July 19, 2012

Bond Market Event Study Methodology

Louis Ederington
Wei Guan
Zongfei (Lisa) Yang

July 2012
Initial Draft: January 2012

Keywords: event studies, bonds, financial econometrics
JEL codes: G14, C1, G10

Corresponding author: Louis Ederington, lederington@ou.edu, Finance Division, Michael F. Price College of Business, University of Oklahoma, 205A Adams Hall, Norman, OK 73019, USA;
Wei Guan, wguan@mail.usf.edu, College of Business, University of South Florida St. Petersburg, 140 Seventh Avenue South, St. Petersburg, FL 33701, USA.
Lisa Yang, lisa1000yang@gmail.com, Finance Division, Michael F. Price College of Business, University of Oklahoma, 205A Adams Hall, Norman, OK 73019, USA.

Bond Market Event Study Methodology

Abstract

Extending the analysis in Bessembinder, Kahle, Maxwell, and Xu (2009), issues in the construction of corporate bond event studies using daily transaction or quote data are explored. We show that the procedures used in studies to date have relatively low power because they fail to control for the substantial heteroskedasticity in bond returns due to differences in term-to-maturity, rating, and other characteristics. Standardization of bond returns by their estimated volatility leads to much more powerful tests. We further find that due to infrequent trading, use of transaction price observations over several days before and after an event, while giving more weight to returns calculated from transactions closest to the event, yields considerably more powerful tests than returns based solely on transactions the day before and day after the event while also increasing coverage of small firms. We find that taking into account the effect on sample size, restricting the sample to institutional size trades does not result in more powerful tests. Nor does restricting the sample to inter-dealer trades.

Bond Market Event Study Methodology

1. Introduction

This paper explores issues in the construction of corporate bond event studies using daily or transactions data. At the urging of the SEC, in July 2002 the National Association of Securities Dealers(NASD) began reporting over-the-counter trades of corporate bonds through its Trade Reporting And Compliance Engine (TRACE). In February 2005, coverage was extended from a subset to almost the entire corporate bond market. Mergent FISD provides similar data for trades involving insurance companies while Datastream provides daily dealer bond quotes. The availability of these relatively new corporate bond databases raises the prospect of conducting the same sort of event studies for corporate debt that have become ubiquitous in equity markets. These could shed light on numerous issues. For instance, is a dividend increase good news for bondholders because it signals that management expects high earnings in the future or bad news because it lowers the cash available to service the debt? Is a new equity issue bad news for bondholders because it signals management thinks the firm is overvalued or good news because leverage is reduced? How does an acquisition impact creditors of the target and acquirer? We show that the test procedures used in bond event studies to date have relatively low statistical power to detect the bond price reaction to an event but that it is possible to construct much more powerful tests.

In the seminal paper in this area, Bessembinder, Kahle, Maxwell, and Xu (2009) (hereafter BKMX) show that among the databases they examine, the TRACE data yield the most powerful bond market event test statistics. For TRACE based event studies, they recommend: 1) eliminating trades of less than $100,000 par value, 2) calculating bond returns from average daily prices where each transaction price is weighted by trade size, 3) calculating abnormal returns using value-weighted rating/maturity benchmarks, and 4) combining returns on a firm's bonds into a single firm return. We extend their study.

In designing a bond market event study, a researcher faces hurdles which are either not present in the equity market or are less serious there. First, bond returns are characterized by considerable cross-sectional heteroskedasticity. Both between and within firms, bonds differ by time-to-maturity, rating and other characteristics. Prices of long-term bonds are much more volatile than prices of short-term notes, prices of low rated bonds are more variable than prices of high rated bonds, and prices of illiquid bonds are more volatile than those of liquid bonds. Also, like equity prices, bond prices are more volatile during periods of financial crisis. Since the t-test assumes bond returns are identically distributed, failure to control for this heteroskedasticity (as is the case in all extant studies to our knowledge) results in mis-specified t-tests. We explore how standardizing event window returns by a bond's estimated return standard deviation impacts the size and power of both parametric and non-parametric tests.

Second, bonds trade much less frequently than equities. On the average day in our January 2005-December 2011 TRACE sample, more industrial bonds do not trade than trade. Since a return calculation requires price observations on two days, the problem is magnified. Supposing an event occurs on day t, two day returns from day t-1 to day t+1 can only be calculated for 31.3% of the bond/days in our sample even after eliminating the least traded bonds. Since some firms have several bonds outstanding, the percentage of firm/days for which returns for the (t-1, t+1) event window are calculable is somewhat higher but is still only 38.6%. Also, these tend to be larger firms with more heavily traded bonds so restricting an event study sample to bonds and firms with calculable returns over the (t-1, t+1) window biases the sample away from smaller firms. Even when bonds trade, the number of trades is small. The median number of trades per day given at least one trade is three and the mean only 6.3 trades. This raises the question of whether it is advantageous to enlarge the event window, i.e., to utilize transaction prices on days t-2, t-3, etc., in addition to t-1, and t+2, t+3, etc., as well as day t+1, in calculating event window returns. Broadening the event window increases the number of observations but also increases the likelihood that forces other than the event of interest will cause bond prices to change so the net impact on test power is not clear a priori. If the event window is broadened, there is also a question of how best to weight price observations on different days since there is more non-event noise in observations further from the event.

Third, since firms may have multiple bonds, an issue is how best to combine the returns on the different bonds into a firm return. Because returns on a firm's long-term debt are more variable than returns on its short-term debt, this task is complicated by the just-discussed heteroskedasticity.

