Review of Accounting and Finance

Emerald Article: More power to you: properties of a more powerful event study methodology

Tarcisio da Graca, Robert Masson

More power to you: properties of a more powerful event study methodology

Tarcisio da Graca
Department of Economics, University of Ottawa, Ottawa, Canada, and
Robert Masson
Department of Economics, Cornell University, Ithaca, New York, USA

Abstract

Purpose – The purpose of this paper is to demonstrate with real data the enhanced statistical power of a GLS-based event study methodology that requires the same input data as the traditional tests.

Design/methodology/approach – The paper uses full sample, subsample and simulated modified sample analyses to compare the statistical power of the GLS methodology with traditional methods.

Findings – The paper finds that it is often the case that traditional tests will not reject the null when a GLS-based test may (strongly) reject the null. The power of the former is poor.

Practical implications – There are many published event studies where the null is not rejected. This may be because of the phenomenon being tested but it may also be because of the lack of power of traditional estimators. Hence, rerunning them with the authors' more powerful test is likely to reject some currently well-accepted null hypotheses of no event effect, stimulating new research ideas. Moreover, as individual stocks have become more volatile, the additional power of the authors' methodology to detect abnormal performance for recent and future events becomes even more important.

Originality/value – There are more than 500 event studies in the top finance journals, which can broadly be split into two subgroups: contemporaneous shocks like changes in regulation and non-contemporaneous events like mergers. GLS contemporaneous modeling of covariances in the former showed little efficiency gains. The paper's GLS modeling of variances for the latter demonstrates potentially huge effects. Practitioners should be skeptical of prior results accepting the null of no event effect and incorporate GLS to be confident of their future findings.

Keywords Event study, Statistical power, Modelling, Stocks and shares

Paper type Research paper

I. Introduction

Event studies of stock returns have been used extensively in the accounting, economics, and finance literature. In a study of the top five finance journals Kothari and Warner (2007) find more than 500 papers using some form of event methodology. These break down into subgroups, such as how bank values may react to a new regulation announced on a single date. For this question one may wonder about covariances between banks at identical dates (contemporaneous cluster effects). We focus instead on the types of models used for estimating the effects of mergers and acquisitions and other events which are firm specific and occur at different times for different firms. The studies we have in mind often conclude that abnormal returns at

JEL classification – G14

The authors would like to thank Yongmiao Hong, George Jakubson, Ari Gerstle, and Deborah Kerley.

Review of Accounting and Finance
Vol. 11 No. 2, 2012
pp. 166-183

1475-7702
DOI 10.1108/14757701211228200

the time of an event do not differ from their predicted values, which are based on their fitted relationship to an index of market returns. We demonstrate that the tests used in these studies are typically less powerful than they could be and illustrate with data that the tests can be highly inefficient. We present a generalized least squares (GLS) estimator which is more powerful for these event studies.

The main source of the benefits derived from the use of this estimator comes from its "right" (in the least square sense) functional form, which incorporates heteroscedasticity in the proper way. This estimator, therefore, is expected to perform better than the traditional ones even in the absence of cluster effects. Nonetheless, if cluster effects are relevant (which is not the case in our data set), this single regression approach can accommodate this feature.

What we demonstrate is that use of GLS in these circumstances may have profound effects on interpreting the effects of events such as M&As. The traditional models, which typically ignore heteroscedasticity, have a bias towards rejecting the null hypothesis of no event effect. We show that GLS estimation in at least some cases vastly increases the power of the tests and in some cases rejects the null at very high significance levels while the traditional estimation models cannot reject the null at anything close to conventional standards.

It is our contention there are published event studies which incorrectly find insignificance of an event effect due to the simple fact that the traditional event study methodologies lack statistical power. Since one cannot differentiate between lack of significance due to weak methodology and that due to no underlying effect, one should be skeptical of such a finding barring repeating the test using GLS.

II. Historical perspective and the intuition of our GLS event study versus the traditional event study

Our initial conclusion of "discovering" a new GLS estimator that is more efficient for event studies followed extensive readings of event studies and consulting with a standard event study textbook, Campbell *et al.* (1997). The event studies in this literature typically employ a common methodology implemented with a two step (or three step in some cases) estimation procedure, using simple means at each step. In our investigation we found that Thompson (1985, section III-B) presents a GLS model for such studies closely related to our model. Our model goes farther, recognizing that implementing this GLS may lead to sparse matrices requiring additional techniques in implementation. Strikingly, little has been done with the Thompson model in section III-B. Chandra and Balachandran (1992) point out that Thompson does not present data and estimation to show that his insights are important in fact. Meanwhile, GLS modeling of cluster effects (contemporaneous events hitting numerous firms at a point in time) have basically shown that the covariance effects from using GLS in preference to less efficient models are empirically not very important. This may be why others have ignored GLS in non-contemporaneous event tests, where the key factor is not *covariances*, but *variances*/heteroscedasticity. The key point to note is that over the last 26 years following Thompson's section III-B model, the only use of a GLS estimator for this type of non-contiguous events that we are aware of are Da Graça (2002, 2008) and Huang (2010). However, as can be seen by the fact that no form of GLS estimator is represented in standard textbooks, the estimators used in the literature are still inefficient and can easily be improved. We provide the intuition concerning this problem and an analysis between the traditional procedure and our GLS estimator and show that this GLS estimation is far more efficient for at least some data sets. We use both a two step inverse variance weighted average (IVWA) approach and a pooled single-stage heteroscedastic regression approach. We illustrate the greater power of GLS using real data. Further, although the two step IVWA model leads to GLS, we illustrate that a preferred pooled single-stage approach potentially involves sparse matrices which may require special computational methods to be numerically stable.

We are now contributing several new important insights into this body of literature. First, we show the nuts and bolts necessary for the estimation of this estimator, not simply the theory: we are writing for the practitioner. Second, in practice the standard estimation procedures may suffer from numerical instability due to sparse matrices. We show why this is the case and how to rectify this problem. Third, we demonstrate exactly how misleading the traditional method can be in terms of testing power by applying these methods to real data. Fourth, by examining data subsamples we illustrate why our technique is more powerful and we demonstrate the robustness of this result. Fifth, we apply a simulation technique similar to Malatesta (1986) which provides another piece of evidence that our heteroscedastic estimator is superior to other procedures. Sixth, we test the normality assumption of the limiting distributions of the statistical tests. Lastly, we provide some evidence that the sparse matrix instability may have affected our analysis and show that our GLS approach handles it properly.

