More Power To You: Demonstrating Increased Power of GLS for Event Studies

by

Tarcisio B. da Graça

and

Robert T. Masson

* University of Brasilia and Cornell University. This expands upon work in da Graça (2002). Thanks are due to Yongmiao Hong, George Jakubson, Ari Gerstle, and Deborah Kerley. Corresponding author Robert T. Masson, Department of Economics, Cornell University, Ithaca, NY 14853, rtm2@cornell.edu.
Abstract

It has long been known that ordinary least squares event study methodologies used by practitioners are inefficient in terms of classical econometric theory (e.g., Thompson 1985 and Malatesta 1986), but practice has failed to rectify this problem (cf. Campbell, Lo, & MacKinlay 1997). When performing event studies one should use a Generalized Least Squares [GLS] approach to estimation rather than the typical traditional approach (which imposes the assumption of homoscedasticity in the second stage estimation).

We demonstrate the power of the GLS model relative to the traditional model using data on Brazilian privatization auctions. We demonstrate that for these data GLS is far more powerful by contrasting p-values for the full data set. We then perform subsample analyses with which we show the intuition for this and find that the traditional estimation techniques used in the literature may not even be able to reject a null hypothesis of no event effect at the 85% level when the efficient GLS estimator finds significance at the 98% level! We also “reverse” a method used by Malatesta to undertake full sample analyses; we simulate using the entire data set after artificially reducing the measured event effects (subtracting from the true stock return in the event window). As we reduce the measured effects until our GLS model falls to the 95 percent significance level, we show that the traditional model has fallen below even a 60 percent significance level.

Importantly, for our one step GLS estimation, we show that sparse matrices (dominated by zeros) can create estimation problems, and in our appendix we show the SAS code to perform GLS which overcomes this problem.

The primary lesson to be learned is that the validity of event studies which have not rejected the null hypothesis of no event effect should be questioned unless they are rerun using GLS methods.
I. Introduction

Event studies of stock returns have been used extensively in the 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 studies often conclude that abnormal returns at 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 inefficient and illustrate with data that the tests can be highly inefficient. We present a Generalized Least Squares (GLS) estimator which is more powerful for event studies. We also show how to calculate this GLS estimator with virtually no increase in computing difficulty.

We illustrate that, in practice, the procedure used to obtain our GLS estimator may require inverting a sparse matrix (a matrix dominated by zeros), which can lead to inefficiency unless one uses sparse matrix inversion algorithms which we show how to implement. With these tools we address the properties of hypothesis tests using financial data on firms that won Brazilian privatization auctions.

We demonstrate that, at least for our data, this GLS estimator is far more powerful, has (much) smaller $p$-values, than the estimators used in the literature. Indeed, we demonstrate that in some samples for which an event is significant at the 98% level using the GLS estimator, the traditional simple average/mean estimator used in practice fails to find significance at the 90% or even the 85% level! We provide an additional piece of evidence suggesting the superiority of our GLS estimator over the traditional estimator by conducting “reverse” Malatesta (1986) simulations. Malatesta’s (1986) simulation technique is one in which he artificially added an event to data where there was no event. We, on the contrary, reverse it and simulate using the entire data set after artificially reducing the measured event effects (subtracting from the true stock return in the event window). These full sample analyses indicate that, for example if we reduce the measured effects to the point at which our GLS estimator drops to a 95 percent
significance level, the traditional model’s power falls to under a 60 percent significance level. These pieces of evidence imply that studies which found abnormal returns not being significantly different from zero could have done so erroneously.

It is our contention there are probably some published event studies which incorrectly find insignificance of an event effect due to the 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 our GLS method.

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. Upon further investigation, we found Thompson (1985) proposed a GLS estimator with similar properties to ours which is more efficient than the OLS estimators used in the literature. 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 powerful. We use both a two step inverse variance weighted average (IVWA) approach and our more complete GLS approach. Furthermore, we furnish a real data set for which this problem is substantial showing the potential consequences of using the traditional procedures in actual estimation. Further, we note that the pooled GLS approach potentially involves sparse matrices which may require special computational methods to be efficient. In our data analysis, we find that the sparse matrix problem can indeed create
estimation problems which can be overcome using sparse matrix computational techniques. In an appendix we provide the SAS code for the interested reader.

The traditional methodology uses the following model to estimate abnormal returns. Suppose an “event” occurs at time zero for each firm (where time 0 is relative to the firm/event and is not a single unique time, and for simplicity assume that the event effect is realized in a single period). Then the analysis would first estimate \( R_{it} = \alpha_i + \beta_i R_{it}^m + \varepsilon_{it} \) for time periods \( t < 0 \), for each event, i, separately, where \( R_{it} \) is the return of firm i at time t (e.g., t measured in days) and \( R_{it}^m \) is the stock market index return for time t associated with the i\(^{th}\) firm’s event.

For the i\(^{th}\) firm, the estimated time zero event effect is then the deviation of the realized return at time zero from its predicted return based on the parameter estimates from the first stage regression, or \( \hat{\delta}_i = R_{i0} - (\hat{\alpha}_i + \hat{\beta}_i R_{i0}^m) \). Without loss of generality, we shall henceforth assume that the “alternative hypothesis” is that the event is “good,” i.e., \( \delta > 0 \). The second step is testing the hypothesis that \( \delta \) is positive by testing if the mean of the \( \delta \)’s is greater than zero, that is \( \overline{\delta} > 0 \). A third step in many studies attempts to “explain” the \( \hat{\delta} \)’s. This is done by regressing them on exogenous determinants which are thought to impact the event effects.

Typical event studies estimate the individual abnormal returns and then compute their simple mean. This mean is tested against the null hypothesis by calculating a ratio of this mean divided by a function of the individual estimated variances.\(^1\)

Returning to the second step, there is an equivalent estimator, the Ordinary Least Squares (OLS) regression of \( \delta_i = \beta_i X_i + \varepsilon_i \). Consider the standard regression notation of \( \beta_i = (X'X)^{-1}X'y \).

When estimating a constant only, \( X \) is the \( N \times 1 \) column vector of 1's, where \( N \) is the number of

\(^1\) This is typical, there are some variants, but they are similar, cf., Campbell et al. (1997) and Patell (1976).
events. Then \( \mathbf{X}'\mathbf{X} = \mathbf{N} \) and its inverse is \( \mathbf{1}/\mathbf{N} \). The dependent variable, \( \mathbf{y} \) in standard notation, is \( \mathbf{\delta} \). So \( \mathbf{y} \) is an \( \mathbf{N} \times \mathbf{1} \) vector containing the \( \delta_i \) measurements. Accordingly, \( \mathbf{X}'\mathbf{y} = \sum_{i=1}^{\mathbf{N}} \delta_i \). So the estimated coefficient is \( \mathbf{\beta} = \overline{\mathbf{\delta}} = \sum_{i=1}^{\mathbf{N}} \delta_i / \mathbf{N} \). The computed intercept is identically equal to \( \overline{\mathbf{\delta}} \).

Notice that this estimator does not depend on any variance components, as opposed to the t-statistics that do depend on them.

