulty was overcome in studies based on market model abnormal returns by using a cumulative abnormal return that covered a multiday window assumed to include the date of actual release of the information. Generally, longer event windows in SUR are implemented by assigning the information value for a given event to several consecutive elements in the information event vector. However, this does not accomplish the same thing that was accomplished with a cumulative abnormal return. Underlying the use of cumulative abnormal returns is the idea that the effect of (possibly unidentified) confounding events occurring during the cumulation period will average out to zero, so the expected value of the cumulative abnormal return will be the abnormal return associated with the event of interest. Assigning the information value to several consecutive event dates in SUR may accomplish quite another thing. In effect, the researcher assumes that the entire stock price change due to an accounting disclosure takes place rapidly, but is unable to identify exactly when the disclosure was made.[4] Assume that all event windows are two days in length (the information event vector has two consecutive non-zero values for each event), and that the return of interest occurs on one of the two days. In terms of equation (1), the portion of r_{it} not explained by $\alpha_i + \beta_i r_{mt}$ has an expected value of zero on non-event days, and is the desired abnormal return on the true event day. Hence, the γ_i coefficient will be estimated using observations half of which have the desired relation, and half of which in effect associate the information variable with a value of zero (or noise). Because γ_i is estimated by minimizing the sum of squared residuals, it will be biased toward zero, and its standard error will be biased upwards, relative to the situation where the information variable is non-zero only on the "correct" date.

In order to assess the behavior of SUR in situations where the event date cannot be identified exactly, two day and five day event windows are used, in addition to the one day window. In the two (five) day simulations, the information variable is assigned the non-zero value two (five) consecutive days, while the non-zero value is added as an abnormal return to the actual return on only one of the two (five) days.[5]

3 Hypotheses tested and statistics used in the tests

There are two basic hypotheses tested in accounting event studies, both having to do with the association between stock price changes and information events. The first null hypothesis is that each of the γ_i coefficients in (1) is equal to zero. The second null hypothesis is that the average stock price response to the information events, across all firms and events, is equal to zero; i.e., there is no information disclosed in the information releases that is associated with stock price changes, on average. Likelihood ratios, χ^2 statistics, several F statistics, and statistics based on sums of coefficients and associated standard errors have been proposed to test these hypotheses. Parks and Teets [1993] presents evidence on several of these statistics. Here, we concentrate on three sets of F-statistics and statistics based on sums of coefficients and associated standard errors.

[4] There may be cases when the researcher assumes that the reaction to the announcement takes place over several days. In these cases, having non-zero information values for consecutive days may be appropriate.

[5] We always added the value to the first return in the multiday window. As the event dates are randomly generated, this should not affect the generalizability of our results to situations where the event window is constructed to precede or surround the uncertain event date.

3.1 $H_{01} : \gamma_i = 0 \; \forall i$

The first null hypothesis is that each of the γ_i coefficients in (1) is equal to zero.[6] This null hypothesis is useful in two situations. The first is when the researcher expects some firms to experience positive returns and others to have negative returns to similar announcements. The average of the γ_i's may be insignificantly different from zero, because positive coefficients for some firms are offset by negative coefficients for others. For example, if the event of interest to the accounting researcher is a change in tax law which will benefit some firms, but will harm others, a test for an effect of the law *on average* is likely to lead the researcher to conclude that the change in tax law does not significantly affect firms, while a simultaneous test of the individual coefficients may lead to the opposite conclusion. The second situation is when an announcement has major implications for only one or two companies, and minor implications for others. Again, the *average* effect may be insignificant, but the effect is not insignificant for all firms.

For each simulation, we calculated three F-statistics that have been suggested to test H_{01}. Two of the statistics are from the true SUR model, while the third is from the consecutive equations model. The F-statistics from the true SUR model are both based on the restricted and unrestricted system mean square errors (SMSE) from the SUR system in equation (2). The first F-statistic is that calculated by the SAS statistical package. It is defined

$$\frac{\mathrm{SMSE}_R - \mathrm{SMSE}_U}{q} \bigg/ \frac{\mathrm{SMSE}_U}{(T-k)*N},$$

where $_R$ and $_U$ denote restricted and unrestricted, q is the number of restrictions, N is the number of firms (equations) in the model, T is the number of time-series observations used for each firm, and k is the number of coefficients estimated for each firm. For testing H_{01}, q is equal to the number of firms in the model, corresponding to the restriction under the null that all γ_i in equation (1) are equal to zero. It is distributed asymptotically $\mathrm{F}(N,(T-k)*N)$.

The second F-statistic, which we denote the Schipper and Thompson F-statistic (see Schipper and Thompson [1985]) is also based on the difference in SMSE's, and is exact in situations where the independent variables are the same across all firms—the multivariate case. It is commonly thought to be more conservative than the SAS F. The Schipper-Thompson (hereafter ST) F to test H_{01} is defined to be

$$(\mathrm{SMSE}_R - \mathrm{SMSE}_U) * \left(\frac{T-k-N+1}{(T-k)*N}\right),$$

and is distributed $\mathrm{F}(N,T-k-N+1)$ in the multivariate case.

In cases where the researcher decides not to incorporate estimates of cross-sectional correlation into coefficient estimates or statistical tests, an F-statistic based on the OLS estimates of equation (1) can be used to test the hypotheses. This F-statistic is equivalent to the SAS F statistic when the covariance matrix is diagonal. However, it can also be

[6] The alternative hypothesis is that one or more of the γ_i coefficients is not equal to zero. It is not that none are equal to zero.

expressed as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^2}{\sigma_{\gamma_i}^2},$$

where N is the number of firms, the γ_i are from equation (1), and $\sigma_{\gamma_i}^2$ are the OLS estimates of the variances of the γ_i.[7] This statistic, which we denote the Theil F, is distributed asymptotically F(N,($T-k$) $*$ N).

3.2 $H_{02}: \sum_{i=1}^{N}\gamma_i = 0$

The second null hypothesis is that the average[8] stock price response to the information events, across all firms and events, is equal to zero; i.e., there is no information disclosed in the information releases that is associated with stock price changes, on average. In cases where the effects of the information events are expected to be in the same direction for all firms, this may provide a more powerful test than the previous test. This is because the accumulation of many small effects, none significant in themselves, may be significant in total.[9]

To test H_{02}, we again calculated two F-statistics from the true SUR model. The SAS F used to test H_{02} is again defined as

$$\frac{\mathrm{SMSE}_R - \mathrm{SMSE}_U}{q} \bigg/ \frac{\mathrm{SMSE}_U}{(T-k)*N}.$$

For testing H_{02}, q is one, the only constraint under the null being that the sum of the γ_i coefficients be zero. This statistic is distributed asymptotically F(1,($T-k$) $*$ N).

The Schipper and Thompson F-statistic used to test H_{02} is simply

$$(\mathrm{SMSE}_R - \mathrm{SMSE}_U),$$

and is distributed F(1,$T-k$) in the multivariate case.

There are three statistics based on the consecutive equations model that can be used to test H_{02}. The first, which we again call the Theil F, is asymptotically distributed F(1,($T-k$) $*$ N). It can be calculated as

$$\frac{\left(\sum_{i=1}^{N}\gamma_i\right)^2}{\sum_{i=1}^{N}\sigma_{\gamma_i}^2}.$$

The final two statistics, which we call SUMT and SUMC, were used by Malatesta[10] [1986]. SUMT is a sum of t-statistics, divided by the square root of the number of t-statistics

[7] See Theil, pp. 314-317, particularly problem 3.1.

[8] Tests of $\sum_{i=1}^{N}\gamma_i = 0$ and $\frac{1}{N}\sum_{i=1}^{N}\gamma_i = 0$ result in the same F statistic. It is convenient to drop the $\frac{1}{N}$ term.

[9] Earlier abnormal returns studies generally tested this hypothesis. While offsetting effects could be handled by creating separate portfolios for firms expected to be affected positively or negatively, this would still give statistics based on groups of firms, rather than a statistic testing all firms individually, but simultaneously.

[10] SUMT and SUMC are the W** and Z** statistics in Malatesta. McDonald reports the W** statistic, but not the Z** statistic.

in the sum:

$$\text{SUMT} = \left(\sum_{i=1}^{N} \frac{\gamma_i}{\sigma_{\gamma_i}}\right) \Big/ \sqrt{N},$$

where N is the number of firms. SUMT is distributed N(0,1) under the null hypothesis of no stock price effect, on average, associated with the information events, and ignoring the covariance structure of the firm returns.

