May 1, 2003

More Power To You: A Stronger Event Study Methodology Illustrated using Brazilian Privatization Auctions

Abstract

Event studies of stock returns have been used extensively in the economics and finance literature, often concluding that abnormal returns are not significantly different from zero. We show that under the maintained hypotheses embodied in the standard pre-event estimation window that the traditional event study methodology is virtually never a *Best Linear Unbiased Estimator* [BLUE]. We introduce a more powerful methodology for event studies that is based on a single regression pooling all pre-event and event data. This methodology is BLUE under the maintained hypotheses. The efficiency gain and power improvement ultimately come from Generalized Least Squares "inverse variance weighting" of firm/event specific abnormal returns.

We show that a potential drawback of our method emerges when inverting the $X'\Omega X$ matrix which becomes sparse (dominated by zeros) as the number of firm/events increases. If not accounted for, this may introduce instability in the estimates. We demonstrate a convenient partition of the sparse matrix and a numerical algorithm to overcome this difficulty, if it arises.

Brazilian privatization auctions are used to illustrate the features and superior performance of the proposed methodology. In particular, our results show that the significance level of positive estimated average abnormal returns increases substantially when the new methodology is applied. It also shows the usefulness of the proposed algorithm for avoiding inversion of sparse matrices, a problem that arose in our estimation.

In light of the enhanced statistical properties of the new methodology, there is at least a *prima facie* case for reexamining standard results; insignificance of abnormal returns in earlier event studies using the traditional method may simply be an artifact of a less efficient methodology.

I. Introduction

Event studies of stock returns have been used extensively in the economics and finance literature, often concluding that abnormal returns are not significantly different from zero. It is our contention that the tests in the literature are inefficient, using classical terms, they are not based on *Best Linear Unbiased Estimators* (not BLUE) except under highly restrictive conditions. We present a new more powerful unbiased methodology for such tests that is based on a single regression rather than the standard "two step" methodology. To implement this methodology one may need to deal with sparse matrix problems (matrices dominated by zeros), as we explain below.

The standard event study methodology for a stock return based model follows a two-step (or three step in some cases) estimation procedure. First one regresses a time series of firm stock returns on market returns during a pre-event "estimation window" in order to establish a prediction equation for "normal" movements in the stock. This is performed for each firm which experiences an event. Then for each event, one finds the "abnormal return" over an "event window" by taking the difference between actual firm returns and the estimated firm returns for this time period as predicted by the first stage regression. Once this has been done, one can evaluate the hypothesis that abnormal returns from the events analyzed are, for example, positive. This is done by treating the abnormal returns as a random variable and testing significance given a null hypothesis of zero abnormal returns. Often this is followed by a third stage in which one looks for the determinants of abnormal returns by treating the abnormal returns as a dependent variable in a regression which includes firm or event specific characteristics as independent variables.

What we demonstrate herein is that for the hypotheses implicit in these two types of tests, the standard tests are almost never efficient and are potentially highly inefficient. This inefficiency causes the tests to be less powerful, potentially failing to reject the null hypothesis of no event effect even when the event would be significant using a BLUE Estimator. We present an alternative more powerful test and demonstrate the enhanced power of this test using an event study of Brazilian privatization auctions.

II. More powerful techniques

The econometric point of this paper can be seen by examining the assumptions employed in traditional event study research. First, there is a pre-event window over which one estimates the returns of an individual stock as a function of the returns of the stock market as a whole. This procedure is followed with separate regressions for each event, such that there are event specific intercepts, slopes *and variances*. Next one assumes that there is a common event effect, e.g., that abnormal returns are a constant $\delta>0$ for each event. One can test whether the returns in the event windows are positive by finding each event's actual returns net of its predicted event returns (using the pre-event regressions extrapolated to the event time period for this prediction) and testing whether the mean of this variable is positive. Implicitly the econometric assumption is that the event effect is a common δ across all such events, and that for the i^{th} event the estimated abnormal return, $\hat{\delta}_i$, is an estimate of the common δ. If one accepts these two propositions as the maintained hypotheses of the estimation, then the normal event study methodology is virtually never BLUE - it is inefficient except under highly restrictive conditions.

The intuition for why the standard methodology is inefficient and lacks power is really quite simple. Suppose that we feel that the abnormal return across events should be equal and positive and that we have estimates of this return for two separate events. Let the abnormal return from one event be estimated as -2.00 and the abnormal return from the second event be 1.00. One can extract the variance on the predicted return from the estimation window model. Suppose one standard deviation for these estimates are 4.00 and 0.10 respectively. The abnormal return for the second event is ten standard deviations above zero while that from the first event is only a half of a standard deviation below zero, and indeed is less than a standard deviation below the positive abnormal return. A simple average of the two estimates would yield a negative value for an overall estimate of the common abnormal return, however the simple average ignores the fact that the second positive estimate is more precise (smaller variance) than the first negative estimate, and so should be given greater weight in the calculation.

Now consider the intuition behind Generalized Least Squares. In GLS one is essentially finding the inverse variance weighted average when considering the estimation of any single parameter. A standard GLS notation is $\beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ where Ω is the covariance matrix. If all of the off diagonal covariances are zero, then one has pure inverse variance weighting.

So, assuming that a two stage procedure is desired for computational simplicity, one might consider using the inverse variance weighted average of the estimated abnormal returns rather than the simple average in the second stage of the estimation.[1] In this case, if one were to think of this in regression terminology, one could think of a regression of the $\hat{\delta}_i$s *solely on a constant term*. In the GLS notation above, then β is the estimate of the common abnormal

[1] Consider the example above with point estimates of -2 and +1 and standard deviations of 4 and 0.10. The simple average is -0.5; the inverse variance weighted mean is +0.998 (the weights are 0.00062 and 0.99938).

returns parameter, δ. The X "matrix" is a vector of ones and the Ω^{-1} matrix is a diagonal matrix of inverse variances for each of the estimated $\hat{\delta}_i$s.[2] The simple average used in the traditional methodology would hence be efficient for this estimation if, and only if, $\sigma_i=\sigma_j$ for all i,j and $\sigma_{ij}=0$ for all $i\neq j$. The inverse variance weighted two stage approach is BLUE if and only if all covariances are zero. In effect the traditional methodology throws out information generated in the first stage regression, the variance of the estimated $\hat{\delta}_i$s. Unless all of these first stage variances are *identical* (which would happen in practice only with probability measure zero), the second stage estimation cannot be BLUE.[3]

One can estimate a BLUE model as we show below. We then compare the results from the traditional methodology (which assumes equal variances as well as zero covariances) and a two stage GLS model or inverse variance weighted mean model (which assumes zero covariances) with a benchmark BLUE Estimator.

What we demonstrate in our modeling below is that the BLUE Estimator in GLS is achieved using dummy variables and a single one stage panel regression combining all of the events' data into one regression covering both the pre-event and event window time periods for every event. 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. The two stage procedure computing the mean using inverse variance

---

[2] $(X'\Omega^{-1}X) = \sigma_1^{-2} + \sigma_2^{-2}$, $X'\Omega^{-1}y = \sigma_1^{-2}y_1 + \sigma_2^{-2}y_2$, so we arrive at the inverse variance weighted mean $\beta = (\sigma_1^{-2}y_1 + \sigma_2^{-2}y_2)\big/(\sigma_1^{-2} + \sigma_2^{-2})$.

[3] OLS is BLUE only if variance is a single σ^2. Often one does not have the power to assume a more general variance structure, but in the first stage estimation of an event study the same program which prints out the $\hat{\delta}_i$s prints out the σ_i for each event.

weights, which we suggest above, is identical to this BLUE Estimator *if* all covariances are zero. In our application, the inverse variance weighted mean is almost identical to the BLUE GLS Estimator, whereas the traditional simple average in the second stage gives far different results.

In principle the traditional method and/or the inverse variance weighted means method *could* yield results very similar to the BLUE Estimator we propose. In practice neither may do so. Accordingly, the simpler two stage procedures *may* provide imprecise estimates of an event effect. Typically when they diverge from providing sensible estimates, they will diverge in a fashion which leads to insignificance of the event effect. That is, if the two stage procedure leads to significance, it is likely that so will the BLUE Estimator we develop below. But if the two stage procedure leads to insignificance, it is not at all clear that this is the correct conclusion, which could be verified by using the BLUE GLS model.