Fourth, many of the trades reported on TRACE are quite small. For instance, the median over our data period is only 25 bonds. Bid-ask spreads on these small bond transactions are often fairly large. For example, Edwards, Harris, and Pinowar (EHP) (2007) estimate the roundtrip spreads on trades of 20 bonds 1.62% versus spreads of 0.48% for trades of 200 bonds. Ederington and Guan (EG) (2011) estimate a spread of 2.25% on trades of 10 bonds and 0.55% on trades of 500 bonds. Together with infrequent trading, this means that return calculations based on small trades contain considerable bid-ask bounce. It is also possible that these represent trades by less informed investors. Accordingly, BKMX (2009) recommend restricting the sample to trades of $100,000 or more par value, i.e. 100 or more bonds. They show that this leads to more powerful tests when the number of observations is held constant. However, since 69.49% of bond trades are for less than 100 bonds, such a restriction sharply reduces the number of observations. As BKMX (2009) acknowledge, but do not explore, this reduction in the number of observations tends to reduce the power of the tests. We explore whether net restricting the sample to large trades increases or decreases the power of event studies considering the sample size effect. Fifth, along the same lines, since November 3, 2008 TRACE has indicated whether the trade was a sale to a non-dealer customer, a purchase from a customer or a trade between dealers. Since spreads are smaller on interdealer trades and these are supposedly informed traders, we examine whether test power is increased by restricting the sample to interdealer trades.

Our results include the following. One, the low t-test power observed by BKMX(2009) is due to violation of the homoskedasticity assumption. Abnormal bond return volatility varies considerably by maturity, rating, and time period. We find that standardizing each bond's abnormal event window return by its estimated volatility based on returns before and after the event leads to a substantial increase in t-test power. While the improvement in the power of the t-test is most dramatic, standardization also leads to a substantial increase in the power of the sign-rank test.

Two, due to infrequent trading, and despite the greater non-event noise over longer windows, we find using transactions on more than one day before and one day after an event to calculate event window returns, while giving greater weight to returns calculated from the transactions just before and the event, substantially increases the power of the tests. It also improves the representativeness of the sample. Part of the increase in test power is due to the increase in the number of firms with calculable event period returns but most is due to averaging out some of the noise in individual trade prices.

Three, restricting the sample to trades of 100 or more bonds has little net effect on the power of the tests. While the noise in abnormal bond returns is reduced, restricting the sample to large trades also considerably reduces the sample size, and the two effects approximately balance out in their impact on the power of the tests. Four, restricting the sample to interdealer trades reduces test power. Five, calculating average bond/day prices by weighting transaction prices by the square root of their trade size results in slightly more powerful tests than either a simple average or the trade size weighting utilized by BKMX (2009). Six, the sign-rank test is more powerful than the t-test or sign test.

As an example of how much our suggested test improvements impact test power, we find that for an event which shifts bond returns downward by 15 basis points, the likelihood of Type II error for a t-test based on a sample of 300 possible firm/event dates (including dates for which event window returns are un-calculable due to a lack of trades) falls from 80.10% for tests based on unstandardized abnormal returns over the (t-1, t+1) window to 36.13% for tests based on standardized returns. Together with a couple of other minor improvements, using transaction price observations over three days before and three days after the event to calculate event window returns (while giving transactions closest to the event day more weight) further reduces the likelihood of Type II error to only 15.47%. For the sign-rank test, the likelihood of Type II error falls from 52.29% for tests based on unstandardized (t-1,t+1) abnormal bond returns to 24.31% for tests based on standardized (t-1,t+1) return and to only 6.97% when the event window is expanded. Thus it is possible to improve considerably on the tests used in bond market event studies to date.

The remainder of the paper is organized as follows. In the next section, we describe our data and return calculations. In section 3, we explore the size and power of various test statistics for returns calculated over the (t-1, t+1) event window following extant procedures. In section 4, we develop our heteroskedasticity correction procedure and show how it improves the power of the test statistics. Issues of trade sampling, in particular whether it is better to restrict the sample to large institutional size trades or inter-dealer trades, are considered in section 5. In 6, we explore how individual trade prices are best weighted in calculating daily bond prices and how individual bond returns are best weighted in calculating firm returns. In section 7, we consider the consequences for the power of the tests of broadening the event window. Section 8 concludes.

Our analysis is conducted using bond transaction prices from TRACE. Our results concerning the need to correct for heteroskedasticity by standardizing returns in section 4 and how best to combine bond returns into firm returns in section 6.2 should apply equally well to dealer quote databases, such as Datastream. The results in sections 5, 6.1, and 7 are specific to transaction data sources, such as TRACE. While we do not specifically explore this issue, out test statistics should be robust to an event which increases volatility. We do not explore event clustering issues.

2. Data and measures

From TRACE we obtain bond trade data from January 1, 2005 to December 31, 2011. Since TRACE reports trades but few bond characteristics, attributes, such as coupon, maturity, and ratings, are obtained from Mergent's FISD database. Statistics on issuer total assets and sales are obtained from COMPUSTAT and LEXIS-NEXIS searches. BKMX(2009) restrict their sample to industrial,[1] non-convertible, non-putable, and non-zero coupon bonds. We impose the same requirements. In addition, we require that the bonds be denominated in US dollars, have $1000 par value, make semi-annual coupon payments, mature in 50 years or less, be rated by Moodys and/or S&P, and neither be in default nor have a tender offer outstanding. Since price changes on short maturity debt are small unless default is imminent, we only include bonds and notes with at least one year to maturity. 7296 bonds issued by 2206 firms meet these requirements.