Crucially, then, we need to examine the assumptions of OLS. In Greene (2003, p. 42) assumption A4 is the homoscedasticity assumption. “Each disturbance, \( \mathbf{\varepsilon}_i \), has the same finite variance, \( \sigma^2 \)” (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 \( \delta_i \), one also derives their respective \( \delta_i \). If these are not a single common \( \sigma \), the classical (OLS) model is not efficient (hence not producing the best linear unbiased estimators). We demonstrate that the conventional modeling which does not weight the \( \delta_i \) by the inverse of their respective estimated \( \delta_i^2 \) 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 \( \delta_i \)’s are not identically equal, but all the covariances remain equal to zero\(^2\). I.e., some stocks follow the market index more closely than others. Consider the standard GLS notation of \( \mathbf{\beta} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} \), where \( \mathbf{\Omega} \) is

\(^2\) See Fama (1970).
the covariance matrix. Again we look only at the case in which the intercept is estimated. \( \mathbf{X} \) is still an \( N \times 1 \) column vector of ones, and assume \( \Omega \) is a diagonal matrix of variances for each of the estimated \( \delta_i \)’s, which in turn implies that the off diagonal covariances are equal to zero. For this illustration, consider a two observation regression. Then \( (\mathbf{X}' \Omega^{-1} \mathbf{X}) = \sigma_1^{-2} + \sigma_2^{-2} \) and

\[
\mathbf{X}' \Omega^{-1} \mathbf{y} = \sigma_1^{-2} \mathbf{y}_1 + \sigma_2^{-2} \mathbf{y}_2,
\]

so we arrive at the inverse variance weighted average of

\[
\hat{\beta} = (\sigma_1^{-2} \mathbf{y}_1 + \sigma_2^{-2} \mathbf{y}_2) / (\sigma_1^{-2} + \sigma_2^{-2}).
\]

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 \( \delta_i \)’s are identical (an outcome which will occur with probability measure of zero). This has been demonstrated by econometricians including Patell (1976) who showed how differences in precision can be incorporated into the estimators.

Obviously, the covariance matrix, \( \Omega \), contains covariance elements that are ignored in the OLS estimation. Some of our GLS efficiency gain may come from incorporating them in the estimator’s structure. But even when the actual covariance elements are negligible, our GLS estimator is more efficient than the equally weighted estimator due to the fact that our GLS structure weights the data in a more “sensible” way.\(^3\) As we shall see, in our data the covariance elements are negligible, nonetheless the use of our GLS estimator achieves a higher level of statistical power when compared with the simple mean (equally weighted) estimator.

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

---

\(^3\) By “sensible,” we mean in the sense of least squares methodology given the maintained hypotheses.
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. And last, we apply a simulation technique similar to
Malatesta (1986) which provides another piece of evidence that our GLS estimator is superior to
other procedures.

III. More powerful techniques

The econometric theory used herein can be demonstrated by examining the assumptions
employed in 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, \( \delta \). Then the
estimated abnormal return for the \( i^{th} \) event, \( \hat{\delta}_i \), is an estimate of the common \( \delta \) and there is a
normal distribution of errors around \( \delta \) across the set of estimates or events. A second possibility
is that there are idiosyncratic \( \delta_s \), but there is a central tendency which is distributed normally
around some \( \bar{\delta} \). (In what follows we will use only the first interpretation except where it is
useful to distinguish between the two.) If either of these alternatives is the true structure of the
\( \delta \)'s, then the traditional event study methodology is virtually never efficient except under highly
restrictive conditions.

The intuition for why our GLS is more efficient and the standard methodology is
inefficient and lacks power is quite rudimentary. Suppose the abnormal return across events
should be equal and positive and we have estimates of this return for two separate events. Let
the estimated abnormal return from the first event be 1.00 and the estimated abnormal return
from the second event be -2.00. Further suppose the first stage estimates lead to estimated
standard deviations for the first and second estimates of 0.10 and 4.00, respectively. The
abnormal return for the first event is ten standard deviations above zero while that of the second event is only one half of a standard deviation below zero and less than a full standard deviation below the first estimated positive abnormal return. A simple average of the two estimates would yield a negative value for the overall estimate of the common abnormal return; however, the simple average ignores the fact that the first positive estimate is more precise (has a smaller variance) than the second negative estimate. Consequently, the first estimate should be given greater weight in the calculation.

When using the two stage procedure for estimation in event studies, one should use the inverse variance weighted average (IVWA) of the estimated abnormal returns rather than the simple mean in the second stage of the estimation. In effect, the traditional methodology which computes the simple mean (i.e., OLS) discards information, the variances of the estimated \( \hat{\delta}_i \)s, generated in the first stage regression. Unless all of the first stage variances are identical (which would only happen in practice with probability measure zero), the second stage estimation cannot possibly be efficient.

Our benchmark GLS model is achieved by using a single one stage panel regression that combines all of the events’ data into one regression that covers both the pre-event and event window time periods for every event and uses dummy variables (also see Thompson 1985, section III-B). This panel not only has an event specific fixed effect, but also has an event specific pre-event slope coefficient between firm returns and market returns (a slope fixed effect) as in the two stage models.

---

4 Consider the example above with point estimates of 1.00 and -2.00 and standard deviations of 0.10 and 4.00. The simple mean is -0.50 whereas the inverse variance weighted average is 0.99812617 using weights of 0.99937539 and 0.00062461.

5 OLS produces the best linear unbiased estimators only if each variance is a constant, \( \sigma^2 \), for all i. Often one does not have the information to assume a more general variance structure; however, during the first stage of estimation in an event study the same program which prints out the \( \hat{\delta}_i \)s will print out the \( \sigma_i \)'s for each event.
We also show there is the potential for another form of inefficiency which arises if one uses our proposed GLS methodology: numerical instability. There is a literature on how sparse matrices may lead to computational inaccuracies (instability) in the standard matrix inversion routines. We explain the relevance to our methodology after presenting more details about our approach. We then show how to overcome this potential estimation instability associated with our proposed model.

In principle the traditional simple means method and/or the inverse variance weighted average (IVWA) method could yield results very similar to the benchmark GLS estimator we propose. In practice neither may do so.

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

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

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 \( i \)th firm’s event,\(^6\)

and

\( \varepsilon_{it} \) is an unsystematic error term with \( E(\varepsilon_{it}) = 0 \) and \( E(\varepsilon'\varepsilon) = \sigma^2 I \), where \( \sigma \) is the assumed common variance and \( I \) 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.

\(^6\) For example, time \( t = -5 \) is five days before event \( i \). Since different events happen at different times, the calendar date for \( t = -5 \) will vary across events.
of stock trading days).

Define the event effect for one period (period zero)\(^7\) for the \(i\)th firm as

\[ \hat{\delta}_i = R_{i0} - (\hat{\alpha}_i + \hat{\beta}_i R_{m0}) \]  

where \(\hat{\alpha}_i\) and \(\hat{\beta}_i\) are the parameter estimates from equation \([1]\). This is the difference between the \(i\)th firm’s realized return in time zero and its predicted return in time zero based on the model in \([1]\) along with the stock market return index in time zero. If the \(\hat{\delta}_i\)'s are independent estimates of a common \(\delta\), then one can test the hypothesis that \(\delta\) is positive (without loss of generality) by looking at the standard event study hypothesis

\[ \bar{\delta} > 0. \]  

By computing the simple mean of the \(\hat{\delta}\)'s, in effect one is equally weighting the measurements, just as in the classical linear model (OLS). As Greene (2003) points out, an underlying assumption of this model is “homoscedasticity,” implying each delta has the same distribution of measurement error, or has a common variance, \(\sigma^2\)'s are the same across all firms/events. If these assumptions are not met, the classical model is inefficient, that is, it does not generate the best linear unbiased estimator of the true underlying \(\delta\). How poorly the classical model performs in estimating the true \(\delta\) is an empirical question which will vary by data set.