As we show below, there is the potential for another form of inefficiency which arises if one uses our proposed methodology: *numerical instability.* There is a literature on how sparse matrices (matrices dominated by zeros) may lead to computational inaccuracies (instability) in standard matrix inversion routines. We explain the relevance to our methodology after presenting more details about our approach. We then show how to overcome the potential estimation instability associated with our proposed model.

A. Estimation in event studies

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

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

where: R_{it} is the market return to firm i at time (day) t,

R_{it}^{m} is the stock market index return for time t associated with the i^{th} firm's event.[4]

There are N independent firm/event specific regressions, each for a time series of length T. (For our applications, we have N=71 firms/events and T=250 days, roughly one year of stock trading days). The superscript m denotes the market return for the time periods associated with the estimation window.

Define the event effect for one period[5] for the i^{th} firm as

$$\hat{\delta}_i = R_{i0} - (\hat{\alpha}_i + \hat{\beta}_i R_{i0}^{m}) \qquad [2]$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the parameter estimates from equation [1]. Supposing that the $\hat{\delta}_i$s are considered to be independent estimates of a common δ, then one can test the hypothesis that δ is positive by looking at the standard event study hypothesis

$$\bar{\delta} > 0 \qquad [3]$$

This has a maintained hypothesis that the variance in $\hat{\delta}_i$s is due to "measurement error."[6]

---

[4] Time t=-5, for example, is five days before event i. Since different events happen at different times, the calendar date for t=-5 will vary across events.

[5] 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.

[6] In a classical statistical model one might have several independent parameter estimates all taken using a single measuring devise which has measurement errors. Then each measurement doesn't have its own variance, each is simply a noisy signal of the size of what is being estimated.

But, suppose that one researcher has a good measuring device and another researcher has a noisier device, and each researcher has 20 estimates of the random variable. One shouldn't throw out the estimates from the researcher with the noisier device, but one shouldn't treat his estimates as having the same weight as those from the researcher with the better device. This then becomes the logic for using GLS or an inverse variance weighted mean. In our case, a

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:

$$R_{it} = \sum_{j=1}^{N} D_{i|j}\,(\alpha_j + \beta_j R_{mt}^{j}) + \sum_{j=1}^{N} D_i^{t=0}\,\delta + \varepsilon_{it} \qquad [4]$$

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

$D_i^{t=0} = 1$ for event i when t=0 (the event window),

δ is the abnormal return which is assumed to be common across all events,

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

ε_{it} is an unsystematic error with the usual assumptions of zero mean, positive variance and covariances of zero.

Under the assumption that the model in equation [1] is BLUE *for each event* and that [2] gives the event specific estimate of the abnormal return, $\hat{\delta}_i$, and that the true abnormal return, δ, is common across events, then [4] is BLUE for δ, and the traditional simple average is not BLUE unless the variances of the estimated $\hat{\delta}_i$s are *identical* across events and all covariances are zero. This methodology simply adds one time period to the pre-event window along with a dummy variable, common to all events, which captures the abnormal returns associated with the event effect.

---

stock which tracks the market more closely will have more precise estimates of δ than one which does not do so.

Equation [4] can in principle be estimated using Maximum Likelihood techniques,[7] but there is a technical problem which can arise in doing so. To see this, consider the same model in matrix form. We illustrate the matrix form with only two events (N=2).

$$\begin{bmatrix} R_{1,-T} \\ \vdots \\ R_{1,-1} \\ R_{1,0} \\ R_{2,-T} \\ \vdots \\ R_{2,-1} \\ R_{2,0} \end{bmatrix} = \begin{bmatrix} 1 & R^m_{1,-T} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & R^m_{1,-1} & 0 & 0 & 0 \\ 1 & R^m_{1,0} & 0 & 0 & 1 \\ 0 & 0 & 1 & R^m_{2,-T} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & R^m_{2,-1} & 0 \\ 0 & 0 & 1 & R^m_{2,0} & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,-T} \\ \vdots \\ \varepsilon_{1,-1} \\ \varepsilon_{1,0} \\ \varepsilon_{2,-T} \\ \vdots \\ \varepsilon_{2,-1} \\ \varepsilon_{2,0} \end{bmatrix} \qquad [5]$$

For the case of N=2, this nests equations [1] and [2] along with the assumption of a common δ. In place of the $\bar{\delta}$ in equation [3], equation [5] directly estimates $\hat{\delta} = \bar{\delta}'$ where $\bar{\delta}'$ is the GLS variance-covariance weighted mean of the $\hat{\delta}_i$s as we shall show below.

This approach is identical to the maintained hypotheses in equations [1] - [3], but can be estimated in GLS to attain the BLUE Estimator. (Referring to a two stage procedure, finding an inverse variance weighted mean of δ will be identical to [5] if the covariances from [5] are all zero.)[8] Note, however, the construction of the data matrix. For simplicity, suppose T=10. Then

---

[7] One could use OLS as well, but then one is again throwing out information having to do with the possibility that different pre-event windows may imply different pre-event precision in the estimation.

[8] Note, suppose that [1] is BLUE ***for each specific event separately***, as it will be under our assumptions. This equation is estimated over time [-T, ..., -1]. Were there no event in time

there are eleven non-zero data elements in the first two columns and eleven elements are identically zero. Similarly for the third and fourth columns. Finally the fifth column has twenty zeros and only two non-zero elements.

Now consider N firms/events. There are now 2N+1 columns in the data matrix. Each of the columns 1 to 2N have T+1 non-zero elements and (N-1)(T+1) identically zero elements. 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. So the ratio of non-zero elements to total elements is $2N(T+1)/2N^2(T+1)$ or 1/N. For N=71, as in our application later, this means there are only about 1.5% of the elements which are non-zero. Note, the final column in the data matrix has N positive elements and NT total elements, so it too is dominated by zeros.[9]

When one has a matrix dominated by zeros, this is called a "sparse matrix." Sparse matrices present estimation problems. Computationally in conventional statistical programs when one is inverting the $X'X$ matrix, it is treating the zeros as floating point *approximations* of zero. This can create rounding errors relative to the case in which the matrix has elements which are identically equal to zero. The computational problems involved with such cases are covered in Thisted (1988).

0, then one would assume that one could estimate [1] for the time period [-T, ...,0] and the model under these assumptions would be BLUE for exactly the same reasons as the original model. Assume instead that there is a shift in time zero, captured by a dummy variable. There is only one observation of this for each event. By assuming a common shift/dummy δ across firm/events, the same properties which makes [1] BLUE for each individual event makes this new pooled regression BLUE if estimated using GLS.

[9] This has been an intuitive explanation. More precisely, if T_1 is the estimation window (250 in our case), T_2 is the event window (1 in our case) and n is the number of events (71) and k is the number of explanatory variables (1 in our base case and 7 later), then the percent zeros in the matrix to be inverted, $Z'\Omega^{-1}Z$, (where Zs are the X deviations from means, as in most panel models) can be expressed as $100 - 100((T_1 + T_2)n + kn)/((T_1 + T_2)n(n+k))$ which is 98.69% in our model with 7 explanatory variables.

We handle the sparse matrix problem by extending an approach pioneered by Mundlak (1961). He noted that the design matrix for a linear regression with dummies had a special structure, so that one could analytically do a partitioned inversion of the $X'X$ matrix. Following this technique we can get an analytic expression for the common coefficient, δ, (as well as the event specific parameters $\{\alpha_i, \beta_i\}$s. Chamberlain (1980) noted that one could use a similar technique in a maximum likelihood setting. If one were using Newton-Raphson (or something similar) to maximize the (log) likelihood, each iteration of the procedure has a structure similar to the linear case. One can analytically simplify the Newton-Raphson procedure to update the common parameters by inverting a matrix of only size (k×k), where k is the number of common parameters (k=1 in this model, but k=7 is an application below). One then can update the estimates of the individual effects one at a time as a function of the update to the common parameters. Iterating this process to convergence maximizes the log likelihood.