A number of bonds in this sample are very rarely traded. To obtain a workable sample, we further require that the bond trade at least 100 times over the 2005-2011 period and that 2-day returns be calculable for at least 10 days over this period. Bonds are dropped from the sample when they default, are called or retired, or maturity drops below one year. Following BKMX we drop canceled, corrected, and commission trade observations. We also eliminate: 1) trades with settlement dates more than a week in the future, 2) "when issued" or "special price" trades, 3) trades with special sale conditions attached, 4) "as of" trades, or 5) trade prices less than $25 per $100 par value (which we regard as effectively in default). Finally, as a check, we compare the yield-to-maturity (YTM) reported on TRACE to the YTM calculated from the trade price reported on TRACE and coupon and maturity date from FISD. While generally the two YTM's are virtually identical, in 1.77% they differ by ten or more basis points. A difference indicates that: 1) either the TRACE price or YTM is incorrect, 2) the bond CUSIP is incorrect, 3) the FISD coupon

[1] FISD classifies as industrial some bonds issued by non-profit institutions, such as universities. However these account for less than 0.5% of the industrial bonds.

or maturity date is incorrect (in a few cases we observe trades after the FISD maturity date), or 4) there was a commission (which is included in the TRACE YTM but not the price) on the trade despite the fact that the TRACE data indicates no commission. We eliminate these trades leaving us with a final sample of 15, 247, 340 trades, 5507 bonds, and 1595 firms

We utilize two return calculations for bond n and event date t. One is the percentage return calculated as:

$$PR(t-1,t+1)_n = \frac{P_{n,t+1} - P_{n,t-1}}{P_{n,t-1}}. \tag{1}$$

where $P_{n,t}$ is the average price of bond n on day t.[2] In section 7, we consider longer event windows. In BKMX (2009), the non-parametric tests show a small tendency to find false evidence of negative events. Suspecting that this bias may be due to positive skewness in percentage returns, we also calculate log returns:

$$LR(t-1,t+1)_n = \ln(P_{n,t+1}) - \ln(P_{n,t-1}). \tag{2}$$

While TRACE and Datastream report what is known as the "clean" price, a bond purchaser pays the "dirty" price which is the clean price plus accrued interest since the last coupon payment.[3] Use varies but most researchers choose to calculate returns in terms of the dirty price. While the dirty price represents what bond traders pay and receive, since bonds trade plus accrued interest, the dirty price, but not the clean price, tends to fall by the amount of the coupon payment on the coupon payment date, thus returns calculated from the dirty price will tend to show a large negative return on the coupon payment date unless the coupon payment is added to equation 1

[2] Since bonds may trade several times a day, one issue considered by BKMX is how to calculate the daily price. They recommend an average of the prices on trades of 100 or more bonds weighted by trade size. We reexamine this issue below initially using an unweighted average of all trades. BKMX measure 1-day returns from t-1 to t. Since trades on event day t may occur before or after the event, we expect that most researchers will choose to test for a bond market reaction to an event on day t using 2-day returns from t-1 to t+1. This increases slightly the non-event noise vis-a-vis 1-day returns so is expected to reduce slightly the power of the tests versus the powers observed by BKMX.

[3] Accrued interest is negative if the settlement date falls between the ex-coupon date, the date after which the coupon payment is paid to the previous holder, and coupon payment date.

which most studies do not indicate that they do. Hence, we prefer and utilize returns calculated using the clean price.[4] Largely following the procedure in BKMX (2009), abnormal returns are calculated as:

$$ABR(t-1,t+1)_n = R(t-1,t+1)_n - BM(t-1,t+1)_n \qquad \textbf{(3)}$$

where $R(t\text{-}1,t\text{+}1)_n = PR(t\text{-}1,t\text{+}1)_n$ in the case of percentage returns or $R(t\text{-}1,t\text{+}1)_n = LR(t\text{-}1,t\text{+}1)_n$ in the case of log returns and $BM(t\text{-}1,t\text{+}1)_n$ is the mean return (percentage or log) on a benchmark rating/maturity matched portfolio corresponding to bond n. For this we utilize 24 benchmark portfolios – six rating classes (Aaa and Aa, A, Baa, Ba, B, and below B) and four maturity groupings (1 to 3 years, 3+ to 5 years, 5+ to 10 years, and over 10 years).[5] Moody's rating is used to assign bonds to portfolios if available. If S&P rates the bond and Moodys does not, S&P's rating is used. We require that at least five bonds in each rating/maturity group trade on days t-1 and t+1 in order to calculate the benchmark return for that group.

A final issue is how to calculate firm returns when a firm has more than one bond outstanding. A few studies (Narayanan and Shastri (1988), Marais et al. (1989), Hand et al. (1992), Warga and Welch (1993), and Cook and Easterwood (1994)) treat each bond as a separate observation. As BKMX observe, this biases the sample toward larger firms with many bonds outstanding. Moreover, returns for bonds of the same firm will be correlated leading to biased test statistics. Hence, we, like BKMX, combine a firm's various bond returns into a single firm return. For this, BKMX weighted each bond return by the number of bonds outstanding. Initially, we use an equally weighted average considering alternatives in section 6.

---

[4] The clean price will also tend to adjust on the coupon payment date if the coupon and market rates of interest differ but the adjustment will be smaller. One disadvantage of returns based on the clean price is that they are not full returns in that they exclude interest but these are minor over short event study intervals.

[5] These differ slightly from BKMX's benchmarks. They do not include bonds rated below B and separate Aaa and Aa rated bonds. Their three maturity classifications are 0 to 5 years, 5+ to 10 years, and over 10. Because of the scarcity of Aaa rated bonds, they split the Aaa category into two instead of three maturity groups.