III. More powerful techniques

First consider traditional event study research. Hypothesis testing is done by treating abnormal returns as a random variable and then testing for significance given the null hypothesis of abnormal returns equaling zero. Implicitly, this hypothesis testing involves several econometric assumptions. First, suppose there exists a common event effect across all events, δ. Then the estimated abnormal return for the ith event, $\hat{\delta}_i$, is an estimate of the common δ and there is a normal distribution of errors around δ across the set of estimates or events. A second possibility is that there are idiosyncratic δ_is, but there is a central tendency which is distributed normally around some $\bar{\delta}$. (We will use the first interpretation except where it is useful to distinguish between the two.) If either of these alternatives is the true structure of the δs, then the traditional event study methodology is virtually never efficient except under highly restrictive conditions.

A. Estimation in event studies

Suppose the "event" occurs at time zero for each firm (where 0 is relative to the firm/event and is not a single unique time). Then the standard event study analysis would first estimate:

$$R_{it} = \alpha_i + \beta_i R_{it}^m + \varepsilon_{it}, \quad t = -T, -(T-1), \ldots, -1, \text{ for each } i = 1, 2, K, N \quad (1)$$

where:

- R_{it} is the market return of firm i at time (day) t.
- R_{it}^m is the stock market index return at time t associated with the ith firm's event.
- ε_{it} is an unsystematic error term with $E(\varepsilon_{it}) = 0$ and $E(\varepsilon'\varepsilon) = \sigma_i{*}\mathbf{I}$, where σ_i is the assumed common variance and $\mathbf{I}$ is a $T \times T$ identity matrix.

There are n independent firm/event specific regressions, each for a time series of length T. (For our illustration, we have $N = 71$ firms/events and $T = 250$ days, approximately one calendar year of stock trading days).

For ease of illustration our "event window," the period of potential abnormal returns is assumed to be one day – multiple days lead to more complex mathematics with no gain in the GLS intuition. The measure of possible abnormal returns is thus:

$$\hat{\delta}_i = R_{i0} - \left(\hat{\alpha}_i + \hat{\beta}_i R_{i0}^m\right) \tag{2}$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the parameter estimates from equation (1). This is the difference between the ith firm's realized return in time zero and its predicted return in time zero based on the model in equation (1) using the stock market return index in time zero. If the $\hat{\delta}_i s$ are independent estimates of a common δ and normally distributed around δ then one can test the hypothesis that δ is positive (without loss of generality) by looking at the p-value for the mean of the estimates as is done in most standard event studies:

$$\bar{\delta} > 0 \tag{3}$$

There is an equivalent estimator, the ordinary least squares (OLS) regression of $\delta_i = \beta X_i + \varepsilon_i$. Consider the standard regression notation of $\beta = (X'X)^{-1}X'y$. When estimating a constant only, X is the $N \times 1$ column vector of 1's, where N is the number of events. Then $X'X = N$ and its inverse is $1/N$. The dependent variable, y in standard notation, is the δ in this context. So y is an $N \times 1$ vector containing the δ_i measurements. Accordingly, $X'y = \sum_{i=1}^{N} \delta_i$. So the estimated coefficient is $\beta = \bar{\delta} = \sum_{i=1}^{N} \delta_i / N$. The computed intercept is identically equal to $\bar{\delta}$. Notice that this estimator does not depend on any variance components and that all δ_is have the same weight in computing the simple average.

Crucially, then, we need to examine the assumptions of OLS. In Greene (2003, p. 42) assumption A4 is the homoscedasticity assumption. "Each disturbance, ε_i, has *the **same*** finite variance, σ^2" (emphasis added). That translates into each individual abnormal return having the same weight when one computes the simple mean. Consequently, as Greene continues, only "*under the very specific assumptions of the classical model*… least squares will be the most efficient use of the data" (emphasis added). In the first stage regressions from which one derives the $\hat{\delta}_i$s, one also derives their respective $\hat{\sigma}_i$s. If variances differ across observations, then the classical OLS model (the simple average here) is not efficient (hence not producing the best linear unbiased estimators). We demonstrate that the conventional modeling which does not weight the $\hat{\delta}_i$s by the inverse of their respective estimated $\hat{\sigma}_i^2$s may significantly reduce the ability of the model to reject the null hypothesis of no event effect.

To illustrate, let us assume the variances of the estimated δ_is are not identically equal, but all the covariances remain equal to zero, i.e. some stocks follow the market index more closely than others. Consider the standard GLS notation of $\beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$, where Ω is the covariance matrix. Again we look only at the case in which the intercept is estimated. X is still an $N \times 1$ column vector of ones, and assume Ω is a diagonal matrix of variances for each of the estimated $\hat{\delta}_i$s, which in turn implies that the off diagonal covariances are equal to zero. For this illustration, consider a two observation regression. Then $X'\Omega^{-1}X = \sigma_1^{-2} + \sigma_2^{-2}$ and $X'\Omega^{-1}y = \sigma_1^{-2}y_1 + \sigma_2^{-2}y_2$, so we arrive at the IVWA of $\hat{\beta} = \sigma_1^{-2^1} y_1 + \sigma_2^{-2}y_2/\sigma_1^{-2} + \sigma_2^{-2}$[1]. That is, from the standard first stage regressions, one can easily compute the GLS estimator under the assumption that all the covariances are equal to zero. Thus, the GLS estimator is efficient, **not** the OLS estimator, as is used in standard practice, unless all the variances of the estimated δ_is are identical (an outcome which will occur with probability measure of zero). Moreover, it is important to notice that this GLS estimator relies on a very specific functional form of σ_is and y_is[2].

B. Alternative event study model specification for a common event effect
Consider the following structure. The same event model can be written as:

$$R_{it} = \sum_{j=1}^{N} D_{i|j}\left(\alpha_j + \beta_j R_{jt}^m\right) + D_{i|j}^{t=0}\delta + \varepsilon_{it} \tag{4}$$

where:

- $D_{i|j} = 1$ if $i = j$, and $= 0$ for $i \neq j$.
- $D_{i|j}^{t=0} = 1$ for event $i = j$ when $t = 0$ (the event window), and zero otherwise.
- δ is the potential abnormal return which is assumed to be common across events.
- α_j and β_j are the parameters which are estimated over the estimation window and have exactly the same interpretation as in the traditional estimation.
- ε_{it} is an unsystematic error term with $E(\varepsilon_{it}) = 0$ and $E(\varepsilon'\varepsilon) = \Omega = \left(\boldsymbol{\sigma}_i^2 \otimes \mathbf{i}\right)^* \mathbf{I}$ where $\boldsymbol{\sigma}_i^2$ is an $N \times 1$ vector of firm/event variances, $\mathbf{i}$ is a $(T+1) \times 1$ vector of 1's, and $\mathbf{I}$ is a $((T+1)^* N) \times ((T+1)^* N)$ identity matrix[3].