In our work, we develop a similar matrix inversion technique for the sparse matrix, one which we find to be computationally easier than that used in earlier papers following this literature. Our algorithm is addressed in Appendix A.

Again, there is a question raised about whether one needs to use the sparse matrix methods for any given data set. It is possible that for a given problem that a simple off-the-shelf maximum likelihood program [ML] 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 all of this, the GLS program with the sparse matrix method remains BLUE and efficient, deviations from this benchmark may or may not be serious, but one cannot know whether one is efficient unless one in fact calculates the GLS using the sparse matrix techniques.

In our implementation of this model using the Brazilian data, with an assumption of a common δ across events, we find that the sparse matrix techniques add little to the results. Even the two stage estimation (using an inverse variance weighted mean, rather than a simple average), approximates the efficient benchmark of our New GLS. But, if we assume that the event effects are firm/event specific, but are determined by a common function of exogenous variables, e.g. $\delta_i = \Phi X_i$ where Φ is common across events, then none of the alternative models match the benchmark case of the New GLS (both a one stage approach and sparse matrix methods are needed to get the efficient BLUE results). With respect to the methodological approach to estimation, one cannot know if one is deviating from the BLUE estimator without undertaking the more complex (but weakly[10] superior) New GLS procedure that we describe.

C. Alternative specification for event specific abnormal returns

Next consider what event studies often do after they test for the sign of δ. They may drop the assumption of a common δ with measurement errors leading to differences in the δ_is.

That is, they hypothesize that the "measurement errors" in the above model are in part explainable by exogenous observable event specific factors. They then regress abnormal returns on firm/event specific explanatory variables as in equation [6].

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

The X_i matrix includes the 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.

[10] Weak in the sense of mathematically $\geq$ rather than $>$ in a theoretical sense, but the likelihood of being identical (=) is of measure zero.

Following the intuition from the last section, in which we examined not only the BLUE GLS but compared this with the inverse variance weighted mean, we could think of the analogue of the inverse variance weighted mean as applied to the regression in [6]. This would be a variance weighted regression in place of [6]. (As we show for our data set, the variance weighted regression is not a close approximation to the BLUE GLS regression.)

As in the last section, given the assumptions of the system (that equations [1] are BLUE for each event, that [2] represents abnormal return and that [6] represents common parameters explaining abnormal returns) a GLS model will be BLUE, whereas the traditional estimation of [6] will not unless all variances are **identical** and covariances are zero.

Now, for this third step, we again for simplicity present an N=2 example with a simple version of [6], $\delta_i = \phi^0 + \phi^1 x_i + \epsilon_i$ with only one explanatory variable. (Note, we use superscripts on these parameters which are assumed to be common across events to avoid confusion with the event specific subscripted α and β terms.)

$$\begin{bmatrix} R_{1,-T} \\ \vdots \\ R_{1,-1} \\ R_{1,0} \\ R_{2,-T} \\ \vdots \\ R_{2,-1} \\ R_{2,0} \end{bmatrix} = \begin{bmatrix} 1 & R^m_{1,-T} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & R^m_{1,-1} & 0 & 0 & 0 & 0 \\ 1 & R^m_{1,0} & 0 & 0 & 1 & x_1 \\ 0 & 0 & 1 & R^m_{2,-T} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 1 & R^m_{2,-1} & 0 & 0 \\ 0 & 0 & 1 & R^m_{2,0} & 1 & x_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \phi^0 \\ \phi^1 \end{bmatrix} + \text{error term} \qquad [8]$$

This again captures the maintained hypothesis of the second set of tests, but is BLUE rather than inefficient. Again, for implementation, one may need to use the sparse matrix methods we introduced above. (We demonstrate that for our data the sparse matrix methods are important for estimation of this third step in the modeling.)

D. The intuition behind the power of our tests

Finally, before turning to an example, we should discuss the issue of power. We address this formally in Appendix B.

To keep the discussion simple, let us do this in the context of the first hypothesis of a uniform δ across events. Suppose that there is a single true δ>0, but there is measurement error so that we observe $\hat{\delta}_i = \delta^{true} + \varepsilon_i$. As the variance of ε_i goes to zero, obviously all observed $\hat{\delta}_i$s will be positive. As its variance becomes large, in the limit as the variance goes to infinity we expect to observe half the observed $\hat{\delta}_i$s being negative and a 0.50 probability of the mean of observed $\hat{\delta}_i$s being negative.

Now consider a sample of events composed of two subsamples, one for which the δs are measured with little error and another where they are measured with substantial error. Those with a substantial error, many of which would be estimated as negative, would be pooled with those with little error. The expected value of $\bar{\delta}$ is unchanged, but the precision of its estimate falls and the probability of finding a negative mean rises relative to the case in which there are only well measured δs. The greater the proportion of observations with substantial errors, the greater the tendency towards failing to reject the null of no significant difference from zero (regardless of whether the point estimate of $\bar{\delta}$ rises or falls). Note, however, the inverse variance weighted mean increases the probability of rejecting the null when it is false as compared to the traditional test; in short, it increases the power. The GLS estimator then improves upon the inverse variance weighted mean by not restricting covariances to be equal to zero.

III. Empirical Examples Demonstrating the Power of the Proposed Tests

Before the analysis, it is useful to provide a brief review of the literature on evaluating acquisitions via event studies, with an emphasis on privatization acquisitions. Then we turn to the analysis. The reader who is only interested in the methodological aspects of the estimation may skip the first three subsections of this section (skip A. - C.) which deal with the institutional aspects of our example.

To preview the analysis, using a two stage approach and inverse variance weighting for finding the mean of the abnormal returns appears to be quite a bit more powerful than the traditional method of using the unweighted mean. Moving to a GLS regression with and without sparse matrix methods adds little to the inverse variance weighted mean results. All three methods have virtually identical results in both parameter estimates and *t*-values. But, conclusions are strongly affected once one turns to modeling the determinants of the abnormal returns.

When modeling the determinants of abnormal returns, on the surface there is no “best fit” to the data between the traditional approach, an inverse variance weighted GLS two stage approach using a canned ML one stage model, or using GLS with sparse matrix inversion methods. We know however that the sparse matrix GLS regression is BLUE (due to the model assumptions) and efficient (due to the matrix inversion accuracy). This is hence the benchmark one must use as it captures the maintained hypotheses and is the most efficient for estimation.

The fact that all of the other three alternatives have results that are somewhat at variance from these *best* estimates suggests that both the modeling strategy and the sparse matrix estimation strategy are important for this problem, at least in our data set.

A. Event studies of acquisitions

There are three strands of literature concerning event studies surrounding acquisitions that are potentially relevant to our analysis. The first looks at private acquisitions or take-overs, a second studies private acquisitions across borders (some of the Brazilian firms in our data were purchased by foreign firms or consortia including foreign firms). The third involves studies of privatization of government enterprises, the type of acquisition that we observe in Brazil.

There are numerous private sector acquisition studies. Franks and Harris (1989) examine the effects of over 1,800 takeovers on shareholder wealth in the United Kingdom in the period 1955-1985. They show that around the announcement date, acquired "targets" gain 25 to 30 percent and acquiring bidders earn zero or modest gains. Jarrel et al. (1988) surveyed many event studies that measure the effects of unanticipated takeover events on stock prices, after correcting for overall market influence on security returns. They summarize by saying that "Acquirers (…) receive at best modest increases in their stock price, and the winners of bidding contests suffer stock-price declines as often as they do gains." Bühner (1991) examines 110 takeovers involving the 500 largest enterprises in the Federal Republic of Germany in the period 1973-1985. His evidence shows that the shareholders of acquiring firms make losses, on average, of 9.83 percent around the takeover. Acquisitions in the United Kingdom from 1977 to 1986 are the subject of study by Limmack (1991). The author uses three counterfactual models in order to evaluate abnormal returns and concludes that bidder firms do suffer wealth decreases. Ding (1999) analyzes acquisition events in an emerging market. Using data from Singapore, he cannot reject the hypothesis that acquiring shareholders make zero abnormal returns around the announcement date. In addition to being consistent across different countries, the evidence that acquirers' shareholders on average at best break even in takeovers seems to be uniform over time as Leeth and Borg (2000) show. They examine the impact of merger announcements in the period 1919-1930 in the United States. Despite the different regulatory and economic environment at that time, their findings also suggest the acquiring firm stockholders do not, on average, make a positive abnormal return on acquisitions.