SUMC is the sum of the γ_i coefficients, divided by the square root of the sum of their OLS variances:

$$\text{SUMC} = \left(\sum_{i=1}^{N} \gamma_i\right) \Big/ \left(\sum_{i=1}^{N} \sigma_{\gamma_i}^2\right)^{(1/2)}.$$

SUMC is also distributed N(0,1), again assuming that firm returns are independent.

4 Results of simulations

In this section, we report on two aspects of the results of the simulations. First, we discuss the Type I error rates of the various statistics for each of the two hypotheses. Next, we discuss the power of the statistics.

4.1 Type I errors: Rejections when no abnormal performance is introduced

4.1.1 $H_{01} : \gamma_i = 0\ \forall i$

In table 1 we present the Type I error rates for the three F-statistics used to test H_{01}. A Type I error occurs when the null hypothesis is rejected when no abnormal performance was introduced into the returns series on the event dates.[11] All of the rejection frequencies are significantly above the nominal levels. Using a normal approximation to the binomial distribution, and n=1649, the 95% confidence intervals should be [3.9%, 6.1%] for the nominal 5% rejection level, and [.5%, 1.5%] for the 1% rejection level. The null hypothesis is rejected far too often. At the nominal 5% level, rejection rates range from 9.2% to 33.9% for the SAS F, 8.2% to 23.5% for the ST F, and 6.9% to 20.3% for the Theil F. The best case at the 5% level occurs for the Theil F statistic for 20 events and 25 firms, using a 1 day window. The rejection rate of 6.9% is about 3.5 standard deviations away from the 5% level. For the nominal 1% level, rejection rates are also high. For the SAS F, rejection rates range from a low of 3.5% to a high of 19.5%. Similar ranges for the ST F and the Theil F are 2.9% to 11.9% and 2.7% to 9.6%. The best case at the 1% level again occurs for the

[11] It may be that the apparent over-rejections are due to our having non-zero elements in the A_i vectors that happen to coincide with dates on which there actually were significant information events for the firms used in the simulations. However, that is a strength of basing the simulations on real returns data, not a weakness. The researcher can never get a truly clean sample, where nothing else has affected the firms on the identified event dates. Only by using returns drawn from such an information rich environment can we get a realistic idea about how the statistical methods will behave in practice.

Theil F, for 20 events and 25 or 75 firms, using a 1 day window. The rejection rate of 2.7% is almost 7 standard deviations from the nominal 1% level.

Note that the 2 event scenarios are independent of the 5 event scenarios, while the 5 event scenarios are selected by choosing every 4th event from the 20 event scenarios. Hence, the 5 and 20 event scenarios are not independent. Also, the statistics themselves are not independent. They are calculated from the same models, using the same returns and simulated events. The statistics differ in whether they incorporate estimates of cross-sectional covariance. They also differ as to degrees of freedom used in finding critical values used to determine significance levels of the calculated statistics, although the statistics may be asymptotically nearly the same.

There are a few regularities to be seen across the three panels. First, the more firms there are in the model, the worse over-rejection is. This holds true across all windows for almost all numbers of events. Second, the Theil F statistic almost always has the lowest over-rejection rate, followed by the ST F, followed by the SAS F, regardless of number of events, firms, or window length. Third, for a given window length and number of firms, the highest over-rejection rate is generally for the 5 event scenario. Generally, the next highest is with 2 events, with 20 events being the lowest, but still very high compared to the nominal Type I error rates. The "best" window length is 5 for scenarios with 2 or 5 events, but is 1 for scenarios with 20 events.

4.1.2 $H_{02}: \sum_{i=1}^{N} \gamma_i = 0$

Table 2 contains the empirical Type I error rates for the 5 statistics used to test H_{02}, that the average of the γ_i coefficients is zero. In general, the empirical Type I error rates for all of these statistics are much closer to the nominal levels than the error rates for the statistics used to test H_{01}. For scenarios with 2 or 20 events, only 14 out of the 180 different statistics are outside the 95% confidence intervals. If the statistics were independent, we would expect 9 to be outside the limits of a 95% confidence interval, strictly due to chance, and these statistics are not independent. Across all scenarios and all 5 statistics, rejection rates at the 5% level, 2 or 20 events, range from 3.5% to 6.9%. At the 1% level, respective values are .4% to 1.7%.

For the scenarios with 5 events, the situation is very different. Only 14 of the 90 statistics presented are within the 95% confidence interval limits. For 5 events, all of the statistics exhibit high Type I error rates, ranging from 5.2% to 17.2% at the nominal 5% level and 1.4% to 6.7% at the 1% level. Most of them lie outside the 95% confidence intervals of [3.9%,6.1%] and [.5%,1.5%]. We have no explanation for why the statistics do relatively well with 2 and 20 events, but very poorly with 5 events. (Recall that the 20 event scenarios essentially take the 5 event scenarios and add an additional 15 events. Therefore, the 5 and 20 event scenarios are not independent. Yet the 20 event statistics appear to be reasonably well specified, while the 5 event statistics reject too often.)

Comparing the true SUR statistics, the rejection rates for the SAS F and the ST F are identical for all scenarios presented. (See Appendix A for an explanation of this apparent equality.) Comparing the statistics calculated using the consecutive equation model, the Theil F and SUMC statistics have essentially identical rejection rates, differing in only 2 cells presented.

At the 5% level, the 95% confidence interval for n=1649 is about 4% to 6%. As shown in table 2, the SUMC statistics reject the null less often than the SUMT statistics but for most of the scenarios with 2 or 20 events per firm, the rejection frequencies for both statistics are within the 95% band. Again, for the 5 events per firm scenarios, both statistics reject significantly more often than would be expected based on the nominal size of the tests. The SUMT rejection rates are particularly high for some scenarios. There is no uniform behavior with respect to the number of firms, or the window length.

Probably these statistics fail (lie outside the 95% confidence band) due to ignoring the covariance structure. The statistics only fail by being above the top of the 95% confidence band, probably indicating that for those cases ignoring the covariance structure produces statistics a little too large.

4.2 Empirical distributions when no abnormal performance is introduced

In this section, we present the first and fifth percentiles of the empirical distributions of the statistics when no abnormal performance is introduced. This is done in order to present meaningful power tables in the next section. Powers of statistical tests must be evaluated for specific Type I error levels. Researchers are generally interested in powers of tests given Type I error rates of 5% or 1%. Since statistics for some scenarios had Type I error rates much higher than indicated by the nominal size, power tables for nominal Type I error rates of 5% and 1% would not be very informative. To provide more meaningful power tables in the next section, we used as critical values the first and fifth percentiles of the empirical distributions generated when no abnormal performance was introduced. We present those empirical values in this section.

Rather than using the calculated F- or z-statistics, we work with the p-values of those statistics. The p-values are those associated with the calculated statistics and their theoretical distributions and degrees of freedom. This allows uniform interpretation of the empirical critical values given. That is, all empirical critical values are p-values, ranging between 0 and 1, rather than being points from different distributions with different degrees of freedom.

Tables 3 and 4 present the first and fifth percentiles of the empirical distributions of the p-values of the statistics generated in the simulations where no abnormal performance was introduced. Table 3 presents the empirical critical p-values for the statistics used to test $H_{01} : \gamma_i = 0 \ \forall i$; table 4 presents similar values for testing $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$.

Each cell entry is based on the empirical distribution of the 1,649 statistics generated for a specific scenario. Consider the cell in table 3, panel A, for the SAS F statistic testing whether all of the γ_i are equal to 0, for 2 events, 25 firms, window length 1 day, 5% level test. The p-values based on theoretical distributions were obtained for each of the SAS F statistics from the 1,649 simulations using 2 events, 25 firms, and a window length of 1 day, where no abnormal performance was added to the firm return vectors on the (false) event dates. The value of 0.219 indicates that 5% of the p-values for these F statistics were smaller than 0.219%. An F statistic with a p-value larger than 0.219% would not represent a rejection of the null hypothesis at the 5% level, based on the empirical distribution.

Another way of looking at tables 3 and 4 is that if the p-values listed in each cell were used as the critical values for rejecting the null hypothesis, the Type I error rates would be exactly 5% (1%) for all scenarios where no abnormal performance was introduced in the simulation process.