**B. Alternative event study model specification for a common event effect**

Consider the following structure. The exact same model as before can be written as

---

\(^7\) The formulae are somewhat more complex if the event window is more than one “day,” but for our purposes there is nothing added from complicating the expression for the event window.
As Fama (1970) notes, under the efficient markets hypothesis, serial correlation should be equal to zero. Thompson (1985) and Malatesta (1986) demonstrate how to deal with contemporaneous covariances in models with contemporaneous time periods using a variance/covariance matrix in the form of $G^q I$, where $G$ is the contemporaneous variance/covariance matrix. We are dealing with non-contemporaneous events, but where our data periods per event are allowed to overlap. However, their method of estimation can easily be incorporated into ours. With a random walk and the inclusion of the market return as a regressor, covariances being equal to zero is a reasonable assumption given Fama’s (1970) observation about efficient markets.

$$R_{it} = \sum_{j=1}^{N} D_{ij} (\alpha_j + \beta_j R_{mt}) + D_{ij}^{0} \delta + \epsilon_{it}$$  

where: $D_{ij} = 1$ if $i = j$, and $= 0$ for $i \neq j$,

$D_{ij}^{0} = 1$ for event $i = j$ when $t = 0$ (the event window), and zero otherwise,

$\delta$ is the abnormal return which is assumed to be common across all events,

$\alpha_j$ and $\beta_j$ are the parameters which are estimated over the estimation window and are used to form the predicted returns during the event window in order to create the difference between the estimated return and the actual return for time $t = 0$ (the event time), and

$\epsilon_{it}$ is an unsystematic error term with $E(\epsilon_{it}) = 0$ and $E(\epsilon' \epsilon) = \Omega = (\sigma_i^2 \otimes \mathbf{i}) \times \mathbf{I}$, where $\sigma_i^2$ 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.

The methodology in [4] adds one time period to the pre-event window and adds a dummy variable which captures the abnormal returns associated with the event effect which is assumed to be constant across all events. One should note, there are numerous terms being multiplied by zeros in [4]. We could have structured [4] to simply reflect zeros for these terms, but the usefulness of constructing the model as above will become apparent when we move to the matrix equivalent model in [5].

If the model in equation [1] is efficient for each event taken separately, [2] gives the

---

As Fama (1970) notes, under the efficient markets hypothesis, serial correlation should be equal to zero. Thompson (1985) and Malatesta (1986) demonstrate how to deal with contemporaneous covariances in models with contemporaneous time periods using a variance/covariance matrix in the form of $\Sigma \otimes \mathbf{I}$, where $\Sigma$ is the contemporaneous variance/covariance matrix. We are dealing with non-contemporaneous events, but where our data periods per event are allowed to overlap. However, their method of estimation can easily be incorporated into ours. With a random walk and the inclusion of the market return as a regressor, covariances being equal to zero is a reasonable assumption given Fama’s (1970) observation about efficient markets.
event specific estimate of the abnormal return, $\hat{\delta}_i$, and the true abnormal return, $\delta$, is common across all events, then our GLS model in [4] is efficient, indeed $\delta$ is BLUE under these assumptions. However, the traditional simple average (or equivalently the OLS estimator) is not efficient unless all the variances of the estimated $\hat{\delta}_i$s are identical across all events.

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 ($N = 2$):

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

This nests equations [1] and [2] along with the assumption of a common $\delta$. In place of the $\bar{\delta}$ in equation [3], the equations in [5] directly estimate $\hat{\delta} = \bar{\delta}'$, where $\bar{\delta}'$ is the inverse variance weighted average 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 inverse variance weighted average
of δ will be virtually identical to [5].

The technical issue arises in the construction of the data matrix. To illustrate, now suppose \( T = 10 \) and \( N = 2 \). Then there are eleven non-zero data elements and eleven elements identically equal to zero in each of the first four columns of the data matrix and the fifth column contains twenty 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% 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

\[ \text{\footnotesize 9 Suppose that [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 \([-T,\ldots,0]\).} \]

\[ \text{\footnotesize 10 This has been an intuitive explanation. More precisely, let } T_1 \text{ be the number of periods in the estimation window (250 in our data), } T_2 \text{ be the number of periods in the event window (1 in our data), } n \text{ be the number of events (71 in our data), and } k \text{ be the number of explanatory variables (1 in our base case and 7 in the expanded case). Then the percentage of elements identically equal to zero out of all the elements in the matrix to be inverted, } X'\Omega^{-1}X, \text{ is } 100 - 100((4n+4nk+k^2)/(2n+k)^2), \text{ which is 89.55\% in our model with 7 explanatory variables. As } n \rightarrow \infty (4n + 4nk + k^2)/(2n + k)^2 \text{ approaches 0, implying the percentage of zeroes approaches 100\%.} \]
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 estimation problems. Computationally, conventional statistical programs treat zeros as floating point approximations of zero when inverting the $X^\prime X$ matrix. In our case we are inverting $(X^\prime \Omega^{-1} X)$, where $\Omega$ is a diagonal matrix so implicitly the number of zeros in the estimation remains the same. This can create rounding errors which increases as the number of zeros increases. The computational problems involved with such cases are covered in Thisted (1988), and we demonstrate that this leads to parameter estimate instability in our second set of tests where we replace $\delta$ with seven independent variables to explain the event effect.

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 $X^\prime X$ matrix. Following this technique we can obtain an analytic expression for the common coefficient (or common coefficients), $\delta$, as well as the event specific parameters $\{\alpha_t, \beta_t\}$. 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 the 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. Since we are inverting the slightly more complex
(\(X'\Omega^{-1}X\)) matrix, this requires a proposed “starting point” for the iteration of the variance estimates as well. (We start with the individual event specific OLS estimates of these variances by using the root mean squared error of the residuals.)

We develop a similar matrix inversion technique for sparse matrices, one which is computationally more convenient than that used in earlier papers following this literature. We introduce this new algorithm in Appendix A and follow with the SAS code a practitioner could use to implement this algorithm. There is always the question of whether or not to use the sparse matrix methods for any given data set. It is possible that for a given problem a simple “off-the-shelf” maximum likelihood (ML) program will give close to identical answers to the sparse matrix GLS which we describe herein. However, it is also possible that the ML program will be inaccurate and provide different estimates. In either case, our GLS program with the sparse matrix method is efficient, yields the best linear unbiased estimators. Deviations from this benchmark may or may not be serious. It cannot be known if one is deviating from the best linear unbiased estimator without undertaking the more complex (but “weakly”11 superior) GLS procedure we describe.

C. Alternative specification for event specific abnormal returns

Event studies often drop the assumption of a common \(\delta\), where measurement errors lead to differences in the \(\delta_s\), after testing for the sign of \(\delta\). They hypothesize that the “measurements” in the above model are, in part, explainable by exogenous observable event

\(^{11}\) As a few more poorly estimated values are added to the sample, the significance level of the simple average can rise or fall. But, the expected value of the significance from adding poorly estimated values falls with each such value. In the limit, if the estimates are very inaccurate in terms of larger \(\sigma\)’s, the model becomes one in which the level of significance will converge to zero, and the probability of the estimated mean carrying the correct sign will converge to 0.5, regardless of the true value of the underlying parameter.
specific factors. They then perform a regression of abnormal returns on firm/event specific explanatory variables such as equation [6]

$$\hat{\delta}_i = \Phi X_i + u_i.$$  \hfill [6]

The $X_i$ matrix includes firm or event specific variables which might explain the differences in the $\delta_i$'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 inverse variance weighted average as applied to the regression in [6]. This would be an inverse variance weighted regression in place of [6]. As in the last section, given the assumptions of the system (the equations in [1] yield the best linear unbiased estimators for each event, [2] represents the abnormal return, and [6] represents the common parameters explaining abnormal returns) our benchmark GLS model will be efficient, whereas the traditional estimation of [6] will not be efficient unless all the variances are identical.