If the model in equation (1) is efficient *for each event taken separately*, equation (2) gives the event specific estimate of the abnormal return, $\hat{\delta}_i$, and the true abnormal return, δ, is common across all events, then our pooled heteroscedastic model in equation (4) is efficient, indeed δ is BLUE under these assumptions.

Equation (4) can, in principle, be estimated using simple maximum likelihood techniques found in any basic statistical software, but there is a technical problem which can arise. Consider the same model in matrix form, illustrated first here with only two events ($N = 2$):

$$\begin{bmatrix} R_{1,-T} \\ \vdots \\ R_{1,-1} \\ R_{1,0} \\ R_{2,-T} \\ \vdots \\ R_{2,-1} \\ R_{2,0} \end{bmatrix} = \begin{bmatrix} 1 & R_{1,-T}^m & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & R_{1,-1}^m & 0 & 0 & 0 \\ 1 & R_{1,0}^m & 0 & 0 & 1 \\ 0 & 0 & 1 & R_{2,-T}^m & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & R_{2,-1}^m & 0 \\ 0 & 0 & 1 & R_{2,0}^m & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,-T} \\ \vdots \\ \varepsilon_{1,-1} \\ \varepsilon_{1,0} \\ \varepsilon_{2,-T} \\ \vdots \\ \varepsilon_{2,-1} \\ \varepsilon_{2,0} \end{bmatrix} \tag{5}$$

This nests equations (1) and (2) along with the assumption of a common δ. In place of the $\bar{\delta}$ in equation (3), the equations (5) directly estimate $\hat{\delta} \equiv \bar{\delta}'$, where $\bar{\delta}'$ is the IVWA of the $\hat{\delta}_i$s. This approach is identical to that of equations (1)-(3) with the same maintained hypotheses, but it can be estimated with our GLS to attain the best linear unbiased estimator. (In the two stage procedure, finding the IVWA of δ will be virtually identical to equation (5)[4]).

The technical issue arises in the construction of the data matrix. To illustrate, now suppose $T = 10$ and $N = 2$. Then there would be eleven non-zero data elements and 11 elements identically equal to zero in each of the first four columns of the data matrix and the fifth column contains 20 elements identically equal to zero and only two non-zero elements. Now consider the general case with N firms/events and T time periods before the event window. The data matrix contains $2N + 1$ columns. Each of the columns 1 through $2N$ have $T + 1$ non-zero elements and $(N - 1)(T + 1)$ elements identically equal to zero. For this part of the matrix there are $2N(T + 1)$ non-zero elements and $2N(N - 1)(T + 1)$ elements identically equal to zero. Therefore, the ratio of non-zero elements to the total number of elements in the first $2N$ columns is $2N(T + 1)/2N^2(T + 1)$ or $1/N$. For $N = 71$ (and $T = 250$), this implies that only 1.4 percent of the elements are non-zero. The final column in the data matrix has N positive non-zero elements and NT elements identically equal to zero, thus it is likewise dominated by zeros. Unlike many consistent estimators whose "accuracy" increases as N rises, when the problem of sparse matrices occurs, the accuracy falls as N increases.

These sparse matrices present unique numerical problems. Computationally, conventional statistical programs treat zeros as floating point approximations of zero when inverting the $X'X$ matrix. We are inverting $X'\Omega^{-1}X$, where Ω is a diagonal matrix so the number of zeros in the estimation remains the same. This can create rounding errors which increase as the number of zeros increases. The computational problems involved with such cases are covered in Thisted (1988), and we demonstrate that this may lead to parameter estimate instability later.

We handle the sparse matrix problem by extending an approach pioneered by Mundlak (1961). He noted that the design matrix for a linear regression including dummy variables had a special structure analytically allowing partitioned inversion of the $X'X$ matrix. Chamberlain (1980) noted a similar technique could be used in a maximum likelihood setting. When using Newton-Raphson (or something similar) to maximize the log likelihood function, each iteration of the procedure has a structure similar to the linear case. The common parameters can be updated by analytically simplifying the Newton-Raphson procedure by inverting a matrix that is only of the size ($k \times k$), where k is the number of common parameters ($k = 1$ in the model above, and $k = 7$ in an extended application below). The estimates of the individual effects can then be updated as a function of the update to the common parameters one at a time. Iterating this process to convergence maximizes the log likelihood.

We develop a similar matrix inversion technique for sparse matrices[5]. We return to the consequence for estimation below.

C. Alternative specification for event specific abnormal returns

Event studies often drop the assumption of a common δ, where measurement errors lead to differences in the δ_i's, after testing for the sign of δ. They hypothesize that the "measurements" in the above model are, in part, explainable by exogenous observable event specific factors. They then perform a regression of abnormal returns on firm/event specific explanatory variables such as equation (6):

$$\hat{\delta}_i = \Phi X_i + u_i. \tag{6}$$

The X_i matrix includes firm or event specific variables which might explain the differences in the $\hat{\delta}_i s$. This creates a third step in the standard analysis.

Following the intuition from the last section in which we examined not only our benchmark GLS estimator but compared this with the two stage IVWA estimator, we could think of the analogue of the IVWA as applied to the regression (6). This would be an inverse variance weighted regression in place of equation (6). As in the last section, given the assumptions of the system (the equations (1) yield the best linear unbiased estimators for each event, equation (2) represents the abnormal return, and equation (6) represents the common parameters explaining abnormal returns) our benchmark GLS model will be efficient, whereas the traditional estimation of equation (6) will not be efficient unless all the variances are ***identical***.