4.2.1 $H_{01} : \gamma_i = 0 \ \forall i$

The high Type I error rates presented in table 1 imply that the first and fifth percentiles of the empirical distributions of nominal p-values must be smaller than 1% and 5%. Table 3 shows that, except for scenarios with 20 events per firm, the first and fifth percentiles are at least an order of magnitude smaller than the nominal 1% and 5% levels. Of the true SUR statistics, the ST F statistic is in general closer to the nominal values than is the SAS F. Empirical 5% rejection rates were achieved for the ST F by using p-values ranging from .047% to 2.421%. To achieve a 1% rejection rate, p-values ranging from 9E-6% to .404% were needed. For the SAS F, empirical p-values used to achieve 5% (1%) empirical rejection rates ranged from .005% (1E-7%) to 1.868% (.276%). The empirical p-values for the Theil F from the consecutive equations model are closer to the nominal p-values than are the ST F at the 5% nominal level, but neither statistic dominates at the 1% nominal level.

For all three statistics, for a given number of firms and window length, increasing the number of events per firm brings the empirical 5% and 1% points closer to the respective nominal points. However, for a given number of events, adding firms moves the empirical critical values farther from the nominal values. That is, as more firms are added, smaller critical p-values must be used to achieve the same nominal size test. Finally, for scenarios with 2 or 5 events per firm, increasing the window length brings the empirical critical p-values up towards the nominal values, but for the 20 event scenarios, increasing window lengths push the empirical critical p-values down, farther from the nominal values.

4.2.2 $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$

Table 4 presents the first and fifth percentiles of the distributions of p-values for the statistics used to test whether the average of the coefficients is significantly different from zero. These points from the empirical distributions are closer to the nominal values than were the respective points for the statistics testing H_{01}. The empirical values are generally below the nominal values, but not always. For scenarios with 20 events per firm, the empirical values are sometimes larger than the nominal values.

For a given number of firms and window length, scenarios with 2 or 20 events have empirical values close to the nominal values, while the empirical values for 5 event scenarios are generally smaller. There is no consistency in the effect of adding firms for a given number of events and window length. For 2 event scenarios, for a given number of firms, increasing the window length brings the empirical critical values closer to the nominal values. There is no such consistency for the 5 and 20 event scenarios.

In summary, from both the tables of Type I error rates and the tables of the 5% and 1% points of the empirical distributions, it is clear that all of the statistics testing H_{01} reject too often when no abnormal performance was introduced. The rejection rates for the statistics used to test H_{02} are much closer to the nominal levels, although there is a slight tendency to over-reject here as well. The next section looks at the other side of the coin—how often the statistics reject the null of no abnormal performance when abnormal performance was added to the returns.

4.3 Power: Rejections of the null when abnormal performance is introduced

Given that the Type I error rates are generally excessive for the statistics examined, simply calculating power as the percentage of rejections at the nominal 5% and 1% levels, when abnormal performance is introduced, would be misleading. Therefore, we present in tables 5 through 10 corrected power tables. The critical p-values used to determine rejection or non-rejection of the null hypothesis are those presented in tables 3 and 4, discussed in the previous section. All corrected power tables contain three panels, as three different ranges of abnormal performance were used in the simulations used to assess power. Abnormal performance metrics used in the simulations reported in Panels A were drawn from a uniform distribution over the range [.00125,.00375], with a mean of .0025. The second range, reported in panels B, was [.00375,.00625], with a mean of .005. The third range also had a mean of .005, but had a larger range, [.0025,.0075].

For all statistics for both hypotheses, for a given number of firms, number of events per firm, and window length, the power of the tests increased as the mean level of abnormal return added on the event dates increased. That is, rejection rates for all scenarios and all statistics, panels A, where the abnormal returns have a mean of .0025, are lower than corresponding cells from panels B and C, where the mean of the abnormal returns added is .005, with differing ranges. The statistics are differentially sensitive to the variance of abnormal returns added, as rejection rates are sometimes higher in panels B than C, and sometimes lower.

One final regularity can be seen across all corrected power tables. Up to this point, we have discussed number of events per firm and number of firms in the model as separate variables. We have done this primarily because these are separate concepts in the realm of accounting research. However, from a purely statistical point of view, changes in either (or both) change the total number of events in the model. An additional column has been inserted in table 5, showing the total number of events in any given scenario. For all of the statistics for both tests, the rejection rates generally increase with the number of total events. However, the rates of increase differ across hypotheses tested, and will be discussed under the appropriate hypothesis.

4.3.1 $H_{01}: \gamma_i = 0 \; \forall i$

The first noteworthy item is that, after correcting for the different empirical rejection rates under the null hypothesis by using the scenario specific rejection rates given in table 3, the corrected power rates for the SAS F and the ST F are identical to three decimal places. (See Appendix A for a reconciliation of these statistics.) Therefore, we present a combined table for the SAS and ST F statistics.

Tables 5 and 6 both show that, for a given number of events per firm and window length, adding additional firms increases the power of the tests. However, the increase in power achieved by adding events for a given number of firms is much greater. For example, for the ST F, using 75 firms instead of 25 firms, each with 2 events, increases the rejection rate from 6.2% to 6.6%. However, having 5 events rather than 2 events for 25 firms increases the rejection rate from 6.2% to 11.0%. In terms of total number of events in the model, 2 events, 25 firms has 50 total events. Moving to 2 events, 75 firms gives 150 total events, while moving to 5 events, 25 firms only gives 125 total events. Yet the increase in power is greater by increasing the number of events per firm.

Finally, window length has a dramatic effect on the powers of the test statistics in the instances where the power for the (correctly specified) one day window is reasonably good. For example, the Theil F in panel B of table 6 rejects 45.5% of the time at the 5% level for the 5 events, 75 firms scenario when a one day window is used. That drops to 26% when a two day window is used, and to 11.2% when a five day window is used.

4.3.2 $H_{02}: \sum_{i=1}^{N} \gamma_i = 0$

The corrected power rates for the SAS F and ST F for testing H_{02} are again identical to three decimal places, after correcting for the difference between nominal and empirical Type I error rates. Therefore, we again present a combined table for the SAS and ST F statistics.

As was true for the statistics testing H_{01}, the total number of events in the model is important. However, for H_{01}, the number of events per firm was very important. Here, that seems to be less true. For example, examine table 10, panel A, the 1% column under the one day window. The power with 2 events per firm, 75 firms, is 16.4%. For 5 events, 25 firms, it is 13.1%. The total number of events goes down for the 5 events, 25 firms scenario, and the power declines. However, numbers of events per firm is not unimportant, either. In the 5% column for the same numbers of events, firms, and window length, the rejection rate increases modestly, from 31.4% to 32.9%, even though the total number of events has decreased. Overall, the total number of events seems more important in determining power for tests of H_{02} than H_{01}. In any case, increasing either the number of events for a given number of firms or increasing the number of firms for a given number of events per firm both result in increases in power.

The effect of increases in window length is not as consistent for statistics testing H_{02} as for those testing H_{01}. Increasing the window length results in substantial declines for all scenarios for the lowest level of added abnormal performance. For the higher level of abnormal performance, both ranges, there is substantial decline for scenarios with 2 or 5 events, or 20 events and 25 firms. However, for 20 events and 50 or 75 firms, the decreases in power are small when moving from a 1 day window to a 2 day window. Moving to a 5 day window results in substantial declines for all scenarios.

Finally, after correcting for the difference in Type I error rates, SUMT appears to be the dominant statistic for testing H_{02}. It always has higher power than the other statistics examined.

5 Research design and interpretation in light of simulation results

In light of our simulations, what can one say about designing and interpreting results of accounting research studies using true SUR or the consecutive equations models? Basically, if one wishes to test hypotheses similar to our H_{01}, that all firm-specific response coefficients are equal to zero, one must use something like the models studied here. However, our evidence indicates that one must use *very* conservative rejection regions, or the probability of Type I errors will be large. Furthermore, if one corrects for high Type I error rates by using rejection regions based on the empirical distributions outlined here, power may be poor, especially if the research examines infrequent events. For example, in table 5, panel A, for scenarios with 2 events per firm, 5% level test, rejection rates range from 6.2 to 6.6%, if the exact event date was known. With event date uncertainty, that range declines to 5.1 to 6.4%. Since rejections of the null hypothesis in this table are based on the empirical critical p-values from table 3, approximately 5% of the simulations would have resulted in rejections *even if no abnormal performance had been introduced.* For scenarios with few events per firm, the power is barely above the number of rejections expected in the absence of introduction of any abnormal performance. However, if the study focusses on recurring events such as quarterly announcements, and there is only slight event date uncertainty, the power is much better. For example, in the same column, if there are 20 events per firm, rejection rates range from 42.6% to 61.2%.