For simplicity, we again present an $N = 2$ version of [6], where $\delta_i = \phi^0 + \phi^1 X_i + \epsilon_i$ 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
Thus, [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.)

**D. The intuition behind the power of our tests**

Finally, before turning to an example, we discuss the issue of power. We address this formally in Appendix B. For clarity of the intuition, we will consider this in the context of the first hypothesis, a uniform \( \delta \) across events. Suppose there is a single true \( \delta > 0 \), but there exists measurement error so that \( \hat{\delta}_i = \delta + \varepsilon_i \) is observed. As the variance of \( \varepsilon_i \) approaches zero, all of the observed \( \hat{\delta}_i \) will be positive. In the limit, as the variance approaches infinity, we expect to observe half of the observed \( \hat{\delta}_i \) being negative and the probability of the mean of the observed \( \hat{\delta}_i \) being negative approaching (but never reaching) 0.50.
Consider a sample of events composed of two subsamples, one for which the $\delta$’s are measured with minimal error and the other with substantial error. Those with substantial error, many of which would be estimated as negative, would be pooled with the minimal error measurements. The expected value of $\bar{\delta}$ remains unchanged, but the precision of its estimate falls, and the probability of finding a negative mean rises relative to the case in which all the $\delta$’s are measured almost correctly. The greater the proportion of observations with substantial errors, the greater the tendency towards failing to reject the null hypothesis of $\delta = 0$ (regardless of whether the point estimate of $\bar{\delta}$ rises or falls\(^{12}\)). Using the inverse variance weighted average model increases the probability of rejecting the null hypothesis when it is false as compared to the traditional test: it increases the power. Thus, it is crucial when doing hypothesis tests to use our GLS estimator in order to draw the correct conclusions.

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, with an emphasis on privatization acquisitions. It is noteworthy to mention here that this review contains numerous publications in which the authors could not reject the null hypothesis of an event effect equaling zero, exactly what we show the traditional (OLS) methodology might be biased to conclude.

As for the previous event study techniques, Thompson (1985) derives various\(^{13}\)

\(^{12}\) This is a global statement, not a local one. Even if $\bar{\delta}$ rises somewhat with this increase in the proportion of poorly estimated observations, its variance will have risen as well.

\(^{13}\) Thompson (1985) presents two GLS 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
estimators for event studies in pooled forms, one of which is similar to 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.

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 correlations, or cluster effects. However, he assumes the events themselves do not have to be contemporaneous. Additionally he assumes each firm experiences a different true $\delta_i$. He performs tests on $\delta$ using simulated data. He collected data on four samples of 80 non-overlapping months. He drew at random 50 firms for each sample and posited 20 firms with no event effects and for 30 firms he artificially created an event effect by adding 1 percent abnormal return on some random date in the 80 month window. He then tested the efficiency of his test statistics. He found no evidence that the more sophisticated estimators which used a form of GLS (with a different structure than ours) were
more efficient than the theoretically less efficient estimators. His simulations suggest that modeling contemporaneous correlations may not significantly improve estimation.¹⁴

Our estimation is different in several ways. First our estimation period for each event starts 250 days (one year) before the event. Given our five years of events, many days across events are not contemporaneous, indeed many pre-event windows do not overlap with some of the other pre-event windows. Second we do not assume contemporaneous correlations between firms, as the right side includes the entire market return. We will, however, do simulations in the spirit of his simulations, i.e., artificially subtracting something from observed abnormal returns to see how this affects results.

B. The data

In 1990, eighty of the 500 largest non-financial enterprises in Brazil belonged to federal or state governments. These companies accounted for 37% of GDP, 63% of total net worth, and 75% of total fixed assets. Although the government allowed for privatization auctions in 1990, due to the political and economic instability, the primary phase of privatization auctions did not occur until the mid 1990s, the period of our study. Most of the privatization auctions occurred from 1995 to 2000. Unlike privatization auctions in many other countries, the Brazilian auctions were first price sealed bid auctions each held on a single day. Participation included not only Brazilian firms, but also foreign firms which were allowed to bid in consortiums with domestic firms. There are 71 acquirer-privatized pairs involving at least two approved bidders in our analysis.

The data set contains stock prices and local market indices. The following

¹⁴ Thompson (1985) suggests that the form of his simulation may be partially responsible for this negative finding.
transformations yielded the stock and market returns,

\[ R_{it} = \ln[p_{it}] - \ln[p_{it-1}], \]

where \( R_{it} \) is firm i’s return on day \( t \) derived from stock prices, \( p_{it} \),

\[ \tau = t, t-1, \text{ and} \]

\[ R_{it}^{m} = \ln[p_{it}^{m}] - \ln[p_{it-1}^{m}], \]

where \( R_{it}^{m} \) is the market index return on day \( t \) derived from

market indices, \( p_{it}^{m}, \tau = t, t-1. \)

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 and its announced winner. This data is necessary for our first analysis based on the
assumption of a uniform privatization effect across auctions. Once we drop the assumption of a
uniform effect across auctions, we then model event effects as a function of the characteristics of
the events and the winners.

Modeling these characteristics includes creating a dummy variable for the nationality of
the acquiring firm. How good a deal is struck in terms of a percent of the value of the acquired
firm when reflected in the acquiring firm’s value suggests that a relative size variable should also
be included.\(^{15}\) We also include a measure of how closely related were the businesses of the
acquiring and acquired firms. The relatedness variable is zero when the acquiring and acquired
firms have matching DataStream Advance industry codes, otherwise this is set to one.

We want to analyze winning competitive bids, so we include a participation indicator
variable. There were 71 auctions for which there were two or more approved bidders. On the
day of auction, if only one bidder actually arrives, the bidder knows it can guarantee the
winning bid at the reservation price. We assume that if the winning bid is equal to the

\(^{15}\) Note, for a bidding consortium, this must be adjusted by share. Also note, five
auctions where the buyer’s stocks had extremely low capitalization are excluded from the study.
government set reservation price, this indicates there was only one firm bidding. If the winning bid is the reservation bid we define the Participation dummy variable to be zero, otherwise this is set to one.

The data source for the location of the acquirers’ headquarters was Bloomberg, an online financial service which provides quotes and technical analysis of securities as well as company and industry information. 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 sub sample 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 treats the 71 observed abnormal returns as a random variable, as if there were 71 independent observations of \( \delta \) measured with noise, which are homoscedastic and the covariance across events equals zero.\(^{16}\) Calculations can then be undertaken to find the simple mean and the \( t \)-values under the null hypothesis of abnormal returns equaling zero versus the mean that was just calculated. We contrast this with the inverse

\(^{16}\) Recall, this is statistically the same as not assuming a uniform effect, but rather assuming the individual \( \delta_i \)'s are unique and distributed normally around a mean \( \delta \).
variance weighted average and with our single equation GLS model in [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. We designate the “Our GLS” model as the same model structure as in ML, but using the sparse matrix methodology presented above. The results are summarized in Table 1.
Campbell et al. (1997) suggested another methodology. They proposed normalizing the measurements by the inverse of their standard deviations as one test. They claimed, “if the true abnormal return is constant across securities then” inverse standard error normalization should be used. But, “if the true abnormal return is larger for securities with higher variance, then the better choice” is the simple mean. Performing the test in this fashion provides the intermediate result of a $t$-value = 3.16 and a $p$-value = 0.008.