For simplicity, we again present an $N = 2$ version of equation (6), where $\delta_i = \phi^0 + \phi^1 x_i + u_i$ with only one explanatory variable. (We use superscripts on these parameters, which are assumed to be common across all events, to avoid confusion with the event specific parameters, subscripted α's and β's.):

$$\begin{bmatrix} R_{1,-T} \\ \vdots \\ R_{1,-1} \\ R_{1,0} \\ R_{2,-T} \\ \vdots \\ R_{2,-1} \\ R_{2,0} \end{bmatrix} = \begin{bmatrix} 1 & R^m_{1,-T} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & R^m_{1,-1} & 0 & 0 & 0 & 0 \\ 1 & R^m_{1,-0} & 0 & 0 & 1 & x_1 \\ 0 & 0 & 1 & R^m_{2,-T} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & R^m_{2,-1} & 0 & 0 \\ 0 & 0 & 1 & R^m_{2,0} & 1 & x_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \phi^0 \\ \phi^1 \end{bmatrix} + \text{ error term} \tag{7}$$

Thus, equation (7) captures the maintained hypothesis of the second set of tests, but is efficient, yields the best linear unbiased estimators, rather than inefficient estimators. Again, for implementation, one may need to use the sparse matrix methods we introduced above. (We demonstrate, at least for our data, the sparse matrix methods are necessary for the estimation in this third step of modeling).

IV. Empirically demonstrating the power of the proposed tests

A. A brief literature review

Da Graça (2002) provides a review of the literature on evaluating acquisitions via event studies. Kothari and Warner (2007) provide a comprehensive survey of the application of the event study methodology in the finance literature.

As for the previous event study techniques, Thompson (1985) derives various[6] estimators for event studies in pooled forms, one of which is like ours (the one in his section III-B). However, he defines time t to be contemporaneous across all events/securities. This would be like examining the effects of a single event (e.g. a declaration of war) on several stocks. For mergers or privatization auctions, the events do not occur on the same day, so we are examining non-contemporaneous events. Thompson also assumes the error terms are serially independent yielding a covariance structure in which there may be contemporaneous correlations or, as some authors say, cluster effects. Despite his relevant contribution, Chandra and Balachandran (1992) note that Thompson does not identify the benefits of using this estimator. We do so herein and empirically demonstrate such benefits.

In the event-study literature, the expression "cluster effects" is associated with two different heteroscedastic features. Some authors refer to "cluster effects" meaning the contemporaneous correlations between the error terms between two firms. This is the effect that, for example, may be relevant when one conducts an event-study designed to measure the impact of a regulatory change, say, in the telecom industry by looking at the change in value of a few telecom firms relative to the movement of the market as a whole.

Knif *et al.* (2008) point out that ignoring the covariances (the contemporaneous cluster effect) may cause over-rejection of the null hypothesis. To the extent this is a potentially relevant concern, there have been attempts to evaluate the degree to which this effect matters in practice.

For instance, Malatesta (1986) simulates a number of competing estimators, including one of Thompson's, but does not simulate the estimator of Thompson's which is similar to ours. Malatesta (1986) assumes time period t is contemporaneous across firms and the covariance matrix structure may reflect contemporaneous cluster effects. His simulations suggest that econometrically modeling contemporaneous correlations may not significantly affect estimation in a broad (as opposed to industry-focused) event study[7].

In our data set we deal mostly with non-simultaneous events. Therefore, the corresponding estimation windows do not overlap perfectly if at all. Moreover, our events encompass a range of industries and the firms whose values may change in response to the events have different nationalities. In any event, in our preliminary analysis, we allowed for contemporaneous cluster effects among the observations that corresponded to a few actions that occurred on the same days, following the procedure suggested in Campbell *et al.* (1997, p. 167). In accordance with Malatesta's simulations, the introduction of the cluster effects did not affect our results.

Some other authors, for example Pynnonen (2005), have focused on time series types of heteroscedasticity that can be handled with GARCH models, instead. These models are appropriate to deal with the so-called volatility cluster effect.

Interestingly, both types of cluster effects can only be evaluated when the researcher approaches the problem as a single pooled regression model. Typically the authors that model each type of cluster effect find that erroneous inferences can be made if one ignores that type of effect as compared with the traditional procedure. Part of the reason why their results are "better" may come from the use of the appropriate firm-specific variance weighting scheme that is implicit in the pooled regression approach. The remaining part may, in fact, come from the postulated cluster effect. However, the latter part may be over-credited if one overlooks the relevance of the former. In our paper, we demonstrate that the former, indeed, matters.

Serra (2002) surveys various forms of parametric and non-parametric approaches to event-studies, but in her footnote 2 she explicitly refers the reader to Thompson (1985) for further details. So she does not present an estimator that is similar to ours. She also discusses that variance (volatility) may increase over the event window. This is a feature that our pooled single regression approach could address directly as well. In fact, Da Graça (2008) models this feature and concludes that, for the same data set as ours, "[. . .] [W]hether variance is allowed to increase [. . .] or not, is immaterial for the statistical significance of the average abnormal returns at [the] 0.1 % [significance] level." [Words in brackets added]

B. The data

In 1990, the Brazilian Government decided to privatize many of its companies. Most of the privatization auctions occurred from 1995 to 2000. The auctions were first price sealed bid auctions each held on a single day. We analyze 71 competitive auctions.

The estimation window is based on the end of the trading day prices for the 250 stock trading days up to one week before the relevant privatization auction. The event window is the day of the auction (Da Graça 2008 extends this window).

We first analyze the average event effect and then model the event effect as a function of event characteristics: Nationality of the acquiring firm; relative size of the bid to the firm; whether the bidder is in the same DataStream Advance industry. The 71 auctions had two or more approved bidders, if only one showed up, that firm could bid the reservation price, so we control for this as well.

The data source for the location of the acquirers' headquarters was Bloomberg.com. DataStream Advance provided industry classifications, market value figures, and exchange rates used to convert all values to the Brazilian Real. The Rio de Janeiro Stock Exchange web site furnished the minimum and final prices for most of the auctions. Dow Jones Interactive and the BNDES Annual Reports complemented this series and also provided, along with Manzetti (1999), data on the shares acquired by each firm.

C. First hypothesis: uniform event effects

In order to carry out this part of the analysis we proceed in two steps. First we examine the full sample to see how well the various models perform. Then we perform subsample forensics to show the sensitivity of the models to the variances of the first stage estimators using the traditional model and to establish the fact that the apparent power of our GLS methodology is not just a sample artifact.

The traditional methodology is based on the simple average of the 71 estimated abnormal returns, which means that each estimated abnormal return is assigned the same weight in estimating the true effect of the event, regardless of whether or not they were measured with different precision levels (variances). The estimator of the variance of this simple average is assumed to be the simple average of the respective variances of the individual abnormal returns. Calculations can then be undertaken to find the *t*-value under the null hypothesis of abnormal returns equaling zero. We contrast this with the IVWA and with our single equation pooled model in equation (5) both using the sparse matrix methodology and without using the sparse matrix methodology. We designate the "ML" (maximum likelihood) model as the estimation performed using the PROC MIXED procedure in SAS which inverts the sparse matrices without any special sparse matrix procedures[8]. The results are summarized in Table I.