All these findings reinforce what Jensen (1986) wrote about the distribution of wealth in takeovers: "it appears that bargaining power of target managers, coupled with competition among potential acquirers, grants much of the acquisition benefits to selling shareholders."

The second branch of the literature on this topic is about cross-border transactions. Corhay and Rad (2000) study the wealth effects of international acquisitions using a sample of foreign acquisitions by Dutch firms during the period 1990-96. Their finding is that cross-border acquisitions create wealth for the Dutch firms, especially for acquisitions in the United States, after controlling for variables such as the relative size of the target with respect to the acquirer's size and the relatedness of their industries. The evidence is weak, though. Examining shareholder wealth gains from domestic and foreign takeover announcements in the U.S. chemical and retail industries, Dewenter (1995) finds that foreigners pay more than domestic investors in hostile transactions, but pay less when there are rival bidders. Among her explanatory variables are exchange rates and taxes. Doukas and Travlos (1988) investigate the effect of international acquisitions on stock prices of U.S. bidding firms. They find evidence supporting the fact that firms expanding into new industries and geographic markets – especially less developed markets – experience larger abnormal returns.

The third branch of literature examines acquisitions of firms which governments are privatizing. There are three basic methods through which governments privatize their assets: fixed-price share sales; tenders or auctions; private placements. In fixed-price share sales, the government splits up the ownership of the company into many shares and sets the unit price of the shares. Interested parties submit the number of shares they are interested in buying at that price. Oversubscribed issues may be allocated pro-rata or by some other criterion. Once the distribution of the shares and payment are effective, the ownership transfer is complete. State-owned activities can also be privatized in competitive auctions where pre-qualified competitors place bids for the price (above some minimum price) and the quantity of shares they want to acquire. Requirements for pre-qualification and rules of the auction vary across countries and over time. A private placement scheme is one in which the government somehow reaches an agreement with a particular investor group on the terms of the sale.

According to Dewenter and Malatesta (1997), the design of a sale may also be a combination of these three basic methods and have other dimensions such as time. For example, all shares of a company may be sold at once or in pieces separated by months or years. Shares may be reserved for employees, managers, institutions, or foreign investors, or limits may be placed on the holdings of some category of investor. Shares may be sold at discounts or with concessionary financing to some investor groups. A government may retain a golden share giving it partial control over some firm decisions or create regulatory bodies through which it exercises further influence.

Dewenter and Malatesta (1997) provide an analysis of initial offer prices in privatizations of state-owned companies compared to initial prices in public offerings of private companies. They test the hypothesis that privatization IPOs are, on average, underpriced more than privately owned company IPOs. Although asymmetric information among the different agents involved in both types of IPOs is the most cited (and modeled) reason for the observed underpricing – which would be a signaling device about the true expected future returns – the sources of asymmetries in the privatization and the privately-owned cases differ. In the latter case, IPOs often involve young firms in relatively new industries, in the former, they commonly consist of old, large and well-known firms, often with aging technologies. Therefore, other things equal, privatization IPOs should be less underpriced. Nonetheless, a distinguishing feature of privatization is that the government can affect the firm value after the initial offer through its policy instruments (regulations). The commitment of the governments to some regulatory environment is exactly what some underpricing is assumed to signal. Potential explanations other than those relying on the maximizing IPO proceeds hypothesis (which is the underlying assumption in the asymmetric information models) have been offered for privatization IPOs' underpricing. Some examples are building domestic political support for a privatization program by promoting widespread direct shareholding among citizens, which would also have the advantage of fostering the development of liquid domestic capital markets. Underpricing can be directed to benefit some specific group of people, such as the firms' employees, who might otherwise deter privatization transactions, or political allies.

For all that, it is not clear whether privatization by IPOs should be more or less underpriced than privately owned company IPOs. This is the empirical issue that Dewenter and Malatesta (1997) tried to answer by performing the test they proposed. Despite their efforts, they did not find any "general tendency for government officials to underprice IPOs to a greater degree than their private company counterparts." Their sample included privatization programs in Canada, France, Hungary, Japan, Malaysia, Poland, Thailand and the United Kingdom. In fact, for Canada and Malaysia the evidence supported the opposite conclusion. Only for the U.K. has evidence been found in favor of the privatization excess underpricing hypothesis. The contribution of D-M also went beyond this somewhat inconclusive result and examined some potential cross-sectional determinants of underpricing in privatization programs that are conducted by the IPO-like method. They found evidence indicating that initial returns are significantly higher in relatively less developed capital markets and for privatized companies in regulated industries.

B. Brazilian privatization auctions[11]

In 1990 eighty of the 500 largest non-financial enterprises in the country belonged to the 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 political and economic instability, the primary phase of privatization auctions did not arrive until the mid 1990s, the period of our study.

Unlike many privatization auctions in other countries, the Brazilian auctions were first price sealed bid auctions 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.

Most of the privatization auctions occurred from 1995 to 2000. There are 71 acquirer-privatized pairs in our analysis of auctions involving at least two approved bidders.

C. The data

The data set contains stock prices and local market indices. The following transformations yield stock and market returns, which are the inputs to the empirical analysis:

[11] Much of the following is attributed to the Getulio Vargas Foundation study reported in Manzetti (1999), and is described in detail in da Graça (2002).

$R_{it} = \ln[p_{it}] - \ln[p_{i,t-1}]$, where R_{it} is firm i's return on day t derived from stock prices, $p_{i\tau}$, τ=t,t-1.

$R_t^m = \ln[p_t^m] - \ln[p_{t-1}^m]$, where R_t^m is the market index return on day t.

The estimation window is based on day end prices for the 250 stock trading days up to one week before the privatization auction in question. The event window is the day of the auction and its announced winner.

These are the data needed for the 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 the event effect as a function of characteristics of the event and the winners.

We examine the winners' nationalities, as determined by the location of the company's main headquarters. A dummy variable for nationality is created with Brazilian firms coded as zero, foreign firms as one.

Assuming that each auction has a premium of a given percent of the value of the acquired firm, that premium will get translated into the acquiring firm's stock price at a percentage based on the size of the acquiring firm relative to the size of the purchase. A relative size variable is based upon market value of the acquiring firm, the total size of the purchase, and its participation share in the consortium for the case where the firm is part of a bidding consortium. Hence the relative size is the product of the purchase value times the participation share of the acquiring firm, divided by its market value.[12]

We also include a measure of how closely related were the businesses of the acquiring and acquired firm. The relatedness variable is zero when the acquiring and acquired firms have

[12] Five auctions with buyer stocks with very low capitalization are excluded from the study.

matching DataStream Advance industry codes, otherwise this is set to one.

We want to analyze winning competitive bids, so we include a participation indicator. As we noted before, there were 71 auctions for which there were two or more approved bidders. If on the actual day of the auction only one bidder actually arrives, that bidder knows it has a winning bid at the reservation price. We assume that if the winning bid is equal to the government set reservation price, that this indicates that there was only one firm bidding. If the winning bid is the reservation bid, we define the participation dummy variable as zero, otherwise this is one.

Bloomberg, an online financial service that provides quotes and technical analysis of securities as well as company and industry information, is the data source for the location of the acquirers' headquarters. DataStream Advance provides industry classification, market value figures and exchange rates to convert all values to the Brazilian Real. The Rio de Janeiro Stock Exchange web site furnishes minimum and final prices for most of the auctions. Dow Jones Interactive and BNDES Annual Reports complement the series and also provide, along with Manzetti (1999), data on the shares acquired by each firm.

D. *First hypothesis: uniform event effects*

The standard methodology treats the 71 observed abnormal returns as a random variable, as if there were 71 independent observations of δ measured with noise One then finds the mean and calculates the *t*-value for the difference between this mean and the null hypothesis of zero abnormal returns. We contrast this with instead using a GLS inverse variance weighted mean and with use of a single equation model as in [5] with and without using the sparse matrix methodology. What we call the "ML" [maximum likelihood] model is estimation using the PROC MIXED procedure in SAS, letting the program invert the sparse matrices without any special sparse matrix procedures. What we call the "New GLS" model is the same model structure, but using the sparse matrix methodology presented above. The results are summarized in Table 1.