Positing these reasons for using a different weighting appears to be *ad hoc*. A different weighting method must be based on a structural change in the data. As we point out (see note 3), only heteroscedastic estimators (*e.g.*, inverse variance weighting) can be BLUE under the maintained hypotheses.

<table>
<thead>
<tr>
<th></th>
<th>IVWA</th>
<th>ML (Proc Mixed)</th>
<th>Our GLS (Benchmark)</th>
<th>Traditional$^{17}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal Return</td>
<td>0.69%</td>
<td>0.70%</td>
<td>0.70%</td>
<td>0.62%</td>
</tr>
<tr>
<td>$t$-value for $H_0$: $\delta = 0$</td>
<td>3.54</td>
<td>3.54</td>
<td>3.58</td>
<td>1.77</td>
</tr>
<tr>
<td>$p$-value (one tailed)</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.038</td>
</tr>
</tbody>
</table>

---

$^{17}$ Campbell et al. (1997) suggested another methodology. They proposed normalizing the measurements by the inverse of their standard deviations as one test. They claimed, “if the true abnormal return is constant across securities then” inverse standard error normalization should be used. But, “if the true abnormal return is larger for securities with higher variance, then the better choice” is the simple mean. Performing the test in this fashion provides the intermediate result of a $t$-value = 3.16 and a $p$-value = 0.008.

Positing these reasons for using a different weighting appears to be *ad hoc*. A different weighting method must be based on a structural change in the data. As we point out (see note 3), only heteroscedastic estimators (*e.g.*, inverse variance weighting) can be BLUE under the maintained hypotheses.
The “Traditional” model simply lacks power relative to the other estimates. Although the results for the traditional model are positive and significant,\textsuperscript{18} they are slightly different in absolute value and are far less significant or precise than any of the three other models. The other three methods have virtually identical results. Utilizing the heterocedastic methods generates \( p \)-values which are remarkably lower, having fallen by over 99\% from the value of the traditional model. Note in particular the power of the Inverse Variance Weighted Average (IVWA) model. This approach simply uses a traditional two step methodology, but rather than evaluating the mean of the abnormal returns random variable from the first stage, it instead looks at the inverse variance weighted average, as should be done for efficiency if the variances are available (as they are in every event study) and the covariances are equal to zero. The information for performing this test is generated when researchers run the traditional model, but they typically do not use this variance information.

To reiterate the above, our GLS model produces the best linear unbiased estimator under our maintained hypotheses and our maintained hypotheses are implicitly the ones underlying the typical traditional event study. A two stage IVWA may almost always be as good as our GLS procedure, which being the “Best,” should be the benchmark. But, this is impossible to know \textit{a priori}. The only way to find out in any specific study is to run our GLS model, which is (at worst weakly) superior to any two stage model. The simple mean traditional model is more

\textsuperscript{18} Many studies hypothesize that the returns to acquiring firms would be zero, due to competition or hard bargaining by sellers, or are positive. That is, one tailed tests would be appropriate. If one felt that most acquisitions were for “empire building,” \textit{e.g.}, growth maximization (as a few studies have claimed), then a two tailed test would be appropriate and the Traditional Model would lead to a less powerful \( p \)-value, one not even significant at the 5\% level in our data.
likely to fail to reject the null hypothesis of no event effect as long as the regression fits in the pre-event windows are not all of equal precision.

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, suppose that the estimated $\delta$’s with low variances have large values and the high variance $\delta$’s have small or negative values. This could be driving the huge gap in the $p$-values between the traditional model and the heterocedastic models. Were this the case, our basic conclusions would remain correct, but the magnitude of the estimated increase in efficiency might be driven by the sample characteristics.

To test for this we regress $\hat{\delta}$ on its estimated variance, $\sigma_\delta^2$. The results are

\[ \hat{\delta} = 0.0068 - 0.57 \sigma_\delta^2. \]

(2.79) (0.56)

There is certainly no significant negative relationship, but were some of the variances extremely large, the negative point estimates could be driving the results. To illustrate this is not the case, we present the scattergram from this regression
There is virtually no 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.

We next examine sample subsets to demonstrate how weak the traditional methodology can be in practice. Although the traditional model has a $p$-value which is significant at the 5% level ($p = 0.038$) for the full sample (on a one tailed test), we use subsamples to illustrate how easily the traditional model might fail to find significance. We ordered the sample observations
from the estimates with the lowest estimated variance of \( \delta \) to the highest estimated variance of \( \delta \). 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 \( \delta \). 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 \( \delta \) and repeating the same procedure. The resulting \( p \)-values are illustrated in the following figure.
The traditional model only meets the 95% significance level for five of the subsamples. Examining the 37 subsamples for which the IVWA model satisfies the 95% significance level, nine of these subsamples do not even satisfy an 85% significance criterion for the traditional model (twenty-four do not even satisfy the 90% level). There is one truncation for which the IVWA estimator is significant at the 98% level, whereas the traditional estimator does not even rise to the 85% significance level.

Another piece of evidence suggesting the superiority of a heterocedastic estimator over the traditional estimator is provided by conducting a “reverse Malatesta” (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 3, for each fraction of the standard deviations, we plot the p-values of three alternative methods: our GLS, the IVWA, and the traditional. As the fraction increases, the greater the disturbance, and the more is subtracted from the original event effect. Consequently, the p-values decrease as the simulated event effect approaches zero.

Figure 3 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 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 illustrated in Figure 3.
In other studies which have found insignificant $t$-values, there is at least a *prima facie* case for reexamining these results and for downgrading their relevance until having done so. For event studies which use simple means and reject the null hypothesis, one would expect to find the rejection of the null hypothesis to be even stronger using heterocedastic techniques (either the IVWA or our GLS method).

To reconcile our findings with Malatesta (1986), our conclusions differ from the ones obtained through Malatesta’s simulations due to the fact that Malatesta does not simulate an estimator like ours. He was concerned with evaluating the impact of the cluster effects on event studies and he examined improving the analysis of contemporaneous covariance effects, he does not look at a GLS like ours. It should be noted that neither Malatesta nor we claim that GLS results would be significantly stronger for every data set, which data sets this would true of cannot be predicted before running the tests.

**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 $\delta_i$s. The first step, estimation of $\bar{\delta}$ as above, is consistent with either the assumption of a single $\delta$ with estimation error of $\delta_i=\delta+\epsilon_i$, where the $\epsilon_i$s are normally distributed or the assumption that the $\delta_i$s vary across firms but are normally distributed. The former interpretation is not consistent with estimating the determinants of the $\delta_i$s, and the latter is.

In estimating the determinants of the $\delta_i$s, we present four models analogous to those we ran for the common $\delta$: 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), our GLS model using sparse matrix methods, and the traditional two stage model for comparison.

The inclusion of variables here follow 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; 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 now examine the results found in Table 2. Our GLS model yields the best linear unbiased estimates, given the statistical assumptions (cf. Thompson 1985), and avoids any problems associated with inverting sparse matrices. Accordingly this model serves as the
benchmark for evaluating the other models. The results are not strong, but provide methodological insights. The main “result” is crucial. Unlike the tests in Table 1, which asks if the true $\delta$ is positive, where the estimated results do not depend on the usage of the sparse matrix routines, for the abnormal returns explained by exogenous data model, our GLS procedure requires sparse matrix methods to assure efficient estimation.