The "Traditional" model simply lacks power relative to the other estimates. Although the results for the traditional model are positive and significant (on a one-tailed test), they are slightly different in absolute value and are far less significant or precise than the other models, all three of which have virtually identical results. Utilizing the heteroscedastic methods generates p-values which are remarkably lower, having fallen by over 99 percent from the p-value of the traditional model. Note in particular the IVWA model. This approach simply uses a traditional two step methodology but uses the variance information which is generated in the traditional model, but not utilized.

We next examine sample subsets to demonstrate how weak the traditional methodology can be in practice. We ordered the sample observations from the estimates with the lowest estimated variance of δ to the highest estimated variance of δ. We then ran subsample models using the traditional model and the IVWA method. The subsamples were constructed by first eliminating the observation with the lowest estimated value for the variance of δ. Then we performed both tests again. This was followed by then eliminating from this new subsample the observation with the next lowest estimated variance of δ and repeating the same procedure[9]. The resulting p-values are shown in Figure 1.

Procedure	IVWA	ML Proc. mixed	Our GLS benchmark	Traditional
Abnormal return (%)	0.69	0.70	0.70	0.62
t-value for H_0: $\delta = 0$	3.54	3.54	3.58	1.77
p-value (one-tailed)	0.0002	0.0002	0.0002	0.038

Table I. Abnormal return estimates and respective statistic derived from various procedures

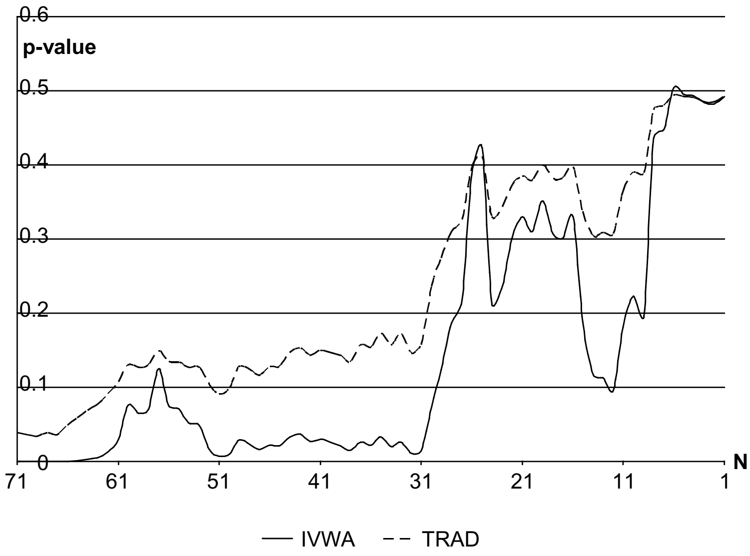

Figure 1. p-value derived from the IVWA and traditional methodologies in the subsamples

The traditional model only meets the 95 percent significance level for five of the subsamples. Examining the 37 subsamples for which the IVWA model satisfies the 95 percent significance level, nine of these subsamples do not even satisfy an 85 percent significance criterion for the traditional model (24 do not even satisfy the 90 percent level). There is one truncation for which the IVWA estimator is significant at the 98 percent level, whereas the traditional estimator does not even rise to the 85 percent significance level.

Another piece of evidence suggesting the superiority of a heteroscedastic estimator over the traditional estimator is provided by conducting a "reverse Malatesta's (1986) simulation. In his paper there was no abnormal performance to begin with. He artificially introduced a 1 percent abnormal performance to his random sample and evaluated the statistical performance of various estimators and corresponding tests in response to this disturbance of the original situation. Here, on the contrary, we detect the presence of some positive abnormal performance. Our "reverse" simulation means that we disturb our original result by artificially subtracting some "quantities" from it. On the event dates we subtract a fraction of the firms' standard deviations from their respective stock returns. The fraction varies from 0 to 1. Then for each fraction, we reconduct the event study, *i.e.* re-estimate the event effect and re-perform the statistical tests. In Figure 2, for each fraction of the standard deviations,[10] we plot the (one-tailed) *p*-values of three alternative methods: our GLS, the IVWA, and the traditional. As the fraction increases, the more is subtracted from the original event effect. Consequently, the *p*-values decrease as the simulated event effect approaches zero.

Figure 2 shows that our GLS method unequivocally improves upon the traditional procedure but it is practically undistinguishable from the IVWA. For a disturbance for which our GLS method still rejects the null hypothesis at the 99 percent

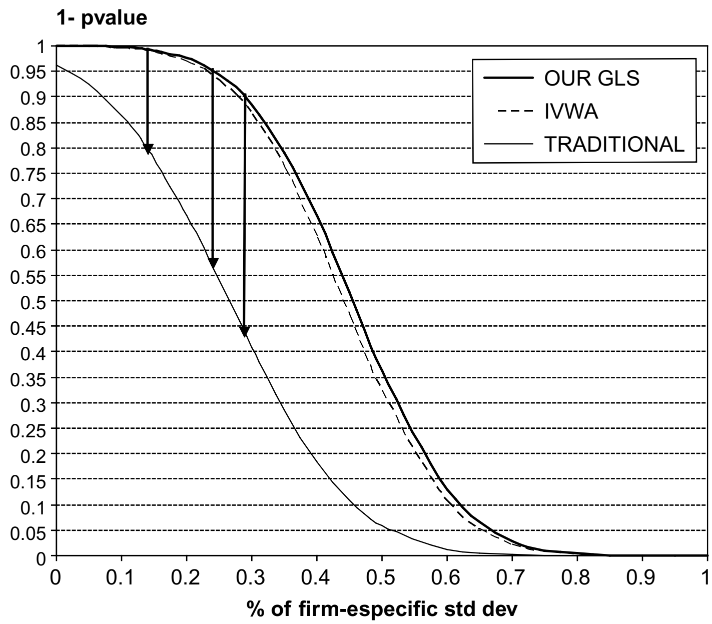

Figure 2.
p-values derived from our GLS, IVWA and traditional methodologies in "reverse Malatesta" simulations

significance level, the traditional method's significance level has fallen below 80 percent. When our GLS model is still rejecting the null at the 95 percent level, the traditional method's significance has fallen below the 60 percent level. And when the disturbance leads our GLS model to 90 percent, the traditional model has fallen to only the 45 percent level. This is all shown in Figure 2, where the p-values correspond to one-tailed tests.