If the researcher is interested only in average effects, the true SUR and the consecutive equations statistics are much better behaved. Type I error rates reported in table 2, while still generally too high, are nowhere near as high as reported in table 1 for tests of H_{01}. The powers of the tests are also higher.

If one is only interested in the average effects, the tendency may be to use a more traditional event study method, such as portfolio abnormal return analysis or cross-sectional pooled regressions of abnormal returns on information variables. However, these methods also have drawbacks. In portfolio abnormal return analysis, the levels of (surprise in) accounting information cannot be included as explanatory variables. In cross-sectional pooled regressions, constraining the coefficients to be the same for all firms or for groups of firms can create problems.[12] Our simulations indicate that true SUR or the consecutive

[12] Teets [1992] presents an example where inference from SUR and inference from a pooled cross-sectional regression are opposite. Several differences between SUR and the pooled cross-sectional approach are examined to see which one(s) drive the difference in inference. The constraint of equality of coefficients under the pooled cross-sectional method appears to drive the difference in inference.

equations model, both of which permit inclusion of levels of information and firm-specific coefficients, may be acceptable alternatives, in terms of Type I error rates and power, for tests of average effects.

We chose the numbers of events per firm to be representative of situations encountered in accounting research. First, the two events per firm scenario corresponds to infrequent events, such as management forecasts of earnings. Second, regular accounting announcements occur annually or quarterly. Over a five year period, that will give rise to 5 or 20 events per firm. To avoid problems with structural change in the sample firms, accountants have frequently restricted their analysis to not more than 5 year periods. Our results suggest that the powers of the test statistics are low for scenarios with only 2 events per firm. This suggests that SUR may not have sufficient power to correctly identify significant information events, *after correcting for high Type I error rates*, in studies of infrequently occurring events. However, for studies using quarterly announcements, SUR may work acceptably.[13]

If the researcher has to choose between adding firms to the sample, where each firm will have only a few events, or identifying additional events for firms already in the sample, our results suggest that either strategy will provide gains in power. If the research question has to do with average effects, either strategy may work equally well. However, if the question has to do with simultaneous tests for all firms (our H_{01}), finding additional events for the existent sample firms will give greater increases in power.

We used different window lengths to determine the extent of the problem caused by event date uncertainty when using SUR. Our results suggest the way that event date uncertainty is typically addressed in SUR can severely affect the test statistics. There is a substantial reduction in power upon moving to a 2 day window from a 1 day window, and a further reduction upon moving to a 5 day window. This reduction in power suggests that SUR as typically implemented may not be appropriate when there is major event date uncertainty.[14]

Finally, what do our results imply about the choice between the true SUR and consecutive equations models? On the whole, there doesn't seem to be a lot to recommend one over the other, empirically. For testing H_{01}, the ST F statistic generally has higher corrected power than the Theil F from the consecutive equations model, but in many of the scenarios, they are very similar. In tests of H_{02}, the SUMT statistic from the consecutive equations model had the highest power in all scenarios. The near equality of the methods may be due to the noise in the cross-sectional covariance estimation process offsetting the gains from incorporating covariance estimates in the coefficient estimates and statistical tests.

[13]This discussion assumes that the average abnormal returns associated with the different frequency events are similar. If the infrequent events are associated with large abnormal returns, and the quarterly announcements are associated with small abnormal returns, these conclusions may not hold.

[14]There are several ways that multiple day event windows could be implemented using firm-specific models. First, if the researcher doesn't need to incorporate cross-sectional correlation, firm-specific models based on equation (1) could be estimated with single day returns outside the event periods and multiple day event period returns, using WLS instead of OLS. If the researcher wants to incorporate cross-sectional correlation, and events are on the same day for all firms, SUR can be used after appropriately scaling the multiple day firm and market returns. Finally, if events happen at different times for different firms, leading to nonsynchronous multiple day event period returns, a method suggested by Marais [1986] may be used.

6 Directions for future work

This simulation study of SUR examined situations where events occurred for different firms at (possibly) different times. It did not examine the situation where a sequence of events (possibly) affects a number of firms on the same event dates. It is in this situation that the Schipper Thompson F statistics are theoretically exact. It would be interesting to simulate this situation, to see if the ST statistics have better Type I error rates and power. This would provide evidence on whether the over-rejection when no abnormal performance was introduced is due to the theoretical distributions holding only asymptotically, or due to leptokurtosis of the returns.

Alternatively, one could generate normally distributed data to use in place of the actual returns data used in this study. Simulations based on this data could provide evidence on the effects of non-normality. One could also use much longer time series of generated data, without having to worry about nonstationarity. This could provide evidence on the effects of the distributions holding only asymptotically.

In these simulations, all coefficients were positive, and equal to unity. In this case tests of the second hypothesis, on the average of the coefficients, should be more powerful than tests of the first hypothesis. Simulations where there is variation across the coefficients relating information to stock prices would give additional evidence on the use of SUR in information event studies.

Finally, comparing pooled cross-sectional methods, portfolio abnormal return analysis, and SUR under a variety of simulated conditions might provide evidence about when researchers could rely on a method, and when assumptions of that method are violated seriously enough that different methods must be contemplated.

A Reconcilation of SAS F and ST F

The SAS F used to test $H_{01} : \gamma_i = 0\ \forall i$ is defined as $\frac{\text{SMSE}_R - \text{SMSE}_U}{q} \Big/ \frac{\text{SMSE}_U}{(T-k)*N}$, and is distributed asymptotically $\text{F}(N,(T-k)*N)$, while the ST F is defined $\text{ST} = (\text{SMSE}_R - \text{SMSE}_U) * \left(\frac{T-k-N+1}{(T-k)*N}\right)$ and is distributed asymptotically $\text{F}(N, T-k-N+1)$ in the multivariate case.

Define $\text{SMSEDF} = \frac{\text{SMSE}_U}{(T-k)*N}$. Consider the ratio of the SAS F to the ST F. It is $\frac{1}{\text{SMSEDF}} \frac{1277}{1278-N}$, since T is 1280 for all simulations, and k is 3. Based on our simulations, the means (standard deviations) of the ratio $\frac{\text{SAS F}}{\text{ST F}}$ for $N = 25$, 50, and 75 are 1.0192 (.000022), 1.040 (.000044) and 1.0616 (.000037). The slight variation in the ratio is due to the $\frac{1}{\text{SMSEDF}}$ factor. Asymptotically, $\text{SMSEDF} \to 1$. The mean (standard deviation) of SMSEDF for $N = 25$, based on 59,364 simulations using 25 firms, is .999967 (.000021). Respective numbers for 50 and 75 firms are .999935 (.000042) and .999916 (.000035). The SAS F is essentially a constant multiple of the ST F; the multiple depends only on the N parameter.

The SAS F is compared to the $\text{F}(N,(T-k)*N)$ distribution, and the ST F is compared to the $\text{F}(N, T-k-N+1)$ distribution. However, if the denominator degrees of freedom are large (greater than 1000), the F distribution is essentially only dependent on its numerator, a χ^2 divided by its degrees of freedom. Hence, both the SAS F and ST F are compared to a χ^2_N divided by N. For $N = 25$, 50, and 75, the critical values for a nominal 5% test are 1.506, 1.350, and 1.293.

Since SAS F $\approx$ 1.0192 ST F for $N = 25$, and both statistics are compared to the same critical value, the empirical rejection rates are different, and always higher for SAS F. Assume that a simulation using 25 firms resulted in an ST F of 1.505. Based on the nominal 5% critical value of 1.506, this ST F would not lead to rejection of the null. However, the corresponding SAS F would be approximately 1.535 (since SMSEDF is not exactly 1), and would lead to rejection of the null.

The reason the rejection rates based on the empirical distributions are the same is that the empirical critical values compensate for the factor relating the SAS F to the ST F. Consider the 5% critical values from table 3 for the SAS F and the ST F, for the scenario with 2 events per firm, 25 firms, using a 1 day window. The SAS F critical p-value is .219%, which translates to an F of 1.9973, while the ST F critical p-value is .325%, which has a corresponding F value of 1.9596. The ratio of these critical values is approximately 1.0192, which is the same as the mean multiplication factor (for 25 firms) used to translate the ST F into the SAS F. The empirical critical values offset the multiplication factor by which the two statistics differ.

The SAS F and ST F statistics used to test $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$ are even more closely related. The SAS F is as defined for the test of H_{01}, but q is now equal to unity. The ST F is simply $\text{SMSE}_R - \text{SMSE}_U$. As indicated previously, $\text{SMSEDF} \to 1$ asymptotically. Therefore, the SAS F and ST F have almost identical values. Although the degrees of freedom used to determine critical values differ across the two statistics (1 and $(T-k)*N$ for the SAS F and 1 and $T-k$ for the ST F), the denominator degrees of freedom are large enough both statistics are essentially compared to a χ^2_1. The minor effect of SMSEDF is offset by the small differences in the p-values of the empirical critical values presented in table 4.