Table 1				
	σ^{-2} Weighted Mean	ML	New GLS	Traditional[13]
Abnormal Return	0.69%	0.70%	0.70%	0.62%
t-value for H_0: δ=0	3.54	3.54	3.58	1.77
p-value (one tailed)	0.0002	0.0002	0.0002	0.038

It is apparent that the "Traditional" model simply lacks power relative to the other estimates. Although the results are positive and significant,[14] they are somewhat different in absolute value and far less significant or precise than any of the three other models. All of the other three methods are virtually identical in results. With the new methods, the p-values are remarkably lower, having fallen by over 99% from the value for the traditional model. Note in particular the power of the σ^{-2} Weighted model. This simply uses a traditional two step methodology, but rather than evaluating the mean of the abnormal returns random variables from

[13] In a text on event study methodologies, Campbell et al (1997), another methodology is suggested. They propose weighting the observations by the inverse of their standard deviations as one test. Their claim is "If the true abnormal return is constant across securities then..." use inverse standard error weighting. But "... if the true abnormal return is larger for securities with higher variance, then the better choice..." is the simple average. Performing the test in this fashion provides the intermediate result of t=3.16, p=0.008.

Positing these reasons for using a different weighting is ad hoc, and must be based on a structural change in the data. As we point out (see note 8), only GLS (inverse variance weighting) can be BLUE under the maintained hypotheses.

[14] Most hypotheses are either that there are returns to acquiring firms or that due to competition or hard selling, that there are zero excess profits. If one felt that most acquisitions were for "empire building," e.g., growth maximization (as a few early studies claimed), then two tails would be appropriate and the Traditional Model would lead to a p value worse than 5%.

the first stage, it instead looks at the inverse variance weighted mean as should be done for efficiency if the variances are available (as they are in every event study).

Indeed, for the test of the first hypothesis, use of the single equation model as presented above does not add much to the magnitude and precision of the inverse variance weighting, regardless of whether we use sparse matrix techniques or let the program invert the entire matrix. We shall see below, however, that this is no longer the case for the second hypothesis. Before moving on, it is useful to consider what differs between the weighted average model and the ML (or New GLS) models. They are statistically virtually identical but for one property. The weighted average model imposes $\mathbf{cov}\{\boldsymbol{\varepsilon}_{i0}, \boldsymbol{\varepsilon}_{j0}\} = \mathbf{0}$ for all i,j, whereas the ML and GLS allow for non-zero covariances (even though the maintained hypothesis is that the covariances are zero). If these covariances are not zero in fact, the ML and New GLS models will compensate for this. At least for these data, the new methodologies, even the simplest of them, are far more powerful.[15] Another way of seeing this is as follows. Two stage modeling estimates the parameters of the problem sequentially, the later stage cannot affect the previous stage. The New GLS estimates all parameters simultaneously, however, which means that all parameter estimates are dependent on the others and this may cause some residual correlation.[16]

Before moving on, it is worth noting one point. The New GLS model is BLUE. It may well be that covariances do not substantially affect any event estimates, that a two stage GLS

---

[15] In principle one could find that the observations with low variance have lower values for abnormal returns than the observations with high variances. In this case the new methodologies would lead to lower *t*-values. It should be noted, however, if the maintained hypotheses are as stated, then these lower values are a more accurate description of the correct statistical significance, the traditional method would be overstating the power of the hypothesis test.

[16] If two events occurred on the same day they would share identical market return data in the pre-event and event windows. It is likely that covariances would be affected in this case.

inverse variance weighted mean is almost always almost as good as our New GLS procedure, which should be the benchmark. But this is hard to know *a priori*. The only way to find out in any specific study is to run the New GLS model, which is (at least weakly) superior to any two stage model. This being said there is, however, an important observation. The simple mean traditional model is more likely to fail to reject the null of no event effect as long as the regression fits in the pre-event windows are not all of equal precision (equal variances). The two stage inverse variance weighted mean approach does not have this tendency towards accepting the null. One suspects, but cannot prove, that if the two stage inverse variance weighted mean model strongly rejects the null that this will be robust to the New GLS BLUE estimation.

The gist of this is, what are we to make of 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 reject the null using simple means, one expects that their rejection of the null will be even stronger with GLS techniques (either the inverse variance weighted mean, which can be calculated from regressions which have already been run, or the "New GLS" method).

E. Second hypothesis: abnormal returns explained by exogenous data

Here we present the analogous four models: a two stage model using inverse variance weighted regression; the ML model; the New GLS model using sparse matrix methods; and the traditional two stage model for comparison. The results are in Table 2.

Before presenting the table, two points should be emphasized. Two of the variables we include in these tests have interpretations suggested by economic theory (relative size and participation), the rest are controls (e.g., one might suppose, from the earlier literature, that whether an acquisition was by a foreign firm might play some role). Accordingly we have *a* *priori* sign expectations on these variables.

The inclusion of the "Relative Size" variable is supported by economic theory. If a larger company makes an acquisition of a given value, its stock abnormal returns should be lower than would be a smaller company's stock abnormal return since the value of the acquisition is small relative to the *ex ante* value of the company. To get some idea of the magnitudes involved, suppose the following. Define the true worth of the acquired firm as W and the auction price as P. Suppose that the size of the acquiring firm is M. Then the abnormal return as in proportion to firm value is $\frac{W-P}{M}$. Consider a model like $\frac{W-P}{M} = \alpha + \beta\left(\frac{P}{M}\right) + \gamma X$ as reflecting the construction of the abnormal return in the model. To gain some intuitive insight, suppose α=0 and γ=0. Then $\beta = \frac{W-P}{P}$, e.g., β is the amount of excess return relative to the price paid. E.g., an excess return of one percent would imply β=0.01.

The other variable supported by theory is participation. One would prepare a bid conditional upon the number of bidders. One might prepare bids for the case of two rivals or one rival, the latter being lower (at least for common value auctions). The participation dummy takes on the value zero if we infer that the bidding firm had no competition. *Ceteris paribus*, facing no rivals should increase the value of the auction to the bidders. This would suggest that the participation variable should have a negative effect on abnormal returns (they should be lessened if there is rivalry). But an equilibrium effect is also worth noting. If participants in an auction have a fixed cost of participation, a lower valued potential acquisition should attract fewer bidders. In this case, one might expect a zero effect from the participation variable. If the endogeneity of why firms participate in equilibrium is considered, there may be a weak negative effect, but certainly less of an effect than would be implied by the *ceteris paribus* case discussed above. One should expect a non-positive effect of this variable, given these theoretical considerations.

While the relative size and participation variables are supported by theory, the other exogenous variables are simply intuitive additions to the model. For example, “What if Nationality plays a role?” as suggested by the work on IPOs cited above.[17] Given the importance of relative size, we not only enter it linearly in the regressions, but also enter it interactively with the other exogenous variables.

Under the model assumptions, the New GLS model is both BLUE given the statistical assumptions and avoids any problems associated with inverting sparse matrices. Accordingly this model serves as the benchmark for evaluating the other models. Results are in Table 2.

First, note that the participation variable is not significantly different from zero in all regressions. If the equilibrium number of bidders is endogenous, this should not be surprising, so we drop any further consideration of this variable.

The GLS model has the Relative Size variable as highly significant and with the correct sign, but the relative size interaction terms are of opposite sign and the two significant terms are potentially offsetting the effects of the relative size term.

Now we can look at the differences between the models. Since we have made the case that the New GLS is BLUE and efficient, it is the “benchmark” model. Again, we want to compare the other models to this benchmark.

---

[17] Greater abnormal returns for foreign acquisitions are found by Corhay and Rad (2000) for acquisitions of Dutch firms; Deventer (1995) chemical and retail rival bidding results (but not hostile takeovers); Doukas and Travlos (1988) U.S. firms’ overseas acquisitions.