Descriptive results are reported in Table 1. Abnormal percentage returns are slightly positively skewed while log returns are slightly negatively skewed. Kurtosis in both log and percentage and both bond and firm returns is very high exceeding that generally observed for stock returns (e.g., Brown and Warner, 1985). We address this kurtosis in section 4.

3. The size and power of extant bond event study test statistics.

To test the size and power of various test procedures and statistics, we employ essentially the same Monte Carlo procedures as in BKMX (2009), Brown and Warner (1980, 1985), Barber and Lyon (1997), and Lyon, Barker, and Tsai (1999), Boehmer et al (1991), and Kolari and Pynnönen (2010). First, to test the size (or Type 1 error probability) and bias of the tests, we choose firms and days at random and calculate how often the tests find evidence of an event when none in fact occurred. Specifically, we choose 300 firm/days at random and test whether the t-test, sign test, and sign-rank test applied to the abnormal 2-day returns find false evidence of an event. Repeating this procedure for 10,000 random samples, we record the percentage of the 10,000 samples in which the tests incorrectly find significant evidence of an event. Our procedure differs from BKMX's in that they choose 200 firm days at random from the sub-sample of firm days with calculable returns, i.e. trades on both days t-1 and t. Below, we consider event study methodologies which impact the number of days with observations, such as restricting the sample to large trades or interdealer trades. In order to consider the impact of these alternatives on test power, we choose 300 firm event days from the sample of all possible firm days whether or not trades occurred on both t-1 and t+1.[6] Hence, the number of return observations in each sample is considerably less than 300, averaging 116 but varying from sample to sample.

---

[6] For each bond, we regard all trading days between the first day the bond trades in our dataset and the last day it trades (or maturity falls below 1 year) inclusive as a possible bond event day. However if the bond traded during the first (last) week in our sample, we treat possible trading days as starting (ending) on the first (last) day in our sample. Possible firm event days include all days from when any bond for that firm first traded to the last trade of any bond.

Test size results for a two-tailed test at a 5% significance level are reported in Panel A of Table 2. A well-specified test is one in which the no-event null is falsely rejected 2.5% or less of the time in favor of finding evidence of a negative event and 2.5% or less of the time in favor of finding evidence of a positive event. As shown in Panel A, although some of the Type I error percentages slightly exceed 2.5%, all three tests are generally well-behaved in that none exceed 3%.[7]

Next we explore the power of various bond event study test statistics measured as the likelihood of detecting an event when one actually occurs. This may also be expressed as one minus the probability of Type II error. For this, we induce an artificial event in which returns from t-1 to t+1 are shifted upward by 15 basis points for a positive event or downward by 15 basis points for a negative event.[8] We again choose 300 firm event days at random from all possible firm days. Using two-tailed tests with a 5% significance level, we record whether each test correctly rejects the no-event null.[9] This is repeated for 10,000 random samples.

Results are reported in Panel B of Table 2. Like BKMX (2009), we find that the sign and sign-rank tests are considerably more powerful than the t-test. The t-test correctly rejects the no-event null less than 25% of the time for both positive and negative shocks and both log and percentage returns. In contrast the sign-rank and sign tests correctly reject the null more than twice as often. The sign test appears slightly more powerful than the sign-rank test. Observing little difference between percentage and log returns in either size or power, in the remainder of the paper, we present results only for log returns.

---

[7] In BKMX's sample, percentage returns were slightly but significantly biased toward finding false evidence of negative returns leading us to consider log returns which are less positively skewed. However, there is little evidence of a negative bias for percentage returns in our sample.

[8] BKMX shock investment grade bond (t-1, t) returns 15 basis points and speculative grade by 25 basis points. We consider varying the shock by rating and maturity below.

[9] While the tests are two-sided, the statistics in Table 4 are one-sided. In other words, for an artificial positive (negative) event, we report the percentage of times the test rejects the no-event null in favor of a positive (negative) event.

4. Abnormal return standardization

A likely candidate for the low t-test power observed in Table 2 is the clear violation of the homoskedasticity assumption. While the t-test assumes that all abnormal returns have the same variance, that is clearly not the case. As BKMX (2009) report, the abnormal return variance is higher for speculative grade than for investment grade bonds. Moreover prices are more variable on long-term bonds than short-term notes. In Table 3, we report abnormal return standard deviations for our 24 rating/maturity classes. Clearly, return volatility rises as the time-to-maturity lengthens and the rating falls. The standard deviation for bonds rated below B maturing in more than 10 years is more than seven times that for Aaa-Aa rated bonds maturing in 1 to 3 years. In addition, our data indicate that returns on illiquid bonds are more volatile and that volatility was higher in the financial crisis period, 2008 and early 2009. An obvious solution to this heteroskedasticity is to standardize each bond's returns by an estimated standard deviation for that bond's returns at that time and base the event study statistics on standardized returns.

4.1. Tests using standardized and unstandardized returns in equity event studies

Both standardized and unstandardized return measures are common in equity event studies. Consider the main non-portfolio procedures. In the non-standardized procedure, the t-statistic is calculated as $T1=AAR/\sigma_{AAR}$ where the average abnormal return $AAR=(1/N)\sum^{N}_{i=1} AR_i$, AR_i is the abnormal return around an event for stock i, and σ_{AAR} is a measure of the standard deviation of the mean return AAR. Two common measures of σ^2_{AAR} are: 1) the time series measure $(1/N) \sum^{N}_{i=1} \sigma^2_i$ where σ^2_i is a time series measure of the stock i's variance, and 2) the cross-sectional measure $((1/(N-1)) \sum^{N}_{i=1} (AR_i - AAR)^2$. Use of the latter (cross-sectional) variance measure has the advantage of allowing the event to impact volatility while the former assumes volatility is unaffected. Both assume independently and identically distributed AR_i. The t-tests in Table 3, like those in BKMX and most extant bond market event studies, follow this T1 procedure with a cross-sectional variance measure.