The Participation variable is not significantly different from zero in all of the regressions. If the equilibrium number of bidders is endogenous, this should not be surprising, thus we cease providing any further consideration to this variable. In our GLS model the Relative Size variable is highly significant and with the correct sign. The Relative Size interaction terms are of the opposite sign, where the two significant terms are potentially offsetting the effects of the Relative Size term.

The important implication of these results, however, is the methodological point: sparse matrix methods are needed for efficiency. This result differs from what we found with the first hypothesis, tested in Table 1. For the earlier hypothesis, the IVWA, the ML, and our GLS models were similar in both parameter estimates and in $t$-values. Only the traditional model had a somewhat different parameter estimate and a considerably higher $p$-value.

For the second hypothesis in Table 2, the models each have results that differ from the parameter estimates in our GLS model, but each looks to be a reasonably good fit ($t$-values) when taken alone and out of context. In general, the magnitudes, signs, and significance levels are fairly comparable to those in our GLS estimation. Notable exceptions are the sign differences on Participation in the inverse variance weighted model and the intercept term in the ML model. At first glance, the Traditional model seems to perform well relative to our GLS results. However, the Traditional model finds the intercept to be significant at the 95% level.
(two tails), whereas our benchmark GLS model fails to have a significant intercept. A Chi-square test of the null hypothesis that the parameter estimates in our GLS and traditional models are the same yields a value for the test statistic of 12.62 ($\sim \chi^2$ with 8 degrees of freedom). This implies that we fail to reject the null hypothesis that the estimates are the same at the 90% confidence level, but just barely. However, we would reject the null hypothesis of equivalence at the 85% level, so the parameter estimates are quite close to being statistically different at generally applied thresholds.\textsuperscript{19, 20}

How general the instability might be in practice is unclear. The independent variables in this data have significant multicollinearity. The correlation matrix reveals that out of the 28 covariances in the data, 16 exceed 0.20 and 4 exceed 0.70. With high multicollinearity the covariances between the coefficients on the variables is presumed to be high. This may have an effect on the inversions of the whole matrix but not on the sparse matrix methods.

If we examine the benchmark model it is clear that the other three approaches are not closely approximating the parameter values from our GLS estimation using the sparse matrix inversion methodology.\textsuperscript{21} The inverse variance weighted model and the Traditional method find one parameter, the intercept, to be negative and significant at the 90% and 95% levels, respectively. Both one-step approaches have insignificant estimates of this same parameter with

\textsuperscript{19} As is usual in Type-I – Type-II error models, we want a much higher standard for rejecting the null if our hypothesis is that the null is incorrect. Here we are asking if the Traditional method replicates our benchmark GLS model, there is only a 0.15 probability that this is the case.

\textsuperscript{20} The rejection levels at the 90% level and the 85% level are 13.36 and 12.03, respectively.

\textsuperscript{21} The traditional model is not so obviously worse in matching the benchmark model for this hypothesis. Why this is so would require analyzing more data sets. If the size of the parameter estimates is important (not simply their $t$-values), it performs poorly relative to the benchmark.
the ML estimator having a positive sign. *A priori* we must assume that some of other event studies are drawing the wrong conclusions from the data by not using what we call our GLS estimator, the uncertainty is only about the frequency of such incorrect inferences.

Before concluding, it is worth noting the contrast between the common $\delta$ and the event specific $\delta$ cases. As we conjectured for the common $\delta$ 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 $\varphi_i$ 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 $\delta$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. We first examine the traditional methodology used by practitioners and demonstrate that it employs the assumptions of the classical homocedastic (constant variance) model, which rarely occurs in reality. We provide the intuition for why one should also examine a particular GLS estimator, as has been rarely done in the past (an exception being Thompson (1985)).

The next question is whether the assumptions necessary for the traditional methods are “close enough” for accurate estimation. Malatesta (1986) addressed the question of whether modeling clusters, or contemporaneous covariances, would improve estimation, and found little effect using a simulation method. Thompson (1985), however, opined that the form of the
simulation may have been one reason for not finding that the GLS estimation led to stronger results. Neither Malatesta nor we would claim that GLS results would be stronger for every data set. But maybe the fact that Malatesta didn’t find an increase in power by using GLS (for a different situation than we model) explains why a leading text, Campbell et al. (1997), a decade later did not pursue Thompson’s (1985, section III-B) GLS model and that practitioners over the last two decades have not taken advantage of his insights.

We look more closely at a model similar to that proposed by Thompson (1985) in his section III-B. We contrast the traditional two step approach to estimation, looking typically at averages of estimated abnormal returns, with a very simple inverse variance weighted average (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. 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.

We also demonstrate that sparse matrix problems may occur when using our pooled GLS 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.

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 heterocedastic methodologies (either IVWA or our GLS). Marginally significant results may in fact have underlying data which would provide much stronger results once one considers the variances of the first stage estimates in a heterocedastic framework.
Appendix A: Computations Given Sparse Matrix Problems

We present the theory underlying our new sparse matrix algorithm and then present the SAS code for the implementation of this new algorithm.

Recall equation [5], an example containing only two events:

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

To further simplify the notation, assume \(T = 2\) so that the estimation window contains only two periods and the event window lasts only one period. Then [5] can be written as

\[
\mathbf{R} = \mathbf{D} \alpha + \mathbf{R}^m \beta + \mathbf{J} \delta + \varepsilon
\]

where:

- \(\mathbf{R}\) is the column vector \([R_{1,-2} \ R_{1,-1} \ R_{1,0} \ R_{2,-2} \ R_{2,-1} \ R_{2,0}]\)',
- \(\mathbf{D}\) is an \((N(T+1) \times N) = (6 \times 2)\) matrix where the first column is the column vector \([1 \ 1 \ 1 \ 0 \ 0 \ 0]'\) and the second column is the column vector \([0 \ 0 \ 0 \ 1 \ 1 \ 1]'\),
- \(\alpha\) is the column vector \([\alpha_1 \ \alpha_2]'\),
\[
\mathbf{R}^m = \begin{bmatrix}
R_{1,-2}^m & R_{1,-1}^m & R_{1,0}^m & 0 & 0 & 0 \\
0 & 0 & 0 & R_{2,-2}^m & R_{2,-1}^m & R_{2,0}^m
\end{bmatrix}', \text{ is an } (N(T+1) \times N) = (6 \times 2)
\]

matrix,

\( \beta \) is the column vector \( [\beta_1 \beta_2]' \).

\( \mathbf{J} \) is a column vector of time dummy variables \( [0 \ 0 \ 1 \ 0 \ 0 \ 1]' \), coded 1 when the event occurs and 0 otherwise,

\( \delta \) is the abnormal returns parameter, assumed to be identical across events, and

\( \mathbf{e} \) is the error term column vector, for which

\[
\mathbf{V} = \begin{bmatrix}
\sigma_1^2 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma_1^2 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma_1^2 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma_2^2 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_2^2 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma_2^2
\end{bmatrix}.
\]

Again, to keep the notation simple, we transform our model to solve an identical problem in a simpler way. Our model is of the form

\[
\mathbf{R}_{it} = \alpha_i + \beta_i \mathbf{R}_{it}^m + \delta \mathbf{R}_{it}^m \mathbf{D}_{it} + \mathbf{e}_{it}.
\]

The term \( \alpha_i \) is mathematically identical to what is called a fixed effect in panel data. In our above formulation
we have five parameters to estimate, $\alpha_1, \alpha_2, \beta_1, \beta_2, \delta$. Both the $\alpha$’s and $\beta$’s are firm/event specific.