Table 2*				
	OLS: σ^{-2} Weights	ML	New GLS	Traditional Method
Intercept	-0.0203 (1.68)	0.00260 (0.83)	-0.00770 (1.21)	-0.02027 (2.12)
Nationality	0.03241 (2.53)	0.01049 (2.33)	0.02086 (2.31)	0.02659 (2.89)
Same Industry indicator	0.02527 (2.03)	0.00709 (1.74)	0.00710 (1.74)	0.01650 (1.82)
Participation: is bid above minimum?	-0.00418 (0.54)	0.00246 (0.83)	0.00240 (0.81)	0.00310 (0.52)
Relative Size	0.00498 (3.67)	0.00016 (0.55)	0.00141 (2.45)	0.00329 (2.89)
Relative Size * Nationality	-0.00340 (2.71)	-0.00104 (1.79)	-0.00207 (1.78)	-0.00275 (2.70)
Relative Size * Industry Indicator	-0.00349 (2.85)	-0.00146 (2.17)	-0.00145 (2.16)	-0.00301 (2.92)
Relative Size * Participation Indicator	-0.00146 (2.16)	-0.00041 (0.80)	-0.00041 (0.80)	-0.00036 (0.62)

* *t*-values in parentheses (based on differences from zero).

The results now are not much like what we had for the first hypotheses, tested in Table 1. For the first hypothesis the inverse variance weighted mean, ML, and New GLS models were similar in both parameter estimates and in *t*-values. Only the traditional model had a somewhat different parameter estimate and a much lower p-value on that estimate. Now each of the models have results which differ in parameter estimates from the New GLS model, but each looks to be a reasonably good fit (*t*-values) when taken alone and out of context.

For the estimation results presented in Table 2, the magnitudes, signs, and significance (*t*-statistics) are fairly comparable to those in the New-GLS estimation. Notable exceptions are the sign differences on “Participation” in the σ^{-2} weighted regressions model and the intercept term in the ML model. On first glance, the "Traditional" model seems to perform well relative to the New-GLS results. A Chi-square test of the null hypothesis that the parameter estimates in the New-GLS and traditional models are the same yields a value for the test statistic of 12.62 (~Chi-square 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 results are quite close to being statistically different at generally applied thresholds.[18]

How general this instability might be in practice is unclear, however. The data have significant multicollinearity. The correlation matrix reveals that for the 28 covariances in the data, 16 exceed 0.20 and 4 exceed 0.70! With high multicollinearity, there is presumed to be high covariances between the coefficients on the variables. The Traditional model and inverse variance weighted model both throw out the covariance information. The ML model does not throw out this information, but its matrix inversion routines may be even less stable with this multicollinearity (which was not a problem in the first model and is probably not a problem in many event studies) combined with the sparse matrix problem.

If, however, we look at the benchmark model and suppose that there is some reason to wish to have accuracy in estimation, it is clear that the other three approaches are not closely approximating the parameter values from this "New GLS" estimation with sparse matrix inversion methodology.[19]

---

[18] As is usual in Type-I – Type-II error models, we want a much high standard for rejecting the null if our hypothesis is that the null is incorrect. Here we are asking if the Traditional method replicates the BLUE GLS model, there is only a 0.15 probability that this is the case.

[19] It might we noted that 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. Still, however, if the size of the parameter estimates is important (not simply their *t* - values), it

Before concluding, it is worth noting the contrast between the common δ and the event specific δ cases. As we conjecture above, in the common δ model we know the effect from using the traditional model is to lead towards accepting the null of no event effect. There, if the inverse variance weighted mean strongly rejects the null, we can be fairly certain that the BLUE estimation using the New GLS model will do so as well. But for the firm/event specific model we can make no such claim. Any ϕ_i could be stronger or weaker in the comparison between any of the models and the Traditional model. All that we know is that the mean of the Traditional model $\hat{\delta}_i$s is less likely to be significantly different from zero than in the case of the inverse variance weighted mean and the New GLS models.

IV. Conclusions

Event studies in the past have typically assumed a model structure in which one first estimates individual event effects and then tests whether these effects have a common tendency. In some cases the common value examined is based on the assumption that the event effect is common across all events. In other cases the researcher may feel that the common effect is a function of explanatory variables, e.g., a greater event effect when some exogenous variable associated with an event is great.

What the traditional approach misses can be intuitively captured by noting that although the first stage in the traditional procedure provides an estimate of the stock value relationship to the market value for each event, it also provides an estimate on how good the fit is of each event's estimated pre-event sensitivity to these value changes. One of the first rules of estimation is, one should not throw out information which might be pertinent to estimation. As

performs poorly relative to the benchmark.

we demonstrate, the variance around the estimates from the pre-event window provide important information for estimation of the event abnormal returns. Indeed, if one assumes that covariances are all zero, if the maintained hypothesis is that the pre-event estimation is best linear unbiased [BLUE], and if the event effect is assumed to be common across events, then the BLUE estimator for the mean of event specific estimated abnormal returns is not the traditional simple mean but the inverse variance weighted mean. More generally, if covariances *may* be non-zero, a one step firm/event specific variance estimation is called for. This estimation may present technical problems due to sparse matrices (matrices dominated by zeros). We examine this intuition by presenting both the econometric structures for BLUE estimators with efficient matrix inversions and, for comparison, the traditional methods for estimation.

We look at two hypotheses which are common in the event literature. The first is that the abnormal returns from an event have a common value (e.g., is a parameter). The second is that the abnormal returns are a common function of exogenous event specific factors (e.g., the abnormal returns for event i can be thought of as explained by ϕX_i plus an error, as in a regression).

When looking at the first hypothesis (a common event effect) using the traditional approach, event effects that are estimated with precision are given the same weight as those estimated with huge errors. This is an inefficient estimation method when one has available the estimated variances for each event's effect. The GLS estimator, in such a case using a two stage process like this, is to use the inverse variance weighted mean of the event effects, not the simple mean. We show why the traditional simple mean leads towards failing to reject a null of no event effect and demonstrate with an event study of Brazilian privatization acquisitions that this loss of power can be substantial.

We further demonstrate that the maintained hypothesis of the traditional event studies need not be estimated in two stages, as is traditional. The maintained hypotheses (either the first or the second one) can be estimated in a single one stage regression pooling all of the pre-event and event data into one regression. This leads to a firm/event specific variance model which yields estimators that are BLUE. We also show that it leads to a sparse matrix estimation problem when the number of events is large. This can lead to computational problems in standard statistical packages. We demonstrate that this does lead to computational problems in our Brazilian acquisition data, which has 71 events, and provide a methodology for overcoming these problems to obtain BLUE estimators which are superior to traditional techniques, and strongly so in our data.

The gist of this has two prongs. Insignificance in earlier event studies using the Traditional method may simply be an artifact of an inefficient methodology which will tend to bias results towards accepting the null of "no event effect." This establishes a *prima facie* case for reexamining these results and for downgrading their relevance until having done so. The second is that parameter estimates explaining abnormal returns using the Traditional methodology may be suspect even if the model fits the data well. And one final note, as the number of events increases, the importance of using a sparse matrix methodology becomes more important.