An alternative t-statistic, T2, suggested by Jaffe (1974), Mandelker (1974) and Patell (1976) is to standardize each AR_i by its estimated standard deviation σ_i yielding the standardized abnormal return, $SAR_i = AR_i/\sigma_i$. Patell (1976) then calculates the t-statistic as $T2 = (1/N)^{.5} \sum_{i=1}^{N} SAR_i$.[10] While T2 assumes that the event does not induce a change in the variance, Boehmer et al (1991) show that this is easily corrected by marrying the standardization technique of Patell (1976) with the cross-sectional standard deviation suggested by Charest (1978) and Penman (1982) and calculating the t statistic as $T3 = [(1/N)^{.5} \sum_{i=1}^{N} SAR_i]/SD\_SAR$ where SD_SAR is the cross-sectional standard deviation of the SAR_i. T2 and T3 assume independent,[11] but not identically distributed, AR_i.

Comparing the power of tests based on standardized and unstandardized returns, Brown and Warner (1985) find (their Table 8) that tests based on standardized returns are more powerful. However, after presenting both test statistics, Campbell, Lo, McKinley (1997) assert, "In most studies, the results are not likely to be sensitive to the choice of J_1 [our T1] versus J_2 [our T2] because the variance of the CAR is of a similar magnitude across securities" (p162). Similarly, according to Kothari and Warner (2006), "While a test using standardized abnormal returns is in principle superior under certain conditions, especially in short-horizon event studies, it typically makes little difference" (their footnote 5). While violations of the homoskedasticity assumption may be minor for equities as these quotes indicate, the evidence in Table 3 indicates it may be much more serious for bonds. Whether return standardization makes much difference to the power of bond market test statistics is the issue to which we now turn.

4.2. Bond return standardization

---

[10] When σ_i is estimated from market model residuals, this is adjusted as $T2= [(D-40)/(D-2)N]^{.5} \sum_{i=1}^{N} SAR_i$ where D is the number of time series observations used to estimate the market model.

[11] In this paper we do not address the lack of independence which would be caused by clustered events. Kolari and Pynnönen (2010) suggest a modification of T3 when events are clustered.

To estimate $\sigma_{n,t}$ for each bond n and possible event date t, we calculate the standard deviation of 2-day returns for bond n around day t using observed 2-day returns from t-25 to t+25 excluding returns across event day t.[12] For this, we require at least six return observations over this 50-day period which reduces the sample from 1,672,765 bond 2-day returns to 1,565,924.

We utilize two different standardization procedures. Standardized abnormal returns, $SABR(t-1,t+1)_n$ are calculated by dividing the abnormal return, $ABR(t-1,t+1)_n$, by the standard deviation of <u>abnormal</u> returns over the (t-25, t+25) period where abnormal returns are calculated as described in section 2. Abnormal standardized returns, $ABSR(t-1,t+1)_n$ are calculated by first calculating standardized raw returns, $SRR(t-1,t+1)_n$ by dividing each raw return by the standard deviation of <u>raw</u> returns over the (t-25,t+25) period, recalculating a standardized benchmark $SBM(t-1,t+1)_n$ using the standardized returns, $SRR(t-1,t+1)_n$ and then calculating the abnormal standardized return $ABSR(t-1,t+1)_n = SRR(t-1,t+1)_n - SBM(t-1,t+1)_n$. The latter procedure gives lower volatility bonds more weight in calculating benchmark returns. Since a series of trades at the same price yields a standard deviation at or close to zero which is unlikely to be repeated, we winsorize both the standard deviations and the standardized return measures at the 0.5% level within each rating/maturity category.

For both ABSR and SABR, we calculate standardized firm returns as equally weighted averages of the firm's *standardized* bond returns. This effectively gives a firm's least volatile bonds a greater weight in calculating firm returns than was the case for unstandardized abnormal returns. A firm's ABR returns will tend to be dominated by its most volatile bonds. Since each standardized return is weighted equally in calculating firm ABSR and SABR, the influence of least volatile notes and bonds on the firm return is effectively increased.

---

[12] Since 2-day returns overlap, they tend to be positively correlated. Thus, while consistent, the sample standard deviation is biased slightly downward in small samples. Also some 1-day returns are in the sample mean twice and others only once. These problems become more serious later when we lengthen the return window. Consequently, instead of using the sample variance calculated as $Var=[1/(T-1)]\sum_t (r_t-mr)^2$ where mr is the sample mean 2-day return, we calculate as the variance as $Var=[1/T]\sum_t r_t^2$.

Statistics for firm SABR and ABSR are reported in Panel C of Table 1. It is noteworthy that excess kurtosis is sharply reduced from 20.89 for ABR to 1.83 for SABR and 2.02 for ABSR. Clearly most of the excess kurtosis observed in non-standardized returns is due to heteroskedasticity. Note that while standard deviations of standardized 2-day *bond* returns are 1.00 by construction, standardized 2-day *firm* return standard deviations are <1.00 due to averaging out some of the noise in individual bond returns. This reinforces BKMX's argument that it is better to use firm returns rather than bond returns.