We now look at data forensics to see how our results are generated in this data set. It is possible that our results are being driven by peculiarities of our sample. For example, outliers with large variances and low abnormal returns may lead to our results.

First we present the scattergrams of the pairs abnormal returns and the corresponding variance in Figure 3.

From Figure 3, it is apparent that there are five observations with very low precision.

To test whether or not our results are numerically driven by the data, we regress the estimated abnormal returns on their respective estimated variances. We run this regression for all observations. Then we drop the five highest variance observations as they could be considered outliers and run the same regression model again. The results are (Table II).

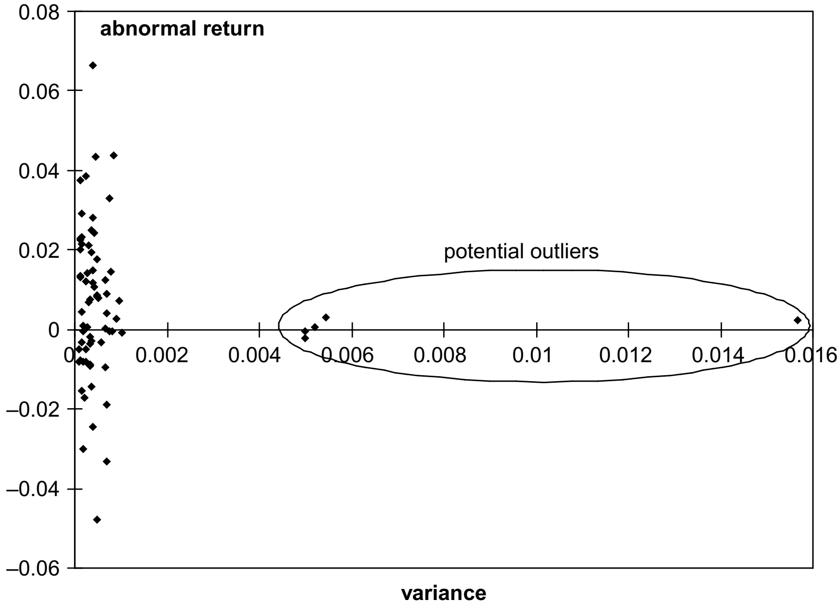

Figure 3.
Abnormal return (vertical axis) vs variance (horizontal axis)

	All observations		All but the five least precise obs.	
	Estimate	t-stat.	Estimate	t-stat.
Intercept	0.007	2.76	0.007	1.60
Variance	-0.564	-0.53	-1.94	-0.20

Table II.
Forensics regressions: abnormal return = a + b variance + error

In both regressions above, there is no statistically significant negative effect within the sample range implying there is no evidence supporting a negative relationship in the data which could be driving the above results. Model characteristics, not sample characteristics, are driving the results.

Having tested for a common δ, which is often done, we move next to trying to explain δ_i with characteristics which are event specific. In this process we find important additional potential sources for bias and the importance of using sparse matrix techniques for estimation.

D. Second hypothesis: abnormal returns explained by exogenous data

The standard event methodology often appears to be schizophrenic in the sense that many studies first estimate a $\bar{\delta}$ and then estimate the determinants of the individual δ_is. The first step, estimation of $\bar{\delta}$ as above, is consistent with either the assumption of a single δ with estimation error of $\delta_i = \delta + \varepsilon_i$ where the ε_is are normally distributed or the assumption that the δ_is vary across firms but are normally distributed. The former interpretation is not consistent with estimating the determinants of the δ_is, and the latter is.

In estimating the determinants of the δ_is, we present two models analogous to those we ran for the common δ: a two stage model using an inverse variance weighted regression, the ML model (using the SAS program's standard matrix inversion methods without using a sparse matrix algorithm) and our GLS model using sparse matrix methods, for comparison.

The inclusion of variables here follows the literature. Without detail (details are in Da Graça (2002)), the included variables are: Nationality of the acquiring firm; a dummy for being in the same Industry; a measure of "Participation," a dummy which takes a value of one if it appears that there was only one bidder (winning bid equal the reservation price); relative size, the size of the acquired firm relative to the acquiring firm (adjusted for the acquiring firm's share of an acquiring consortium); Relative size interacted with each of Nationality, Industry, and Participation.

We report results in a table with four different estimation procedures. First we use a two stage procedure in which the second stage is an inverse variance weighted regression. Second we use the single stage procedure, imbedding the explanatory variables into a regression like our GLS formulation, but use SAS ML (maximum likelihood) without sparse matrix techniques. Then we have our benchmark GLS, the one stage estimation procedure using sparse matrix estimation techniques. Finally we present the traditional model, an OLS regressing δ_i on the exogenous variables. The results are in Table III.

From Table III, we observe that the four models have different results. Also of interest, although the earlier results for $\bar{\delta}$ showed a much lower t-value for the traditional model, there is no such pattern here. The benchmark GLS model is the correct model assuming that variances in estimation of δ_is differ across events. The traditional model differs from the benchmark model, especially on the "same industry" indicator. The ML model and the GLS model should be identical if there are no sparse matrix estimation problems, but clearly there are sparse matrix problems because they differ so much (e.g. nationality in the ML is about half of that in the GLS benchmark). And unlike the estimation of $\bar{\delta}$, IVWA two stage regression parameter estimates are also deviating quite a bit from the benchmark.

	Inverse variance weighted regression	ML	Our GLS (benchmark)	Traditional (OLS)
Intercept, φ^0	−0.0203 (−1.68)	0.0026 (0.83)	−0.0077 (−1.21)	−0.0203 (−2.12)
Nationality	0.0324 (2.53)	0.0105 (2.33)	0.0209 (2.31)	0.0266 (2.89)
Same Industry Indicator	0.0253 (2.03)	0.0071 (1.74)	0.0071 (1.74)	0.0165 (1.82)
Participation: 1 if the Bid is Above the Reservation Price	−0.0042 (−0.54)	0.0025 (0.83)	0.0024 (0.81)	0.0031 (0.52)
Relative Size: Acquisition Size Over Purchaser Size	0.0050 (3.67)	0.0002 (0.55)	0.0014 (2.45)	0.0033 (2.89)
Relative Size * Nationality	−0.0034 (−2.71)	−0.0010 (−1.79)	−0.0021 (−1.78)	−0.0028 (−2.70)
Relative Size * Industry Indicator	−0.0035 (−2.85)	−0.0015 (−2.17)	−0.0014 (−2.16)	−0.0030 (−2.92)
Relative Size * Participation Indicator	−0.0015 (−2.16)	−0.0004 (−0.80)	−0.0004 (−0.80)	−0.0004 (−0.62)

Table III.
Parameter estimates and t-stats (in parentheses) for the IVWA, Our GLS and OLS procedures for our complete sample ($n = 65$)

Some of the instability may be coming from high multicolinearity, nationality and industry are correlated 0.68 and the interacted relative size with nationality is correlated 0.75 with nationality.