Appendix A: Computations Given Sparse Matrix Problems

Recall equation [5], an example for only two events.

$$\begin{bmatrix} R_{1,-T} \\ \vdots \\ R_{1,-1} \\ R_{1,0} \\ R_{2,-T} \\ \vdots \\ R_{2,-1} \\ R_{2,0} \end{bmatrix} = \begin{bmatrix} 1 & R^m_{1,-T} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & R^m_{1,-1} & 0 & 0 & 0 \\ 1 & R^m_{1,0} & 0 & 0 & 1 \\ 0 & 0 & 1 & R^m_{2,-T} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & R^m_{2,-1} & 0 \\ 0 & 0 & 1 & R^m_{2,0} & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,-T} \\ \vdots \\ \varepsilon_{1,-1} \\ \varepsilon_{1,0} \\ \varepsilon_{2,-T} \\ \vdots \\ \varepsilon_{2,-1} \\ \varepsilon_{2,0} \end{bmatrix} \qquad [5]$$

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

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

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)×N)=(6×2) matrix with the first column being the column vector $[1\ 1\ 1\ 0\ 0\ 0]'$ and the second column is $[0\ 0\ 0\ 1\ 1\ 1]'$,

α is the column vector $[\alpha_1\ \alpha_2]'$,

$$\mathbf{R}^{\mathbf{m}} \equiv \begin{bmatrix} R^m_{1,-2} & R^m_{1,-1} & R^m_{1,0} & 0 & 0 & 0 \\ 0 & 0 & 0 & R^m_{2,-1} & R^m_{2,-1} & R^m_{2,0} \end{bmatrix}'$$, two columns by 6 rows,

β is the column vector $[\beta_1\,\beta_2]'$,

J is a column vector of time dummies $[0\,0\,1\,0\,0\,1]'$, coded 1 for an event,

δ is the abnormal returns parameter, assumed to be identical across events,

ε is the error term column vector, for which

$$E(\varepsilon\varepsilon') \equiv V = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_2^2 \end{bmatrix}$$

It will also be useful to define

$$\mathbf{R}^{\mathbf{m}}_{\mathbf{1}'} \equiv \begin{bmatrix} R^m_{1,-2} & R^m_{1,-1} & R^m_{1,0} \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{R}^{\mathbf{m}}_{\mathbf{2}'} \equiv \begin{bmatrix} 0 & 0 & 0 \\ R^m_{2,-1} & R^m_{2,-1} & R^m_{2,0} \end{bmatrix}$$

Now consider the following transformation

$$\mathbf{M_d R} = \mathbf{M_d D}\alpha + \mathbf{M_d R^m}\beta + \mathbf{M_d J}\delta + \mathbf{M_d}\varepsilon$$

$M_d \equiv \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix}$ where M is a 3×3 matrix (as are the matrices of zeros).

The elements of the $\mathbf{M}$ matrix are set to scale each variable as the difference from its mean value (e.g., $\mathbf{M} = \mathbf{I} - (\mathrm{ii}')/3$.

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

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

The more simple notation reveals in standard form how to obtain a GLS estimator for the parameters as $\hat{\gamma} = \left[\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\right]^{-1}\left[\mathbf{Z}'\mathbf{V}^{-1}\widetilde{\mathbf{R}}\right]$.

Returning to the more complex notation, this estimator is

$$\begin{bmatrix}\hat{\beta}\\ \hat{\delta}\end{bmatrix} = \left[\begin{bmatrix}\mathbf{R^m}'\\ \mathbf{J}'\end{bmatrix}\mathbf{M_d V^{-1} M_d}[\mathbf{R^m}'\ \mathbf{J}]\right]^{-1}\begin{bmatrix}\mathbf{R^m}'\\ \mathbf{J}'\end{bmatrix}\mathbf{M_d V^{-1} M_d R}$$

Defining $\widetilde{\mathbf{R}}^{\mathbf{m}} \equiv \mathbf{M_d R_m}$ and $\widetilde{\mathbf{J}} \equiv \mathbf{M_d J}$ then

$$\begin{bmatrix}\hat{\beta}\\ \hat{\delta}\end{bmatrix} = \left[\begin{bmatrix}\widetilde{\mathbf{R}}^{\mathbf{m}\prime}\\ \widetilde{\mathbf{J}}'\end{bmatrix}\mathbf{V^{-1}}[\widetilde{\mathbf{R}}^{\mathbf{m}}\ \ \widetilde{\mathbf{J}}\ ]\right]^{-1}\begin{bmatrix}\widetilde{\mathbf{R}}^{\mathbf{m}\prime}\\ \widetilde{\mathbf{J}}'\end{bmatrix}\mathbf{V^{-1}}\widetilde{\mathbf{R}}$$

$$\begin{bmatrix}\hat{\beta}\\ \hat{\delta}\end{bmatrix} = \left[\sum_{i=1}^{N}\begin{bmatrix}\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}\prime}\\ \tilde{\mathbf{J}}_{\mathbf{i}}^{\prime}\end{bmatrix}\frac{1}{\sigma_i^2}\,\mathbf{I}\,[\,\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}}\quad \tilde{\mathbf{J}}_{\mathbf{i}}\,]\right]^{-1}\left[\sum_{i=1}^{N}\begin{bmatrix}\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}\prime}\\ \tilde{\mathbf{J}}_{\mathbf{i}}^{\prime}\end{bmatrix}\frac{1}{\sigma_i^2}\,\mathbf{I}\,\tilde{\mathbf{R}}_{\mathbf{i}}\right] \text{ recalling } N=2 \text{ in our example.}$$

where

$$\tilde{\mathbf{R}}_{\mathbf{1}}^{\mathbf{m}\prime} \equiv \begin{bmatrix} R_{1,-2}^{m} - \bar{R}_1^{m} & R_{1,-1}^{m} - \bar{R}_1^{m} & R_{1,0}^{m} - \bar{R}_1^{m} \\ 0 & 0 & 0 \end{bmatrix},\ \tilde{\mathbf{R}}_{\mathbf{2}}^{\mathbf{m}\prime} \equiv \begin{bmatrix} 0 & 0 & 0 \\ R_{2,-1}^{m} - \bar{R}_2^{m} & R_{2,-1}^{m} - \bar{R}_2^{m} & R_{2,0}^{m} - \bar{R}_2^{m} \end{bmatrix}$$

so

$$\begin{bmatrix}\hat{\beta}\\ \hat{\delta}\end{bmatrix} = \left[\sum_{i=1}^{N}\begin{bmatrix}\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}\prime}\\ \tilde{\mathbf{J}}_{\mathbf{i}}^{\prime}\end{bmatrix}\frac{1}{\sigma_i^2}\,\mathbf{I}\,[\,\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}}\quad \tilde{\mathbf{J}}_{\mathbf{i}}\,]\right]^{-1}\left[\sum_{i=1}^{N}\begin{bmatrix}\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}\prime}\\ \tilde{\mathbf{J}}_{\mathbf{i}}^{\prime}\end{bmatrix}\frac{1}{\sigma_i^2}\,\mathbf{I}\,\tilde{\mathbf{R}}_{\mathbf{i}}\right]$$

$$\begin{bmatrix}\hat{\beta}\\ \hat{\delta}\end{bmatrix} = \left[\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\begin{bmatrix}\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}\prime}\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}} & \tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}\prime}\tilde{\mathbf{J}}_{\mathbf{i}}\\ \tilde{\mathbf{J}}_{\mathbf{i}}^{\prime}\,\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}} & \tilde{\mathbf{J}}_{\mathbf{i}}^{\prime}\tilde{\mathbf{J}}_{\mathbf{i}}\end{bmatrix}\right]^{-1}\left[\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\begin{bmatrix}\tilde{\mathbf{R}}_{\mathbf{i}}^{\mathbf{m}\prime}\,\tilde{\mathbf{R}}_{\mathbf{i}}\\ \tilde{\mathbf{J}}_{\mathbf{i}}^{\prime}\,\tilde{\mathbf{R}}_{\mathbf{i}}\end{bmatrix}\right]$$

where

$$\tilde{R}_i \equiv \left[R_{i,-2}-\bar{R}_i,\ R_{i,-1}-\bar{R}_i,\ R_{i,-1}-\bar{R}_i\right]^{\prime}$$

$$
\begin{bmatrix} \hat{\beta} \\ \hat{\delta} \end{bmatrix} =
\begin{bmatrix}
\dfrac{\sum\limits_t \left(\tilde{R}_{1,t}^{m}\right)^2}{\sigma_1^2} & 0 & \dfrac{\tilde{R}_{1,0}^{m} - \dfrac{\sum\limits_t \tilde{R}_{1,t}^{m}}{T+1}}{\sigma_1^2} \\
0 & \dfrac{\sum\limits_t \left(\tilde{R}_{2,t}^{m}\right)^2}{\sigma_2^2} & \dfrac{\tilde{R}_{2,0}^{m} - \dfrac{\sum\limits_t \tilde{R}_{2,t}^{m}}{T+1}}{\sigma_2^2} \\
\dfrac{\tilde{R}_{1,0}^{m} - \dfrac{\sum\limits_t \tilde{R}_{1,t}^{m}}{T+1}}{\sigma_1^2} & \dfrac{\tilde{R}_{2,0}^{m} - \dfrac{\sum\limits_t \tilde{R}_{2,t}^{m}}{T+1}}{\sigma_2^2} & \sum\limits_i \dfrac{1}{\sigma_i^2}
\end{bmatrix}^{-1}
\begin{bmatrix}
\dfrac{\sum\limits_t \tilde{R}_{1,t}^{m} \tilde{R}_{1,t}}{\sigma_1^2} \\
\dfrac{\sum\limits_t \tilde{R}_{2,t}^{m} \tilde{R}_{2,t}}{\sigma_1^2} \\
\sum\limits_i \dfrac{R_{i,0} - \dfrac{\sum\limits_t R_{i},t}{T+1}}{\sigma_i^2}
\end{bmatrix}
$$

(Recalling T=2 for our example.)