To demonstrate the calculations with a minimum amount of notation, eliminate the $\alpha$’s from the formulation, using standard panel firm/event fixed effect notation, by transforming the data using differences from means. Or

$$M_d R = M_d D \alpha + M_d R^m \beta + M_d J \delta + M_d \varepsilon,$$

$$M_d = \begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix}$$

where $M$ is a $3 \times 3$ matrix (as are the matrices of zeros).

The $M$ matrix, $M = I - 1/3 \ast (i i’)$ where $i$ is a $(3 \times 1)$ column vector of 1’s, transforms each element of the matrix it is multiplied with to the difference from its mean value.

Noting that $M_d D \alpha = 0$, $M_d \varepsilon = \xi$, and $E[M_d \varepsilon] = E[\varepsilon] = 0$, we can write

$$M_d R = M_d [R^m J] \begin{bmatrix} \beta \\ \delta \end{bmatrix} + M_d \varepsilon \text{ or, more simply, } \tilde{R} = Z \gamma + \tilde{\xi}.$$

This notation reveals in standard form how to obtain the GLS estimator for the parameters, $\hat{\gamma} = [Z’V^{-1}Z]^{-1}[Z’V^{-1}\tilde{R}]$.

Returning to the more complex notation, the coefficients

$$\begin{bmatrix} \hat{\beta} \\ \hat{\delta} \end{bmatrix} = \left[ M_d^T \varepsilon \varepsilon^{-1} M_d [R^m J] \right]^{-1} \begin{bmatrix} R^m \\ J’ \end{bmatrix} M_d \varepsilon^{-1} M_d R.$$

42
Defining $\mathbf{R}^m_d = \mathbf{M}_d \mathbf{R}_m$ and $\mathbf{J} = \mathbf{M}_d \mathbf{J}$, then

$$
\begin{bmatrix}
\mathbf{\hat{R}}_m' \\
\mathbf{\hat{J}}'
\end{bmatrix}
= \left[ \begin{bmatrix}
\mathbf{R}^m_d \\
\mathbf{J}'
\end{bmatrix}
\right] V^{-1} \left[ \begin{bmatrix}
\mathbf{R}_m \\
\mathbf{J}
\end{bmatrix}
\right]^{-1} \left[ \begin{bmatrix}
\mathbf{R}^m_d' \\
\mathbf{J}'
\end{bmatrix}
\right] V^{-1} \mathbf{R}.
$$

This implies

$$
\begin{bmatrix}
\mathbf{\hat{R}}_m \\
\mathbf{\hat{J}}'
\end{bmatrix}
= \left[ \begin{bmatrix}
\mathbf{R}_i \\
\mathbf{J}_i'
\end{bmatrix}
\right] \frac{1}{\sigma_i^2} I \left[ \begin{bmatrix}
\mathbf{R}_i \\
\mathbf{J}_i'
\end{bmatrix}
\right]^{-1} \left[ \begin{bmatrix}
\mathbf{R}_i \\
\mathbf{J}_i'
\end{bmatrix}
\right] \frac{1}{\sigma_i^2} I \mathbf{R}_i,
$$

where

$$
\tilde{\mathbf{R}}_i^m' = \begin{bmatrix}
R_{i,0}^m & R_{i,1}^m & R_{i,2}^m \\
0 & 0 & 0
\end{bmatrix},
\tilde{\mathbf{R}}_2^m' = \begin{bmatrix}
0 & 0 & 0 \\
R_{2,0}^m & R_{2,1}^m & R_{2,2}^m
\end{bmatrix}
$$

and

$$
\mathbf{J}_i = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}.'
$$

So

$$
\begin{bmatrix}
\mathbf{\hat{R}}_m \\
\mathbf{\hat{J}}'
\end{bmatrix}
= \left[ \begin{bmatrix}
\mathbf{R}_i \\
\mathbf{J}_i'
\end{bmatrix}
\right] \frac{1}{\sigma_i^2} I \left[ \begin{bmatrix}
\mathbf{R}_i \\
\mathbf{J}_i'
\end{bmatrix}
\right]^{-1} \left[ \begin{bmatrix}
\mathbf{R}_i \\
\mathbf{J}_i'
\end{bmatrix}
\right] \frac{1}{\sigma_i^2} I \mathbf{R}_i.
$$
where

\[
\tilde{R}_i = [R_{i,-2} - \tilde{R}_i, R_{i,-1} - \tilde{R}_i, R_{i,0} - \tilde{R}_i].
\]

Therefore,

\[
\begin{bmatrix}
\begin{bmatrix}
\Sigma (\tilde{R}_{1,t})^2 / \sigma_1^2 \\
0 \\
\Sigma (\tilde{R}_{1,t})^2 / \sigma_2^2 \\
\Sigma (\tilde{R}_{2,t})^2 / \sigma_1^2 \\
\Sigma (\tilde{R}_{2,t})^2 / \sigma_2^2
\end{bmatrix}
& \begin{bmatrix}
\tilde{R}_{1,0} - \frac{t}{T+1} \\
\tilde{R}_{2,0} - \frac{t}{T+1} \\
\tilde{R}_{1,t} \\
\tilde{R}_{2,t}
\end{bmatrix}
& \begin{bmatrix}
\Sigma \tilde{R}_{1,t} \\
\Sigma \tilde{R}_{2,t} \\
\Sigma \tilde{R}_{1,t} \\
\Sigma \tilde{R}_{2,t}
\end{bmatrix}
& \begin{bmatrix}
\sigma_1^2 \\
\sigma_2^2 \\
\sigma_1^2 \\
\sigma_2^2
\end{bmatrix}
& \begin{bmatrix}
\Sigma (\tilde{R}_{1,t})^2 / \sigma_1^2 \\
\Sigma (\tilde{R}_{1,t})^2 / \sigma_2^2 \\
\Sigma (\tilde{R}_{2,t})^2 / \sigma_1^2 \\
\Sigma (\tilde{R}_{2,t})^2 / \sigma_2^2
\end{bmatrix}
& \begin{bmatrix}
\tilde{R}_{1,0} - \frac{t}{T+1} \\
\tilde{R}_{2,0} - \frac{t}{T+1} \\
\tilde{R}_{1,t} \\
\tilde{R}_{2,t}
\end{bmatrix}
& \begin{bmatrix}
\sigma_1^2 \\
\sigma_2^2 \\
\sigma_1^2 \\
\sigma_2^2
\end{bmatrix}
& \begin{bmatrix}
\Sigma \tilde{R}_{1,t} \\
\Sigma \tilde{R}_{2,t} \\
\Sigma \tilde{R}_{1,t} \\
\Sigma \tilde{R}_{2,t}
\end{bmatrix}
& \begin{bmatrix}
\sigma_1^2 \\
\sigma_2^2 \\
\sigma_1^2 \\
\sigma_2^2
\end{bmatrix}
\end{bmatrix}^{-1}.
\]

More generally, for larger samples of events, this matrix is an \((N+1)\times(N+1)\) matrix in which only \(3N + 1\) elements are non-zero and \(N^2 - N\) are identically equal to zero. The non-zero elements as a fraction of total elements is \((3N+1)/(N^2+2N+1)\). As \(N\) increases, the number of non-zero elements dominated by the number of zero elements, and had we not used the
differences from means transformation, there would be even more zero elements relative to the non-zero elements in the relevant matrix. Once N becomes sufficiently large, the so-called sparse matrix problem must be addressed. Computation can become inaccurate due to rounding errors, despite the high accuracy of computers.