Still what this indicates is that for some data sets, this one in particular, the only way to get efficient estimates is to use the benchmark one step GLS estimator with sparse matrix estimation techniques. A priori we must assume that some other event studies are drawing the wrong conclusions from their data by not using what we call our GLS estimator, the uncertainty is only about the frequency of such incorrect inferences.

Without commenting in detail about the model, Huang (2010) offers a very recent and compelling piece of evidence that favors our modeling approach. He explains abnormal returns in M&As, returns of targets, of acquiring firms, and of their combinations (synergies). His explanatory variables are an index of corporate governance and control variables (e.g. relative size). Using traditional two stage modeling, his measure of worse corporate governance has a positive, but insignificant, influence on the value of the acquiring firms and on synergies. Using our modeling taking into account heteroscedasticity all results become statistically strong and the poor corporate governance variable switches sign and becomes statistically significant indicating that poor corporate governance leads to the expected lower value of the acquisition to the acquiring firm and lower synergies from mergers. The traditional model failed relative to the application of our model.

Before concluding, it is worth noting the contrast between the common δ and the event specific δ cases. As we conjectured for the common δ model the use of the traditional model tends towards accepting the null hypothesis of no event effect existing. If the inverse variance weighted model strongly rejects the null hypothesis, we can be fairly certain that the best linear unbiased estimation using our GLS model will do so as well. But, for the firm/event specific models we can make no such claim. Any ϕ_i could be more or less significant in a comparison between the Traditional model and any of the other models. All we know with certainty is that the mean of the

Traditional model's $\hat{\delta}_i$s is less likely to be significantly different from zero than that of the inverse variance weighted or our GLS models.

V. Conclusions

This paper addresses a gulf between the econometric theory of event studies and how event studies are typically conducted in practice. Among the 500 plus event studies published in the finest finance journals, it is highly likely that some of them have incorrectly found insignificance of an event effect due to the simple fact that the traditional event study methodologies lack statistical power. Looking forward, our contribution goes towards reducing the probability that this type of mistake happens from now on. It also raises concerns about the credibility of many published event studies where the no effect hypothesis is not rejected. Obviously this may be because of the phenomenon being tested is indeed innocuous but it may also be because of the lack of power of the traditional test. Without rerunning the data from these tests using GLS there is no way to know if these published results are misleading or not.

In the finance literature, event studies can broadly be split into two subgroups depending on whether the subject firms are all affected at the same time, like banks reacting to a new regulation, or the different firms are affected at different times, like in mergers and acquisitions studies. For the former subgroup, the so-called cluster effects (related to the contemporaneous covariances among the firms' stock returns) have been modeled with GLS but some studies (in particular, Malatesta (1986)) have shown that embodying these effects leads to little practical impact. This may be the reason why no-one has looked a GLS modeling of variances (precision of estimated firm's stock return counterfactual model) for the latter subgroup. This is precisely the gap that our work intends to bridge.

We first examine the traditional methodology used by practitioners and demonstrate that its event effect estimator equally weights the firms' abnormal returns, which is equivalent to the OLS estimator in a regression context. It is as if all the firms' abnormal returns are estimated with the same precision. But this rarely occurs in practice. Typically each firm's abnormal return is estimated with its own (specific) precision level. Intuitively the more precise estimates should be given more weight in computing the overall event effect. The exact functional form that blends together all abnormal returns and respective variances so that statistical efficiency is maximized is given by a particular GLS estimator which is similar to a model proposed by Thompson (1985) in his section III-B. Despite Thompson's proposed model, we are unaware of any applications of the model in the ensuing 26 years other than Da Graça (2002 and 2008), Huang (2010) or this paper.

In this paper we contrast the traditional two step approach to estimation, looking typically at averages of estimated abnormal returns, with a very simple IVWA estimator. We show that there is relevant information from the first stage estimation which is not being used efficiently in the second stage when one employs the traditional methodology, specifically the measurement error variance. The efficient use of the information about variances leads to a more powerful test and potentially to stronger results.

We "evaluate" the models using Brazilian privatization auction events and demonstrate that for our data there are substantial differences between the traditional estimation procedures and our GLS procedures. These auctions occurred over the

1990s and they were, for the most part, non-simultaneous. This is the feature that approximates our application to the M&A subgroup of event studies.

By using subsample analysis, we demonstrate the intuition for the substantial gap between the inverse variance weighting procedures and the traditional method. In one subsample the IVWA had a 98 percent significance level, while the traditional method failed to be significant at the 85 percent level. We also provide full sample tests, "reversing" Malatesta's (1986) simulation technique in which he artificially added an event to data where there was no event. We instead artificially reduce the measured event effects in our data and find, for example if we reduce the measured effects to the point at which our GLS estimator drops to a 95 percent significance level, that the traditional model's power has fallen to under a 60 percent significance level. Additionally we check the possibility that our strong results are driven by some "special" feature of our data set. For instance, it could be the case that, coincidentally, the highest abnormal returns had the lowest variances. We test for this pattern and do not find any relationship between measurement and precision.

The event tests we evaluate all converge theoretically to a normal distribution as the sample size increases. Because individual stock returns may not fit the normality assumption well, it is appropriate to check as to whether these tests adhere to the normality assumptions. We find that this assumption cannot be rejected for our heteroscedastic test but it is violated for the traditional test.

We also demonstrate that sparse matrix problems may occur when using our pooled single regression model. We show that in our data, the model to explain the differences between events with exogenous explanatory variables is plagued by instability problems due to sparse matrices. The GLS estimator we propose is built so as it does not suffer from this potential numerical/computational problem.

By looking at our subset analysis and our simulation analysis we demonstrate that studies using the typical traditional analysis which do not find a set of events to be significant, may in fact be simply finding that their methodology was inadequate to find the underlying significance. Insignificant results should, at least, be questioned until models have been rerun using heteroscedastic methodologies (either IVWA or our GLS).