References

[1] Stephen J. Brown and Jerold B. Warner. Measuring security price performance. *Journal of Financial Economics*, 205–58, June, 1980.

[2] Stephen J. Brown and Jerold B. Warner. Using daily stock returns. *Journal of Financial Economics*, 14(1):3–31, 1985.

[3] Cynthia J. Campbell and Charles E. Wasley. Measuring security price performance using daily NASDAQ security returns. October 1991. Working paper, Washington University in St. Louis.

[4] C.S. Agnes Cheng, William S. Hopwood, and James C. McKeown. Non-linearity and specification problems in unexpected earnings response regression model. *The Accounting Review*, 67(3):579–598, July 1992.

[5] Daniel W. Collins and Warren T. Dent. A comparison of alternative testing methodologies used in capital market research. *Journal of Accounting Research*, 22(1):48–84, Spring 1984.

[6] Thomas Dyckman, Donna Philbrick, and Jens Stephan. A comparison of event study methodologies using daily stock returns: a simulation approach. *Journal of Accounting Research, Supplement*, 22:1–30, 1984.

[7] P. Malatesta. Measuring abnormal performance: the event parameter approach using joint generalized least squares. *Journal of Financial and Quantitative Analysis*, 21:27–38, March 1986.

[8] M. Laurentius Marais. On detecting abnormal returns to a portfolio of nonsynchronously traded securities. 1986. Working paper, University of Chicago.

[9] Bill McDonald. Event studies and systems methods: some additional evidence. *Journal of Financial and Quantitative Analysis*, 22(4):495–504, December, 1987.

[10] Robert P. Parks and Walter R. Teets. Evidence on misspecification of seemingly unrelated regression statistics. September 1993. Working paper, Washington University in St. Louis.

[11] Grace Pownall, Charles Wasley, and Gregory Waymire. The stock price effects of alternative types of management earnings forecasts. September 1992. Washington University.

[12] Grace Pownall and Gregory Waymire. Voluntary disclosure choice and earnings information transfer. *Journal of Accounting Research*, 27 Supplement:85–105, 1989a.

[13] Grace Pownall and Gregory Waymire. Voluntary disclosure credibility and securities prices: evidence from management earnings forecasts, 1969-1973. *Journal of Accounting Research*, 27(2):227–245, Autumn, 1989b.

[14] Katherine Schipper and Rex Thompson. The impact of merger-related regulations using exact distributions of test statistics. *Journal of Accounting Research*, 23:408–15, 1985.

[15] Walter R. Teets. An empirical investigation into the association between stock market responses to earnings announcements and regulation of electric utilities. Forthcoming in *Journal of Accounting Research*, Autumn, 1992.

[16] Walter R. Teets. A methodological consideration on the choice between event study estimation procedures. November 1991. Working paper, University of Illinois at Urbana-Champaign.

[17] A. Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–368, 1962.

Table 1: Type I error rates: Percentage of 1,649 simulations where $H_{01}: \gamma_i = 0\ \forall i$ was rejected when no abnormal performance was introduced.

Model[a]: $r_{it} = \alpha_i + \beta_i\, r_{mt} + \gamma_i\, A_{it} + \varepsilon_{it}$

Panel A: SAS F Statistic[d]

Number of events/firm	Number of firms (N)	1 Day Window[b]		2 Day Window[b]		5 Day Window[b]	
		5%[c]	1%	5%	1%	5%	1%
2	25	13.2	7.7	12.7	7.3	10.7	5.5
2	50	19.6	11.5	18.1	9.8	13.9	7.8
2	75	25.5	15.0	23.4	13.9	19.3	11.6
5	25	14.9	7.6	15.1	7.0	12.5	4.8
5	50	23.4	12.0	24.9	12.3	18.5	8.4
5	75	31.2	18.3	33.9	19.5	25.5	13.2
20	25	9.2	3.5	11.3	3.8	12.6	4.7
20	50	13.6	5.0	18.0	6.9	16.7	7.3
20	75	21.2	9.7	23.7	9.9	22.0	10.6

Panel B: Schipper-Thompson F Statistic[e]

Number of events/firm	Number of firms (N)	1 Day Window		2 Day Window		5 Day Window	
		5%	1%	5%	1%	5%	1%
2	25	12.0	7.0	11.6	6.9	9.7	5.0
2	50	15.7	9.0	14.4	7.7	11.5	5.9
2	75	17.9	10.4	16.2	9.7	13.7	7.2
5	25	13.7	6.5	13.7	6.0	10.9	4.3
5	50	18.1	8.7	19.3	9.0	13.6	6.1
5	75	21.9	10.7	23.5	11.9	16.6	8.4
20	25	8.2	3.0	9.6	2.9	10.9	3.5
20	50	9.7	3.5	12.7	4.2	13.4	5.0
20	75	12.4	4.1	14.1	5.2	12.9	5.5

Panel C: Theil F-statistic[f]

Number of events/firm	Number of firms (N)	1 Day Window		2 Day Window		5 Day Window	
		5%	1%	5%	1%	5%	1%
2	25	12.2	7.0	11.0	6.6	9.3	4.8
2	50	14.3	8.5	13.6	7.5	10.4	5.4
2	75	14.5	8.7	13.2	7.9	11.9	6.2
5	25	12.9	6.4	13.3	5.8	11.5	5.0
5	50	14.9	7.9	17.5	8.4	14.7	7.2
5	75	17.2	8.6	20.3	9.6	16.1	7.9
20	25	6.9	2.7	9.4	3.0	11.7	4.1
20	50	8.5	3.0	11.2	3.5	12.9	4.7
20	75	8.9	2.7	10.9	3.8	11.0	3.8

[a] Models were estimated using 1,280 daily firm (r_{it}) and market (r_{mt}) returns from the periods 12/06/74–12/31/79, 12/10/79–12/31/84, or 12/06/84–12/29/89. Within a simulation, all returns were from the same time period. To simulate accounting information events, randomly selected elements of the A_{it} vector were assigned randomly generated non-zero values from one of the ranges [.00125,.00375], [.00375,.00625], or [.0025,.0075]. Values varied across firms and events, but all were generated from the same range within a simulation. In simulations used to determine powers of the tests, abnormal performance was introduced into firms' returns by adding the non-zero values to the firms' returns corresponding to the first date in each event window.

[b] For an x day window, the A_i vector has x consecutive nonzero elements for each event. Under the null hypothesis, no abnormal returns are added to the returns vectors.

[c] Nominal size of the test—the expected percentage of rejections of H_{01} due to chance when no abnormal performance is introduced, if the test statistic is well specified.

[d] The SAS F-statistic is defined as

$$\frac{\text{SMSE}_R - \text{SMSE}_U}{q} \bigg/ \frac{\text{SMSE}_U}{(T-k)*N},$$

where SMSE denotes system mean squared error, $_R$ and $_U$ denote restricted and unrestricted, q is the number of restrictions, N is the number of firms (equations) in the model, T is the number of time-series observations used for each firm, and k is the number of coefficients estimated for each firm. The statistic to test H_{01} is asymptotically distributed $\text{F}(N, 1277*N)$.

[e] The Schipper-Thompson F-statistic is

$$(\text{SMSE}_R - \text{SMSE}_U) * \left(\frac{T-k-N+1}{(T-k)*N}\right),$$

and is asymptotically distributed $\text{F}(N,1277-N+1)$.

[f] The Theil F-statistic is

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_i^2}{\sigma_{\gamma_i}^2},$$

where $\sigma_{\gamma_i}^2$ are the OLS estimates of the variances of the γ_i. The statistic to test H_{01} is asymptotically distributed $\text{F}(N,1277*N)$.