Note that the above inverse matrix can be partitioned into the form

\[
A^{-1} = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}^{-1} = \begin{bmatrix}
F_1 & -F_1 A_{12} A_{22}^{-1} \\
-A_{22}^{-1} A_{21} F_1 & A_{22}^{-1} \left(I + A_{21} F_1 A_{12} A_{22}^{-1}\right)
\end{bmatrix}
\]

where \(F_1 \equiv [A_{11} - A_{12} A_{22}^{-1} A_{21}]^{-1}\).

Only \(A_{11}\), the upper left partition of \(A^{-1}\), is sparse. Since the other partitions are not sparse, we no longer have to invert a matrix dominated by zeroes.

These matrix manipulations suggest a complexity which really is more apparent than real in implementation. We used the following procedures to implement our GLS model (here we present the case of a common event effect).

Our SAS code estimates the weighted average abnormal returns avoiding the direct inversion of the \(X'\Omega^{-1}X\) matrix. In order to execute this code the user needs:

1) to fill in the spaces we commented out in the code;
2) to create three SAS data sets with the following structures:

I) “panel” data set

II) “guess” data set: this data set contains the initial guess for the numerical procedure. A natural initial guess is the parameter estimates of the standard (less efficient) procedure.
where \( rmse_i \) is the root mean square error of the regression \( R_i = \alpha_i + \beta_i R_i^m \) for \( i = 1, 2, ..., N \)

III) “initial_value” data set: containing your original conjecture for the common abnormal return, called “f” here. A natural initial conjecture would be the estimate from the standard (less efficient) procedure.

Once steps 1 and 2 are completed, the user submits the appended SAS code and SAS prints the weighted average abnormal return (parameter “f” in the code) for each iteration of the numerical procedure.

The SAS code is

```sas
proc iml;
use panel;
read all var{ret avabret mktret1 mktret2 /*...*/ mktretN}; /*N is the number of firms/events*/
use guess;
read all var{_rmse_ intercept mktret};
use initial_value;
read all var{f};
N=nrow(_rmse_);
T=nrow(mktret1)/N-1;
y=ret;
w=avabret;
```
dummies=I(N) @ J(T+1,1,1);

z=mktret1 || mktret2 || /*...*/ mktretN || dummies; /*N is the number of firms/events*/
sqsigma=_rmse_##2;
a=intercept;
b=mktret;
pi=b/a;
theta=pi//f;
u=y-z*pi-w*f;

do i=1 to 10; /* users may change the maximum number of
iterations according to their numerical requirements */
isqsigma=sqsigma##(-1);
iv=isqsigma@j(T+1,1,1);
s1=z*(u#iv);
s2=w*(u#iv);
h22=w*(w#iv);
h21=w*(z#iv);
h12=h21';
h11=z*(z#iv);

k11=inv(h11-h12*inv(h22)*h21);
k12=-k11*h12*inv(h22);
k22=inv(h22)+inv(h22)*h21*k11*h12*inv(h22);
deltapi=k11*(s1-h12*inv(h22)*s2);
pi=pi+deltapi;
f=f+inv(h22)*(s2+h21*deltapi);
sdf=k22**0.5;
tvalue=f/sdf;
u=y-z*pi-w*f;
sqsigma=(i(N)@j(1,T+1,1/(T+1)))*(u##2);

print f; /* f is the estimated average abnormal return*/
print sdf; /* sdf is the estimated standard deviation of the estimated f*/
print tvalue; /* tvalue is the t-value of the estimated f*/
end;
print T; /*T is the number of time periods in the estimation window*/
print N; /*N is the number of firms/events*/
quit;

quit;
Appendix B: The Power of the Tests Compared

Suppose $\delta_i = \delta + \varepsilon_i$, where $\varepsilon_i \sim \text{Normal}(0, \sigma_i^2)$ and independently distributed.

The Traditional Methodology:

$\overline{\delta} = \frac{\sum_{i=1}^{n} \delta_i}{n} = \frac{\sum_{i=1}^{n} (\delta + \varepsilon_i)}{n} = \delta + \frac{\sum_{i=1}^{n} \varepsilon_i}{n}$. Thus, $E[\overline{\delta}] = E\left[\delta + \frac{\sum_{i=1}^{n} \varepsilon_i}{n}\right] = \delta$ and

$\text{Var}[\overline{\delta}] = E[(\overline{\delta} - E[\overline{\delta}])^2] = E\left[\left(\frac{\sum_{i=1}^{n} \varepsilon_i}{n}\right)^2\right] = \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2}$. Therefore,

$\overline{\delta} \sim \text{Normal}\left(\delta, \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2}\right)$.

The Inverse Variance Weighted Methodology:

$\delta = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \delta_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} (\delta + \varepsilon_i)}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \delta + \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \varepsilon_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$. Thus,
Let us now introduce a version of Hölder's Inequality (See Casella & Berger, Statistical Inference p. 181)

\[ \sum_{i=1}^{n} |a_i b_i| \leq \left( \sum_{i=1}^{n} \left( \frac{a_i}{\sigma_i} \right)^p \right)^{1/p} \left( \sum_{i=1}^{n} \left( \frac{b_i}{\sigma_i} \right)^q \right)^{1/q}, \]

where \( \frac{1}{p} + \frac{1}{q} = 1 \). Applying to our case, let

\[ a_i = \sigma_i, \quad b_i = \frac{1}{\sigma_i}; \quad p = q = 2, \]

to obtain

\[ \var \left( \bar{\sigma} \right) = \left[ \sum_{i=1}^{n} \left( \sigma_i^{-1} \right)^{-1} \right]^{-1} = \frac{\sum_{i=1}^{n} \sigma_i^{-2}}{n^2} = \var \left( \bar{\delta} \right). \]
Thus, the variance of the inverse variance weighted average abnormal return is smaller than the corresponding simple average. As a consequence, a statistical test based on the inverse variance weighted methodology is more powerful than the corresponding traditional method, as is shown subsequently.

When testing \( H_0: \delta = 0 \) versus \( H_1: \delta > 0 \)
the power function of the traditional test is
\[
\bar{\beta}(\delta) = \text{Prob}\left(Z > c - \frac{\delta}{\text{Var}[\bar{\delta}]}\right),
\]
where \( Z \) is a standard normal random variable and \( c \), a constant, can be any positive number. When the size of the test is 5\%, for example, \( c = 1.65 \). Likewise, the power function of the weighted test is
\[
\bar{\beta}(\delta) = \text{Prob}\left(Z > c - \frac{\delta}{\text{Var}[\bar{\delta}]}\right)
\]
(see, for example, Casella & Berger, Statistical Inference p. 360). So
\[
\text{Var}[\bar{\delta}] \geq \text{Var}[\bar{\delta}^\ast] \Rightarrow \bar{\beta}(\delta) \leq \bar{\beta}(\delta), \ \forall \delta > 0.
\]
Therefore, the inverse variance weighted test is more likely to reject the null hypothesis when it is false than the traditional test, that is, the weighted \( \sigma_i^{-2} \) test is more powerful.
REFERENCES