4.3. Size and power results

Size and power results are presented in Table 4 for standardized log returns. These are structured the same as in Table 2, i.e., 10,000 random samples of 300 firm/days each with 15 basis point shocks for the power tests. The t-statistic is calculated as in Boehmer et al(1991) (T3 in section 4.1). Again the significance level for all tests is 5%. Since the sample is reduced slightly from that in Table 2 due to our requirement for sufficient observations to calculate the standard deviations, we repeat the tests for unstandardized returns (ABR) for this sample for comparison. As reported in panel A, although the t and sign-rank tests show a slight tendency to find false evidence of a positive event too often, all three tests are reasonably well-behaved.

As reported in Panel B, the improvement in the power of the t-test when returns are standardized is dramatic. The likelihood of finding evidence of an actual positive or negative 15 basis point event approximately triples. The likelihood of detecting a negative event jumps from 19.89% for ABR to 66.01% for SABR and 63.87% for ABSR. Clearly the t-test's low power which we and BKMX observe for unstandardized returns is due to the violation of the homoskedasticity assumption.

Standardization also leads to a sizable increase in the power of the sign-rank test. The likelihood of detecting a negative event increases from 47.71% for unstandardized returns to 75.43% for SABR and 75.69% for ABSR. Note that this means that the likelihood of Type II error is cut in half. The results for a positive event are similar. Since standardized and unstandardized bond returns have the same sign, the power of the sign test is unchanged at the bond level. However, because standardization gives greater weight to less volatile bonds in calculating firm returns, some firm returns switch signs leading to a modest improvement in the power of the sign test as well. Nonetheless, while the sign test appeared the most powerful of the three tests for unstandardized returns, the sign-rank test clearly is more powerful for standardized returns.

Powers differ little between standardized abnormal returns, SABR, and abnormal standardized returns, ABSR. For the t-test, SABR's power appears slightly higher; for the non-parametric tests, ABSR's. While each measure has its advantages, we marginally prefer ABSR since it attaches more weight to less volatile bonds in calculating benchmark returns. Since ABSR and SABR results are roughly the same, henceforth we present results for ABSR only to save space.

In summary, while, to our knowledge, all bond event studies to date have utilized tests based on unstandardized returns, we find that bond return volatilities differ greatly so that standardizing each bond's event window returns by a return standard deviation estimate based on returns before and after the event leads to much more powerful t-tests and sign-rank tests and slightly more powerful sign tests. There is little difference in the size and power of the two standardized return measures, ABSR and SABR. While we have provided results for tests based on the TRACE database, results should be very similar for quote databases, such as Datastream, and also for monthly databases since volatilities should differ by time-to-maturity, rating, liquidity, etc regardless of the database.

In our power tests, we have tested the effect of a 15 basis point return shock. According to Table 3, 15 basis points represents .356 standard deviations for an Aa rated bond maturing in 1 to 3 years but only .048 standard deviations for a bond rated below B maturing in 10 or more years. Consequently, the likelihood of Type II error is much less for the former. This is illustrated in

Table 5, where we report powers of the sign-rank test based on abnormal standardized returns by rating/maturity category for across-the-board 15 basis point return shocks averaging powers for positive and negative return shocks. Since conducted by term-to-maturity and rating, these tests are conducted at the bond level, rather than the firm level as in Table 4.[13] For each maturity/rating category, we run 10,000 simulations. Clearly the likelihood of Type II error varies greatly by both maturity and rating from 0.0% for Aa rated bond maturing in less than 3 years to 87.71% (100%-12.29%) for bonds rated below B maturing in more than 10 years. Powers do not exceed 30% for bonds maturing in 10+ years regardless of their rating nor for bonds rated below B regardless of maturity. Results for the t-test are very similar.

This raises the issue whether it is realistic to model an event as impacting all bond prices equally. An event which causes a given shift in the yield-to-maturity has a much greater impact on prices (and therefore returns) of a bond maturing in 20 years than on prices of a note maturing in 2 years. As an alternative to across-the-board shocks, in the remainder of the paper we also present results for a proportional shock where the proportions are based on the standard deviations in Table 3. Specifically, $SHOCK_{r,m} = .0015[SD_{r,m}/.01568]$ where $SHOCK_{r,m}$ is the shock in basis points to bonds in rating category r and maturity category m, $SD_{r,m}$ is the standard deviation of abnormal returns of bonds of rating r and maturity m as reported in Table 3 and 1.568% is the standard deviation of all abnormal returns. So the shocks vary from a low of .0015(.00421/.01568) = 4.03 basis points for 1-3 year Aa rated bonds to a high of .0015(.03094/.01568) = 29.61 bp for 10-30 year bonds rated below B. Under this proportional event shock format, type II errors are roughly the same across all rating/maturity categories.

[13] Since bonds trade infrequently and firms may have multiple bonds, there are more missing (t-1, t+1) bond returns than firm returns. Also some ratings and maturities trade more than others. For comparison with the test powers in Table 2 and each other, sample sizes were adjusted so that average number of observed bond returns was the same for each maturity/rating category and equal to the firm/return average.

5. Trade Sampling

Next we turn attention to the issue of trade sampling.[14] To this point we have included all trades in calculating bond returns. Many trades reported on TRACE are quite small. In our sample 69.4% are for less than 100 bonds (< $100,000 par value) and 46.5% are for 25 bonds or less. It is clear that there is more bid-ask bounce in smaller trades. Edwards, Harris, and Pinowar (EHP) (2007) estimate the roundtrip spreads on trades of 20 bonds at 1.62% but only 0.55% for trades of 500 bonds. Ederington and Guan (2011) estimate a spread of 2.25% on trades of 10 bonds and 0.48% on 200 bond trades. Also, small investors may be less informed so deviations from equilibrium prices may be larger on small trades. Consistent with this, in our sample, the standard deviation of 2-day firm ABRs is 1.16% when calculated using only trades of 100 or more bonds versus 1.57% for all bonds. Reasoning that prices on institutional size trades are more reliable and finding evidence of higher test power, BKMX recommend calculating returns using only trades of 100 bonds or more. Since it consists solely of insurance company trades, the Mergent FISD transaction database effectively imposes a similar restriction.