Notes

1. Consider, for example, $y_1 = 1.00$, $y_2 = -2.00$, $\sigma_1 = 0.10$ and $\sigma_2 = 4.00$. The simple mean is -0.50 whereas the GLS estimate (or the inverse variance weighted average – IVWA) is 0.99812617 using weights of 0.99937539 and 0.00062461.
2. A different and, consequently, less efficient functional form, is presented in Patell (1976). He proposes a methodology that relies on normalizing the measurements (y_is) by their standard deviations (not their variances) and then taking the simple average of the normalized abnormal returns. Using the same two-observation illustration as for the GLS estimator in the text, Patell's measure is expressed as $\left(\sigma_1^{-1}y_1 + \sigma_2^{-1}y_2\right)/2$.
3. As Fama (1970) notes, under the efficient markets hypothesis, serial correlation should be equal to zero.
4. Suppose that equation (1) generates the best linear unbiased estimator *for each specific event separately*. This equation is estimated over time $[-T, \ldots, -1]$. Were there no event in time 0 and [1] could be estimated for time periods $[-T, \ldots, 0]$, then the model would yield the best linear unbiased estimator for exactly the same reasons as for the original time frame.

Assume instead there is a shift at time zero captured by a dummy variable. Thus, there is only one observation of this shift per event. By assuming a common shift/dummy δ across all firm/events, the same properties which make [1] yield the best linear unbiased estimators for each individual event of the period $[-T, \ldots -1]$ make this pooled regression yield the best linear unbiased estimators if estimated by GLS over the time period $[-T, \ldots, 0]$.

5. The algorithm and its derivation are available from the authors upon request.
6. Thompson (1985) presents two pooled single regression estimators, one for the case in which the average event effect is estimated directly (in his section III-B) and one for the case in which the individual event effects are estimated jointly (in his section III-A). Malatesta (1986) simulates the latter model and tests the simple mean of these individual event effects. He showed that the latter modeling approach does not improve upon the traditional two stage estimation approach *when the covariance elements, or cluster effects, are small*. On the other hand, in our work we will discuss the former model and show, that *at least for our data*, our GLS estimator *focusing on variances* is far more efficient than the traditional two stage estimation approach.
7. Thompson (1985) suggests that the form of his simulation may be partially responsible for this negative finding.
8. We modeled (4) assuming a cluster effect $\sigma_{ij} = \rho\sigma_i\ \sigma_j$ and also allowed for first order autocorrelation $\varepsilon_{it} = \rho_i\varepsilon_{it-1} + v_{it}$. Neither of had significant effects (available from the authors upon request). We also modeled the Campbell et al. (1997) suggested standard error normalization which yielded $t\ -3.16$.
9. Subsequent subsamples contain abnormal return estimates that are measured with the highest estimated variance (lowest precision). Had we proceeded in the opposite direction, i.e. eliminating the next highest estimated variance, later subsamples would contain the most precise estimates. Very precise estimates tend to yield very low p-values, converging to zero. The difference between methodologies could become practically indiscernible. Therefore, by truncating the way we did, the subsampling reveals that the IVWA method outperforms the traditional method exactly when there is more imprecision in the estimation stage and the researcher needs to extract all the available statistical power. To this point, Kothari and Warner (2007) note that individual stocks have become more volatile over time. Consequently, the authors continue, the power to detect abnormal performance for more recent events is lower than for earlier periods.
10. Throughout the paper we have been writing in terms of variance rather than standard deviation as measure of variability because variance is the essence of GLS estimation. For the simulations discussed in this paragraph and presented in Figure 2, we use standard deviation. This is because for simulating a "reverse Malatesta" test we (as did he) want our adjusted valuations to be in the same units as the data, standard deviation is in data units, variance is in their square.

References

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

Chamberlain, G. (1980), "Analysis of covariance with qualitative data", *Review of Economic Studies*, Vol. 47, pp. 225-38.

Chandra, R. and Balachandran, B. (1992), "More powerful portfolio approaches to regressing abnormal returns of firm-specific variables for cross-sectional studies", *The Journal of Finance*, Vol. 47 No. 5, pp. 2055-70.

Da Graça, T. (2002), "Stock market performance of winners in privatization auctions: a new event study methodology", Doctoral dissertation, Cornell University, Ithaca, NY.

Da Graça, T. (2008), "Distribution of gains in the Brazilian privatization program: evidence from an event study", working paper, available at: http://ssrn.com/abstract=1107331

Fama, E. (1970), "Efficient capital markets: a review of theory and empirical work", *Journal of Finance*, Vol. 25, pp. 383-417.

Greene, W. (2003), *Econometric Analysis*, 5th edition, Prentice-Hall, Upper Saddle River, NJ.

Huang, H. (2010), "Essays on firm behavior and efficiencies", Doctoral dissertation, Cornell University, Ithaca, NY.

Knif, J., Kolari, J. and Pynnonen, S. (2008), "Stock market reaction to good and bad inflation news", *The Journal of Financial Research*, Vol. 31, pp. 141-66.

Kothari, S. and Warner, J. (2007), "Econometrics of event studies", in Eckbo, E. (Ed.), *Handbook of Corporate Finance – Empirical Corporate Finance*, Vol. 1, North-Holland, Amsterdam, pp. 3-36.

Malatesta, P. (1986), "Measuring abnormal performance: the event parameter approach using joint generalized least squares", *The Journal of Financial and Quantitative Analysis*, Vol. 21, pp. 27-38.

Manzetti, L. (1999), *Privatization South American Style*, Oxford University Press, New York, NY.

Mundlak, Y. (1961), "Empirical production function free of management bias", *Journal of Farm Economics*, Vol. 43, pp. 44-56.

Patell, J. (1976), "Corporate forecasts of earnings per share and stock price behavior: empirical tests", *Journal of Accounting Research*, Vol. 14, pp. 246-76.

Pynnonen, S. (2005), "On regression based event study", working paper, Vaasa University, Vaasa.

Serra, A. (2002), "Event study tests: a brief survey", working paper, Universidade do Porto, Porto.

Thisted, R. (1988), *Elements of Statistical Computing*, Chapman & Hall, New York, NY.

Thompson, R. (1985), "Conditioning the return-generating process on firm-specific events: a discussion of event study methods", *Journal of Financial and Quantitative Analysis*, Vol. 20, pp. 151-68.

About the authors

Tarcisio da Graca is Senior Economist at the Canadian Competition Bureau (the opinions and ideas expressed in this paper do not reflect necessarily those of the Competition Bureau).

Robert Masson is Professor at the Department of Economics, Cornell University, Ithaca, New York, USA. Robert Masson is the corresponding author and can be contacted at: rtm2@cornell.edu