Table 2: Type I error rates: Percentage of 1,649 simulations where $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$ was rejected when no abnormal performance was introduced

Model: $r_{it} = \alpha_i + \beta_i r_{mt} + \gamma_i A_{it} + \varepsilon_{it}$

Panel A: SAS F Statistic[a]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	5.6	1.7	5.5	1.5	4.5	1.1
2	50	5.6	1.6	5.1	1.3	4.7	1.0
2	75	5.9	1.4	5.2	1.4	4.5	1.0
5	25	5.3	2.0	6.1	1.4	5.8	1.6
5	50	6.8	1.6	6.5	1.8	6.8	1.8
5	75	6.4	2.2	6.5	1.5	7.9	2.2
20	25	5.4	0.9	6.2	0.7	4.8	1.0
20	50	5.3	1.2	5.5	1.0	5.4	1.3
20	75	5.9	1.3	5.9	1.3	5.3	1.0

Panel B: Schipper Thompson F Statistic[b]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	5.6	1.7	5.5	1.5	4.5	1.1
2	50	5.6	1.6	5.1	1.3	4.7	1.0
2	75	5.9	1.4	5.2	1.4	4.5	1.0
5	25	5.3	2.0	6.1	1.4	5.8	1.6
5	50	6.8	1.6	6.5	1.8	6.8	1.8
5	75	6.7	2.2	6.5	1.5	7.9	2.2
20	25	5.4	0.9	6.2	0.7	4.8	1.0
20	50	5.3	1.2	5.5	1.0	5.4	1.3
20	75	5.9	1.3	5.9	1.3	5.3	1.0

Panel C: Theil F Statistic[c]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	5.4	1.2	5.0	1.3	4.1	1.2
2	50	4.4	1.1	4.4	0.9	3.7	0.8
2	75	4.5	0.7	4.4	0.8	3.5	0.6
5	25	5.2	1.8	5.8	1.7	6.7	2.1
5	50	6.3	2.2	7.3	2.4	10.0	3.2
5	75	7.0	2.0	7.7	2.4	12.4	4.4
20	25	4.7	0.6	5.2	0.4	4.5	1.0
20	50	4.5	1.0	4.7	0.8	4.9	1.0
20	75	4.9	0.8	4.9	0.8	4.6	1.0

Panel D: SUMT[d]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	6.9	1.3	6.2	1.2	5.0	1.5
2	50	5.1	1.1	5.2	1.0	4.5	1.2
2	75	5.6	1.0	5.4	0.8	4.1	0.7
5	25	6.5	1.7	7.2	2.1	8.6	3.0
5	50	7.0	2.5	9.3	3.0	13.2	4.9
5	75	7.0	2.4	10.5	2.7	17.2	6.7
20	25	5.0	1.0	4.9	1.0	5.8	1.3
20	50	4.9	0.8	4.6	1.1	5.6	1.5
20	75	4.1	1.0	5.4	1.0	6.6	1.4

Panel E: SUMC[e]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	5.4	1.4	5.0	1.3	4.1	1.2
2	50	4.4	1.1	4.4	0.9	3.7	0.8
2	75	4.5	0.7	4.4	0.8	3.5	0.6
5	25	5.2	1.8	5.8	1.7	6.7	2.1
5	50	6.3	2.2	7.3	2.4	10.0	3.2
5	75	7.0	2.0	7.7	2.4	12.4	4.4
20	25	4.7	0.6	5.2	0.5	4.5	1.0
20	50	4.5	1.0	4.7	0.8	4.9	1.0
20	75	4.9	0.8	4.9	0.8	4.6	1.0

See notes to table 1 for definitions of symbols and description of model.

[a] See notes to table 1 for formula. The statistic to test H_{02} is asymptotically distributed F(1,$(T-k)*I$).

[b] Statistic is calculated as $\text{SMSE}_R - \text{SMSE}_U$, and is asymptotically distributed F(1,$T-k$).

[c] Statistic is calculated $\left(\sum_{i=1}^{N}\gamma_i\right)^2 \Big/ \sum_{i=1}^{N}\sigma^2_{\gamma_i}$, and is distributed asymptotically F(1,$(T-k)*I$).

[d] Statistic is calculated $\left(\sum_{i=1}^{N}\frac{\gamma_i}{\sigma_{\gamma_i}}\right) \Big/ \sqrt{N}$, and is distributed asymptotically N(0,1).

[e] Statistic is calculated $\left(\sum_{i=1}^{N}\gamma_i\right) \Big/ \left(\sum_{i=1}^{N}\sigma^2_{\gamma_i}\right)^{(1/2)}$ and is distributed asymptotically N(0,1).

Table 3: First and fifth percentiles of empirical distributions of p-values of statistics testing $H_{01} : \gamma_i = 0 \; \forall i$. Each distribution based on 1,649 simulations where no abnormal performance was introduced.

Model: $r_{it} = \alpha_i + \beta_i \, r_{mt} + \gamma_i \, A_{it} + \varepsilon_{it}$

Panel A: SAS F statistic

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	0.219[b]	<0.001	0.190	<0.001	0.735	<0.001
2	50	0.017	<0.001	0.063	<0.001	0.272	<0.001
2	75	0.005	<0.001	0.007	<0.001	0.062	<0.001
5	25	0.347	0.002	0.485	0.001	1.041	0.001
5	50	0.083	<0.001	0.089	<0.001	0.239	0.002
5	75	0.014	<0.001	0.017	<0.001	0.037	<0.001
20	25	1.868	0.128	1.430	0.276	1.047	0.093
20	50	0.966	0.050	0.580	0.029	0.438	0.010
20	75	0.297	0.023	0.211	0.009	0.198	0.003

Panel B: Schipper-Thompson F Statistic

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	0.325	<0.001	0.285	<0.001	1.010	0.001
2	50	0.059	<0.001	0.186	<0.001	0.658	0.001
2	75	0.047	<0.001	0.063	<0.001	0.357	<0.001
5	25	0.501	0.004	0.685	0.002	1.400	0.002
5	50	0.236	0.001	0.251	0.002	0.589	0.010
5	75	0.106	0.001	0.129	0.002	0.237	0.004
20	25	2.421	0.197	1.885	0.404	1.408	0.146
20	50	1.974	0.151	1.271	0.096	0.996	0.036
20	75	1.240	0.162	0.947	0.078	0.898	0.033

Panel C: Theil F Statistic

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	0.386	<0.001	0.367	<0.001	1.140	<0.001
2	50	0.126	<0.001	0.218	<0.001	0.758	0.001
2	75	0.090	<0.001	0.092	<0.001	0.600	<0.001
5	25	0.464	0.002	0.707	0.002	0.997	0.002
5	50	0.247	<0.001	0.262	0.001	0.481	0.003
5	75	0.189	0.001	0.172	0.001	0.215	0.007
20	25	2.964	0.156	1.968	0.340	1.517	0.116
20	50	2.582	0.128	1.703	0.110	1.219	0.079
20	75	2.195	0.227	1.638	0.096	1.455	0.029

See notes to table 1 for definitions.

Table 4: First and fifth percentiles of empirical distributions of p-values of statistics testing $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$. Each distribution based on 1,649 simulations where no abnormal performance was introduced.

Model: $r_{it} = \alpha_i + \beta_i \, r_{mt} + \gamma_i \, A_{it} + \varepsilon_{it}$

Panel A: SAS F Statistic

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	4.044[a]	0.424	4.626	0.634	6.023	0.686
2	50	4.154	0.810	4.947	0.815	5.237	1.022
2	75	4.194	0.791	4.795	0.768	5.349	0.885
5	25	4.752	0.408	4.114	0.414	4.342	0.376
5	50	3.530	0.497	3.872	0.458	3.592	0.518
5	75	3.622	0.269	3.574	0.756	2.824	0.352
20	25	4.435	1.156	4.331	1.296	5.313	0.904
20	50	4.661	0.852	4.649	1.063	4.574	0.699
20	75	4.285	0.707	4.260	0.650	4.423	1.019

Panel B: Schipper-Thompson F Statistic

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	4.064	0.430	4.646	0.642	6.046	0.695
2	50	4.174	0.820	4.968	0.825	5.259	1.033
2	75	4.215	0.801	4.817	0.778	5.372	0.895
5	25	4.772	0.415	4.134	0.421	4.363	0.382
5	50	3.549	0.504	3.892	0.466	3.612	0.526
5	75	3.643	0.275	3.594	0.766	2.842	0.358
20	25	4.456	1.168	4.351	1.308	5.334	0.915
20	50	4.683	0.862	4.671	1.074	4.596	0.709
20	75	4.306	0.716	4.281	0.659	4.445	1.030

Panel C: Theil F Statistic

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	4.614	0.483	5.015	0.519	6.070	0.664
2	50	5.434	0.948	5.685	1.032	6.937	1.386
2	75	5.375	1.330	5.686	1.106	6.788	1.717
5	25	4.657	0.356	4.179	0.369	3.634	0.267
5	50	3.340	0.385	2.932	0.234	1.674	0.165
5	75	3.297	0.356	2.653	0.480	1.160	0.090
20	25	5.477	1.375	4.934	1.578	5.509	1.046
20	50	5.680	0.922	5.153	1.354	5.158	1.000
20	75	5.329	1.228	5.091	1.044	5.174	1.002