While there is less noise in larger trades, restricting the sample to trades of 100 or more bonds eliminates 69.44% of trades in our sample. Hence, the number of firm days with calculable returns from t-1 to t+1 is reduced from 38.61% to 22.98%. BKMX (2009) acknowledge this sample size reduction but do not explore its implication for test power holding their sample size constant at 200 firm/days. To explore whether net the power of event studies is increased by restricting the sample to trades of 100+ bonds because the noise reduction outweighs the sample size reduction or whether it is reduced because the sample size reduction outweighs the noise reduction, we calculate bond and firm returns from t-1 to t+1 for all days t and bonds n, based on only trades of 100 or more bonds. Maturity/rating benchmarks for the abnormal return calculations are also recalculated based on 100+ trades only. For the size and power tests we

[14] This is not an issue for quote databases.

again choose 300 firm/days at random among all possible firm/days and repeat for 10,000 random samples.

Size results are reported in Panel A of Table 6 and power results in Panel B. Power results are reported both for an across-the-board 15 bp return shock and for shocks proportional to the relative standard deviations for that rating and maturity as described in the previous section. Results for all trades are repeated for easier comparison. As shown in Panel A, all three tests are reasonably well-behaved for both samples. As reported in Panel B, the t-test is slightly more powerful when the sample is restricted to trades of 100+ bonds, the sign test is slightly less powerful, and results for the sign- rank test are mixed. In other words it appears that for the t-test, the noise reduction effect slightly outweighs the sample size reduction effect and that this is reversed for the sign tests. However, for all three tests, the power differences are small.

As one would expect, restricting the sample to trades of 100 or more bonds tends to eliminate smaller firms from the sample. When the sample is restricted to trades of 100 bonds or more, median total assets are $17.9 billion. For the firm days in the broader sample, but not in the 100+ sample, median total assets are $11.3 billion.[15]

Since November 2008, TRACE reports whether the trade is a dealer sale to a customer, a dealer purchase from a customer, or an interdealer trade. Edwards, Harris, Piwowar (2007) and Ederington and Guan (2011) show that bid-ask spreads are considerably smaller in interdealer trades. Moreover, dealers may be more informed so deviations from equilibrium prices could be

[15] Determining issuer characteristics, such as total assets, is complicated by the fact that only about half were available on COMPUSTAT using the bond CUSIPs from TRACE. Some bonds are issued by firms where the equity is privately held so that they are not required to file statements with the SEC except when they issue public debt. Various firms, such as Dun and Bradstreet, collect or estimate total assets, for some but not all of these firms. We searched for these on Lexis-Nexis. Also, after mergers and acquisitions, the debt often continues to trade under the original name and CUSIP although it is now an obligation of the acquiring firm. Thus, we had to match bonds and firms using tickers, S&P Bond Guide and other sources. In addition, foreign firms may issue public debt in the US. Again, we searched for these on Nexis-Lexis. If possible, we collected total asset figures as of December 31, 2009. If only available for a different date, we adjusted to a 12/31/2009 basis using the producer price index. We were unable to obtain total asset figures for 5.21% of the sample (by firm/day).

less.[16] Hence, we expect less non-event noise in returns calculated from interdealer trades only and indeed the standard deviation of firm ABR returns based on only interdealer trades is 1.45% versus 1.57% for returns based on all trades. Of course since only 27.79% of the trades in our sample are interdealer trades, this (like the restriction to trades of 100+ bonds) sharply reduces the sample size so the net effect on test power is unclear giving rise to our next tests.

Size results and power results based on data from November 2008 (when TRACE started identifying interdealer spreads) through December 2011 are reported in Table 7. Results for tests using all trades for this same subperiod are shown for comparison. As shown in panel B, restricting the sample to interdealer trades reduces the power of the tests somewhat. This restriction also tends to bias the sample toward smaller firms.

6. Price and bond averages

6.1 Calculating average prices

In calculating bond prices each day, prices on different trades have to be combined into an average price for the day. Then in calculating returns on a firm's bonds over an event window, the returns on the various bonds must be converted to a firm average. Comparing the power of: 1) using only the last trade of the day, and 2) an average where each trade is weighted by trade size, BKMX find that the latter is more powerful so recommend it. To this point, we have followed their procedure, but, since their trade size weighting gives a trade of 1000 bonds a weight 100 times that of a trade of 10 bonds, we decided to explore whether an alternative weighting in which the difference in weights between different size trades is reduced might yield more powerful tests.

---

[16] It is commonly assumed in the microstructure literature that market makers are less informed than at least some traders so (abstracting from their bid-ask spreads) tend to lose on trades with more informed investors. While this may be the case with regard to firm specific information, given the opaque nature of the corporate bond market, dealers may well be more informed than most – especially regarding equilibrium price of similar bonds.

Accordingly, we compare test powers when the trades are weighted by trade size and the square root of trade size. We also consider equal weights.

Test powers are reported in Table 8. Since all three tests are reasonably well-behaved for all three weightings, we do not report size results. Since the power stats are not very different for the three price weightings, any of the three weighting schemes seems acceptable. Nonetheless, for both across-the-board and proportional positive and negative shocks, powers for all three test statistics are slightly higher for the square root of trade size weighting. Consequently, we utilize it in the remainder of the paper.