More generally, for larger samples of events, this matrix is (N+1)×(N+1) matrix in which only 3N+1 elements are non-zero and N^2-N are identically zero. Non-zero elements as a fraction of total elements is $(3N+1)/(N^2+2N+1)$. As N gets large, the non-zero elements are dominated by zeros. Once N gets to be sufficiently large, one must address the so-called sparse matrix problem. Computation can become inaccurate due to rounding errors, despite high accuracy of computers, this still can become a major problem, as noted in text.

Note that the above inverse matrix of the form $A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1}$ can be partitioned as

$$\begin{bmatrix} A_{11}^{-1}\left(I + A_{12}F_2A_{21}A_{11}^{-1}\right) & -A_{11}^{-1}A_{12}F_2 \\ -F_2A_{21}A_{11}^{-1} & F_2 \end{bmatrix} \text{ where } F_2 \equiv \left((A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1}.$$

The upper left partition is a diagonal matrix which is simple to invert using diagonal matrix techniques (which are not confounded by the dominance of zeros) Greene (2003). The other partitions are not sparse. The remaining inverse operations are not affected by the sparse matrix problem.

Appendix B: The Power of the Tests Compared

Suppose $\delta_1 = \delta + \varepsilon_i$, where $\varepsilon_i \sim$ *Normal* $(0; \sigma_i^2)$ *and independently distributed.*

<u>The Traditional Methodology</u>

$$\bar{\delta} = \frac{\sum_{i=1}^{n} \delta_i}{n} = \frac{\sum_{i=1}^{n} (\delta + \varepsilon_i)}{n} = \delta + \frac{\sum_{i=1}^{n} \varepsilon_i}{n}. \quad \text{Therefore: } E[\bar{\delta}] = E\left[\delta + \frac{\sum_{i=1}^{n} \varepsilon_i}{n}\right] = \delta \text{ and}$$

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

$$\bar{\delta} \sim \text{Normal}\left(\delta; \frac{\Sigma\, \sigma_i^2}{n^2}\right).$$

<u>The Inverse Variance Weighted Methodology</u>:

$$\bar{\bar{\delta}} \equiv \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \delta_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}(\delta + \varepsilon_i)}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \delta + \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\varepsilon_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

Therefore:

$$E\left[\bar{\bar{\delta}}\right] = E\left[\delta + \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\varepsilon_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}\right] = \delta \text{ and}$$

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

$$\bar{\bar{\delta}} \sim \mathrm{Normal}\left(\delta\,;\left(\sum_{i=1}^{n} \left(\sigma_i^2\right)^{-1}\right)^{-1}\right)$$

Let us now introduce a version of Holder's Inequality (See Casella & Berger, Statistical Inference p. 181):

$$\sum_{i=1}^{n} |a_i\, b_i| \le \left(\sum_{i=1}^{n} \left(a_i\right)^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} \left(b_i\right)^q\right)^{\frac{1}{q}},$$

where $\frac{1}{p} + \frac{1}{q} = 1$

Applying to our case, do $a_i = \sigma_i$: $b_i = \frac{1}{\sigma_i}$; $p = q = 2$, to obtain:

$$\mathrm{Var}\left[\bar{\bar{\delta}}\right] = \left(\sum_{i=1}^{n} \left(\sigma_i^2\right)^{-1}\right)^{-1} \le \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2} = \mathrm{Var}\left[\bar{\delta}\right]$$

Thus the variance of the σ_i^2 weighted average abnormal return is smaller than the corresponding simple average. As a consequence, a statistical test based on the σ_i^2 weighted methodology is more powerful than the corresponding traditional method as shown below:

We want to test H_0: $\delta = 0$ versus H_1:$\delta > 0$.

The power function of the traditional test is: $\bar{\beta}(\delta) = \text{Prob}\left(Z > c - \dfrac{\delta}{\text{Var}\left[\bar{\delta}\right]}\right)$, where

Z is a standard normal random variable and the constant c can be any positive number. For instance,
c = 1.65 when the size of the test is 5%.

Likewise, the power function of the σ_i^2 weighted test is: $\bar{\bar{\beta}}(\delta) = \text{Prob}\left(Z > c - \dfrac{\delta}{\text{Var}\left[\bar{\bar{\delta}}\right]}\right)$.

(See, for example, Casella & Berger, Statistical Inference p. 360).

Therefore $\text{Var}\left[\bar{\delta}\right] \geq \text{Var}\left[\bar{\bar{\delta}}\right] \Rightarrow \bar{\beta}(\delta) \leq \bar{\bar{\beta}}(\delta), \forall \delta > 0.$ This means that the σ_i^2 weighted test is more likely to reject the null hypothesis when it is false than the traditional test, i.e., it is more powerful.

REFERENCES

Bühner, R., "The Success of Mergers in Germany". *International Journal of Industrial Organization* 9:513-532, 1991.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, "The Econometrics of Financial Markets," Princeton University Press, Princeton, NY, 1997.

Casella, G., R. Berger, *Statistical Inference*, Duxbury Press, Belmont, CA, 1990.

Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* 47, Jan. 1980: 225-38.

Corhay, A., and A.T. Rad, "International Acquisitions and Shareholder Wealth: Evidence from the Netherlands," *International Review of Financial Analysis* 9: 163-174, 2000.

da Graça, T. B., "Stock Market Performance of Winners in Privatization Auctions: A New Event Study Methodology," *Doctoral Dissertation*, Cornell University, 2002.

Dewenter, K., "Does the market react differently to domestic and foreign takeover announcements? Evidence from the U.S. chemical and retail industries." *Journal of Financial Economics* 37:421-441,1995.

Dewenter, K., and P. Malatesta, "Public Offerings of State-Owned and Privately-Owned Enterprises: An International Comparison" *Journal of Finance* 52: 1659-1679, 1997.

Ding, D., "Acquisition Announcements and Shareholder Wealth: Evidence from an Emerging Market." *Journal of Emerging Markets* 4:5-26, 1999.

Doukas, J., and N. Travlos, "The Effect of Corporate Multinationalism on Shareholders' Wealth: Evidence from International Acquisitions," *Journal of Finance* 43:1161-1175, 1988.

Franks, J., and R. Harris, " Shareholder Wealth Effects of Corporate Takeovers: The U.K. Experience 1955-1985." Journal of Financial Economics 23:225-249, 1989.

Greene, W., *Econometric Analysis*, 5th edition, Prentice Hall, 2003.

Jarrell, G., J. Brickley, and J. Netter, "The Market for Corporate Control: The Empirical Evidence Since 1980." *Journal of Economic Perspectives* 2: 49-68, 1988.

Leeth, J. D., and J. R. Borg, "The Impact of Takeovers on Shareholder Wealth during the 1920s Merger Wave," *Journal of Financial and Quantitative Analysis* 35:217-38,2000.

Limmack, R. J., "Corporate Mergers and Shareholders Wealth Effects: 1977-1986." *Accounting* *and Business Research* 21:239-251,1991.

Manzetti, L., "Privatization South American Style," NY: Oxford University Press, 1999.

Mundlak, Y., "Empirical Production Function Free of Management Bias," *Journal of Farm Economics* 43: 44-56, 1961.

Thisted, Ronald A., *Elements of Statistical Computing*, Chapman and Hall, NY, 1988.