Panel D: SUMT

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	3.903	0.880	3.841	0.804	5.004	0.576
2	50	4.586	0.895	4.852	1.002	5.545	0.853
2	75	4.102	0.924	4.900	1.344	5.703	1.435
5	25	3.582	0.399	3.506	0.407	2.567	0.157
5	50	2.498	0.300	2.072	0.235	1.047	0.094
5	75	3.376	0.302	2.047	0.462	0.579	0.061
20	25	4.959	1.054	5.252	0.945	4.420	0.644
20	50	5.187	1.145	5.233	0.961	4.527	0.513
20	75	6.162	1.060	4.490	0.996	3.739	0.733

Panel E: SUMC

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	4.618	0.488	5.040	0.524	6.127	0.674
2	50	5.436	0.989	5.688	1.040	6.979	1.418
2	75	5.400	1.340	5.702	1.145	6.923	1.720
5	25	4.664	0.367	4.225	0.414	3.642	0.298
5	50	3.360	0.391	2.940	0.264	1.690	0.172
5	75	3.344	0.432	2.655	0.492	1.173	0.094
20	25	5.506	1.376	4.936	1.607	5.527	1.103
20	50	5.706	1.014	5.193	1.378	5.188	1.067
20	75	5.361	1.238	5.100	1.059	5.245	1.007

See notes to table 1 for model description.
See notes to table 2 for statistic definitions.
[a]All p-values are expressed as percentages

Table 5: Power: Percentage of 1,649 simulations where $H_{01} : \gamma_i = 0 \; \forall i$ was rejected by the SAS and Schipper Thompson F-statistics when abnormal performance was introduced Rejections based on empirical distributions summarized in Table 3

Model: $r_{it} = \alpha_i + \beta_i \, r_{mt} + \gamma_i \, A_{it} + \varepsilon_{it}$

Panel A: Level 1, Abnormal returns from [.00125,.00375]

Number of events/firm	Number of firms (N)	Events × firms	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	50	6.2	1.3	5.1	1.0	5.2	1.1
2	50	100	6.4	1.3	6.4	1.2	5.6	1.0
2	75	150	6.6	1.2	5.9	1.2	6.0	1.3
5	25	125	11.0	2.5	8.9	1.3	6.7	1.2
5	50	250	15.5	3.2	9.9	2.1	6.4	1.6
5	75	375	17.6	4.7	10.9	2.7	7.0	1.6
20	25	500	42.6	27.6	23.9	14.4	10.4	3.0
20	50	1000	56.2	40.6	34.9	20.1	14.4	3.1
20	75	1500	61.2	48.9	40.2	25.6	16.9	4.1

Panel B: Level 2, Abnormal returns from [.00375,.00625]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	11.5	1.5	6.7	1.4	5.8	1.5
2	50	13.8	2.9	9.2	1.5	6.9	1.3
2	75	19.5	3.9	9.9	1.5	7.3	1.3
5	25	35.7	15.1	20.7	4.4	11.8	1.5
5	50	48.2	27.9	30.1	11.9	13.8	3.2
5	75	52.5	35.7	34.8	18.3	14.0	3.9
20	25	87.0	73.9	62.3	51.1	31.5	18.2
20	50	94.5	89.5	76.0	60.2	41.3	25.7
20	75	97.2	94.2	81.5	70.5	47.6	32.0

Panel C: Level 3, Abnormal returns from [.0025,.0075]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	12.7	1.8	6.9	1.6	5.9	1.3
2	50	15.7	4.0	10.2	1.6	6.9	1.5
2	75	20.4	4.9	10.6	1.8	7.2	1.5
5	25	37.6	17.2	21.5	4.9	12.3	1.5
5	50	49.9	30.0	31.5	13.3	14.7	3.8
5	75	54.4	37.7	36.6	20.3	15.7	4.6
20	25	87.9	76.8	64.4	52.7	32.7	19.7
20	50	95.5	90.4	78.0	62.6	43.0	27.2
20	75	97.5	95.3	83.0	73.0	48.5	33.4

See notes to table 1 for definitions.

Table 6: Power: Percentage of 1,649 simulations where $H_{01} : \gamma_i = 0\ \forall i$ was rejected by Theil F-statistic when abnormal performance was introduced
Rejections based on empirical distributions summarized in Table 3

Model: $r_{it} = \alpha_i + \beta_i r_{mt} + \gamma_i A_{it} + \varepsilon_{it}$

Panel A: Level 1, Abnormal returns from [.00125,.00375]

Number of events/firm	Number of firms (N)	1 Day Window		2 Day Window		5 Day Window	
		5%	1%	5%	1%	5%	1%
2	25	6.1	1.3	5.7	1.0	5.3	1.0
2	50	7.0	1.4	5.8	1.2	5.6	1.0
2	75	6.8	1.3	5.6	1.1	5.8	1.1
5	25	10.7	1.9	9.2	1.6	6.9	1.2
5	50	15.0	2.2	9.9	1.6	7.5	1.3
5	75	18.7	4.8	11.6	2.5	7.2	2.4
20	25	42.6	25.6	23.7	12.9	11.1	2.9
20	50	54.0	37.7	33.5	18.2	13.6	4.0
20	75	59.9	46.0	40.1	22.4	16.2	3.3

Panel B: Level 2, Abnormal returns from [.00375,.00625]

Number of events/firm	Number of firms (N)	1 Day Window		2 Day Window		5 Day Window	
		5%	1%	5%	1%	5%	1%
2	25	9.0	1.3	6.4	1.5	5.5	1.0
2	50	9.6	1.9	7.2	1.4	6.5	1.1
2	75	11.3	2.2	6.9	1.2	6.9	1.2
5	25	24.4	5.8	16.2	2.5	9.6	1.5
5	50	38.5	11.4	20.1	3.7	10.9	2.2
5	75	45.5	21.8	26.0	7.6	11.2	3.8
20	25	86.0	69.3	56.1	41.7	21.9	8.8
20	50	95.0	87.0	72.8	55.1	31.2	14.8
20	75	97.1	94.1	79.3	64.2	39.5	15.3

Panel C: Level 3, Abnormal returns from [.0025,.0075]

Number of events/firm	Number of firms (N)	1 Day Window		2 Day Window		5 Day Window	
		5%	1%	5%	1%	5%	1%
2	25	12.6	1.9	7.7	1.5	6.3	1.2
2	50	16.1	3.1	10.1	1.5	7.0	1.5
2	75	19.5	4.4	9.3	1.6	7.6	1.5
5	25	37.5	15.3	22.7	5.6	11.9	1.8
5	50	48.8	26.9	32.3	10.0	15.7	3.2
5	75	54.5	36.5	37.3	19.2	16.3	7.1
20	25	88.1	74.3	63.0	51.2	32.4	17.9
20	50	95.9	89.1	77.5	61.1	42.6	28.0
20	75	97.6	94.6	82.7	69.8	48.6	30.7

See notes to table 1 for definitions.

Table 7: Power: Percentage of 1,649 simulations where $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$ was rejected by the SAS and Schipper Thompson F-statistic when abnormal performance was introduced
Rejections based on empirical distributions summarized in Table 4

Model: $r_{it} = \alpha_i + \beta_i \, r_{mt} + \gamma_i \, A_{it} + \varepsilon_{it}$

Panel A: Level 1, Abnormal returns from [.00125,.00375]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	13.6	4.3	10.0	3.3	7.2	1.5
2	50	23.2	10.7	15.9	5.5	9.3	2.9
2	75	31.3	15.8	20.2	7.6	10.9	2.7
5	25	32.9	13.8	20.9	6.2	13.6	3.3
5	50	50.4	32.4	34.0	16.1	20.4	7.7
5	75	60.1	37.2	43.6	26.8	24.3	9.6
20	25	67.3	51.6	44.5	33.1	24.1	11.7
20	50	87.0	74.1	66.6	51.0	36.5	20.3
20	75	92.9	83.7	75.1	59.6	47.1	31.5

Panel B: Level 2, Abnormal returns from [.00375,.00625]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	41.0	19.1	26.4	11.1	14.3	3.5
2	50	60.5	44.5	43.5	24.8	21.6	10.0
2	75	69.5	55.4	52.6	33.7	30.4	13.6
5	25	70.6	46.9	50.9	28.1	30.5	11.0
5	50	86.4	73.1	70.2	50.5	48.0	28.0
5	75	92.9	80.1	79.4	67.7	56.1	36.2
20	25	95.5	91.4	82.7	74.0	57.1	39.4
20	50	99.5	98.5	95.6	90.5	75.4	58.5
20	75	99.8	99.8	98.2	95.1	85.3	73.5

Panel C: Level 3, Abnormal returns from [.0025,.0075]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	38.9	17.8	25.5	10.4	13.5	3.3
2	50	58.5	41.5	41.0	24.6	19.5	9.5
2	75	67.7	52.8	49.8	32.1	28.9	12.5
5	25	71.4	48.2	51.8	28.1	30.7	11.3
5	50	86.5	73.1	70.5	51.4	48.5	28.2
5	75	93.1	80.5	80.5	67.9	56.1	36.9
20	25	95.9	92.5	84.1	75.4	58.2	41.4
20	50	99.6	98.8	96.2	91.5	77.6	60.3
20	75	99.8	99.8	98.5	95.8	86.8	75.3

See notes to table 2 for definitions.

Table 8: Power: Percentage of 1,649 simulations where $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$ was rejected by Theil F-statistic when abnormal performance was introduced
Rejections based on empirical distributions summarized in Table 4

Model: $r_{it} = \alpha_i + \beta_i\, r_{mt} + \gamma_i\, A_{it} + \varepsilon_{it}$

Panel A: Level 1, Abnormal returns from [.00125,.00375]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	14.5	4.2	10.3	2.4	6.9	1.5
2	50	24.8	10.2	16.5	5.5	10.2	3.2
2	75	31.4	16.4	20.6	8.0	11.2	3.8
5	25	32.9	13.1	21.4	6.7	13.5	3.1
5	50	50.9	29.3	33.7	13.3	17.0	6.4
5	75	61.8	39.5	43.0	24.7	21.2	7.6
20	25	68.0	52.6	44.9	32.3	23.2	10.7
20	50	87.1	71.7	64.0	50.2	34.2	19.0
20	75	92.6	84.9	73.6	58.5	41.6	25.1

Panel B: Level 2, Abnormal returns from [.00375,.00625]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	33.4	11.7	20.6	5.7	11.1	2.1
2	50	55.4	33.9	34.7	16.6	18.1	7.0
2	75	66.0	49.5	43.8	23.9	23.7	10.2
5	25	67.3	36.7	44.1	18.8	25.3	7.3
5	50	85.7	67.0	65.1	39.2	36.4	16.3
5	75	92.8	81.7	77.0	61.0	44.3	21.7
20	25	95.5	91.0	80.8	69.9	48.9	29.7
20	50	99.5	98.6	95.1	89.7	68.9	51.1
20	75	99.8	99.8	98.2	95.6	80.8	64.7

Panel C: Level 3, Abnormal returns from [.0025,.0075]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	39.0	17.5	25.4	9.2	13.0	3.1
2	50	60.2	41.0	41.2	22.7	21.8	9.8
2	75	67.7	54.3	49.5	31.4	29.5	14.7
5	25	71.4	47.4	52.6	27.2	31.0	10.7
5	50	87.0	70.3	70.3	47.5	44.3	23.2
5	75	93.6	82.7	79.0	65.9	51.5	31.7
20	25	96.1	92.2	84.1	75.3	57.4	40.8
20	50	99.6	98.7	96.2	91.0	73.6	59.6
20	75	99.9	99.8	98.5	96.2	83.1	69.6

See notes to table 2 for definitions.

Table 9: Power: Percentage of 1,649 simulations where $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$ was rejected by SUMT statistic when abnormal performance was introduced
Rejections based on empirical distributions summarized in Table 4

Model: $r_{it} = \alpha_i + \beta_i\, r_{mt} + \gamma_i\, A_{it} + \varepsilon_{it}$

Panel A: Level 1, Abnormal returns from [.00125,.00375]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	021.0	09.3	012.9	05.8	08.2	02.1
2	50	36.5	21.2	23.1	10.1	12.9	3.5
2	75	44.6	28.4	29.7	16.7	16.7	6.7
5	25	45.3	23.1	30.1	12.1	16.4	4.6
5	50	63.3	43.5	44.8	25.2	23.2	10.2
5	75	75.7	55.8	56.2	40.3	29.3	14.4
20	25	81.4	67.1	60.3	43.7	31.0	14.3
20	50	94.5	89.5	78.7	63.3	46.9	27.0
20	75	98.6	94.7	86.2	75.3	56.1	40.4

Panel B: Level 2, Abnormal returns from [.00375,.00625]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	54.9	38.1	35.1	20.4	19.6	5.3
2	50	74.1	58.8	56.0	38.4	32.1	14.6
2	75	82.2	69.6	65.9	51.8	41.9	26.2
5	25	80.7	64.2	64.6	44.0	39.5	15.2
5	50	93.1	84.8	79.1	64.0	55.2	33.4
5	75	98.5	92.5	87.8	79.4	61.0	46.0
20	25	99.2	97.8	94.1	85.7	67.1	50.3
20	50	100.0	99.9	99.3	97.5	86.1	69.7
20	75	100.0	99.9	99.9	99.3	92.3	83.1

Panel C: Level 3, Abnormal returns from [.0025,.0075]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	54.8	39.7	35.7	21.0	19.6	5.6
2	50	74.8	60.5	56.8	39.7	33.3	14.5
2	75	83.4	70.5	67.3	52.8	43.3	27.3
5	25	81.6	65.9	65.8	45.2	40.9	15.4
5	50	93.5	85.7	80.3	65.5	56.4	34.9
5	75	98.7	93.5	88.9	80.5	62.0	47.2
20	25	99.4	98.2	95.0	86.8	69.3	51.5
20	50	100.0	99.9	99.5	97.8	87.9	71.9
20	75	100.0	99.9	99.9	99.5	93.3	85.0

See notes to table 2 for definitions.

Table 10: Power: Percentage of 1,649 simulations where $H_{02} : \sum_{i=1}^{N} \gamma_i = 0$ was rejected by SUMC statistic when abnormal performance was introduced Rejections based on empirical distributions summarized in Table 4

Model: $r_{it} = \alpha_i + \beta_i r_{mt} + \gamma_i A_{it} + \varepsilon_{it}$

Panel A: Level 1, Abnormal returns from [.00125,.00375]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	14.5	4.2	10.4	2.4	7.0	1.5
2	50	24.8	10.4	16.5	5.5	10.2	3.2
2	75	31.4	16.4	20.6	8.2	11.3	3.8
5	25	32.9	13.1	21.5	7.1	13.6	3.3
5	50	51.1	29.8	33.9	13.7	17.0	6.4
5	75	62.0	40.9	43.0	24.9	21.3	7.8
20	25	68.0	52.6	44.9	32.8	23.2	10.9
20	50	87.2	72.7	64.0	50.3	34.3	19.5
20	75	92.6	84.9	73.6	58.5	41.8	25.1

Panel B: Level 2, Abnormal returns from [.00375,.00625]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	41.3	18.8	27.2	9.2	13.8	3.0
2	50	62.3	44.1	43.2	24.3	23.4	10.4
2	75	69.3	57.0	52.3	33.7	31.2	16.6
5	25	71.3	46.6	52.7	28.1	31.4	11.0
5	50	86.7	70.2	69.4	48.2	44.3	24.4
5	75	93.1	83.5	78.8	66.1	52.0	32.0
20	25	95.6	91.6	82.8	73.8	55.7	39.4
20	50	99.5	98.6	95.3	90.3	72.3	57.4
20	75	99.8	99.8	98.2	95.6	81.7	67.8

Panel C: Level 3, Abnormal returns from [.0025,.0075]

Number of events/firm	Number of firms (N)	1 Day Window 5%	1 Day Window 1%	2 Day Window 5%	2 Day Window 1%	5 Day Window 5%	5 Day Window 1%
2	25	39.0	17.5	25.4	9.2	13.0	3.1
2	50	60.3	41.2	41.2	22.8	21.9	9.9
2	75	67.7	54.4	49.5	31.8	29.8	14.7
5	25	71.5	47.6	52.9	27.9	31.0	11.3
5	50	87.1	70.5	70.3	48.7	44.5	23.4
5	75	93.7	83.7	79.0	66.0	51.7	31.8
20	25	96.1	92.2	84.1	75.6	57.4	41.2
20	50	99.6	98.8	96.2	91.0	73.6	60.0
20	75	99.9	99.8	98.5	96.2	83.2	69.7

See notes to table 2 for definitions.