6.2. Averaging bond returns into firm returns

Because some firms have several bonds outstanding, another issue is how to combine the abnormal returns on the different bonds into a single firm return. A few bond market event studies have simply treated each bond as a separate event observation. By giving firms with many bonds multiple observations, this biases the sample toward larger firms and, because the returns on different bonds of the same firm are correlated, the usual test statistics are biased toward finding false evidence of an event. Also, as reported in Table 1, by averaging out some of the noise in individual bond returns, firm returns have a slightly lower standard deviation so should yield more powerful tests. Hence we, like BKMX (2009) recommend coalescing returns on different bonds of the same firm into a single firm return.

To this point, we have utilized an unweighted average of standardized bond returns for the firm return. BKMX (2009) use an average of unstandardized bond returns in which the various bond returns are weighted by the relative numbers of bonds outstanding. As noted above, while an average of unstandardized returns tends to be dominated by a firm's most volatile bonds, an average of standardized returns effectively gives less volatile notes and bonds (i.e. bonds with less price noise) greater weight. Consequently, we view an average of standardized returns as more appropriate.

The issue to which we now turn is whether test power can be improved by weighting standardized bond returns in calculating firm returns and, if so, what weights. One possibility we consider is to weight bond ABSRs by their relative numbers outstanding ala BKMX (2009). Since due to bid-ask bounce one would expect average prices based on many trades to be more reliable than prices based only a couple of trades, another alternative we consider is weighting each bond's ABSR by the square root of the number of trades on which the bond return is based relative to all of the firm's bonds. Finally, since results in section 6.1 indicate individual trade prices are best weighted by the square root of trade size, we calculate an average in which each bond's ABSR is weighted by the square root of the total trade volume on which the bond return is based relative to all the firm's bonds.

Power tests are reported in Table 9. Again we dispense with reporting size results since all are fairly well-behaved. As shown in Table 9, test power differs little among the four bond ABSR weighting schemes. For most (but not all) test statistics and cases, power is slightly highest when each bond ABSR is weighted by the relative square root of the number trades on which ABSR is based. Nonetheless, the power differences are slight. Since little seems to be gained by going to a more complicated weighting, we stick with an unweighted average of bond ABSRs.

7. The event window choice

The final issue we consider is the event window choice when working with transaction databases, such as TRACE or Mergent FISD. As noted above, of the possible firm-event days, only 38.6% have trades on both t-1 and t+1 so that returns over the (t-1, t+1) event window can be calculated. In particular, returns from t-1 to t+1 for many smaller firms go unobserved. Moreover, there may be only a trade or two on days t-1 and t+1, so given the bid-ask bounce in transaction prices, the calculated returns may contain considerable noise. Thus a question is whether more powerful and representative tests can be constructed by incorporating trades over a longer event window, such as (t-2, t+2) or (t-3, t+3) and, if so, how. Of course, the longer the window, the greater the non-event-induced fluctuations in bond returns which will tend to reduce test power. While BKMX find that single day event windows yield more powerful tests than monthly, there are possibilities in between.

Within longer event windows there are several possible return calculations. For instance, within the (t-2, t+2) window, returns could be calculated from t-2 to t+2 , from t-2 to t+1, or from t-1 to t+2, as well as from t-1 to t+1. We construct a composite return from all four giving more weight to the shorter window returns. Henceforth, we use $R(t\text{-}x,t\text{+}y)_i$ to indicate the return from t-x to t+y for firm i and $R\{t\text{-}x,t\text{+}y\}_i$, i.e., dates in braces, to designate a composite return based on all possible returns within the t-x to t+y window.

We begin by examining the consequences of expanding the event window from t-1 to t+1 to t-2 to t+2. Statistics for unstandardized non-composite abnormal firm returns $ABR(t\text{-}1,t\text{+}1)_i$, $ABR(t\text{-}1,t\text{+}2)_i$, $ABR(t\text{-}2,t\text{+}1)_i$, and $ABR(t\text{-}2,t\text{+}2)_i$ are reported in Panel A of Table10. In calculating abnormal returns for each event window, benchmark returns for bonds in the same maturity/rating category over the same window are subtracted from raw returns. As expected, the standard deviation of abnormal returns rises with the number of days over which the return is calculated. However, due to the negative correlation in daily returns (possibly due to bid-ask bounce), the increase in the return standard deviation as the event window is expanded from (t-1, t+1) is less than would be expected if daily returns were independent. If independent, the standard deviation of ABR(t-2, t+2) window should be approximately $2^{.5}$=1.414 times the standard deviation of ABR(t-1,t+1). However, the standard deviation ABR(t-2, t+2) is only 1.137 times that of ABR(t-1, t+1). Statistics for an average of the four (or as many as are calculable) ABRs are reported in the last row of Panel A.

Statistics for standardized abnormal returns, ABSR, are reported in Panel B where statistics for the composite ABSR{t-2, t+2} calculated as an average of the four (or as many as are calculable) individual ABSRs are reported in the final row. As compared with the ABR average, the composite ABSR{t-2, t+2} effectively puts more weight on the (t-1,t+1) return since its volatility is normally lowest and less weight on the (t-2, t+2) returns since its standard deviation is normally highest.[17] Note also that the standard deviation of ABSR{t-2,t+2} is .7259 which is less than that of any of the non-composite ABSRs. There are two reasons for this: 1) for an individual bond, some of the noise in individual window returns gets averaged out over different windows, and 2) since returns for some of a firm's bonds may be observed for one window and other bonds for another window, the firm return may be based on more bonds. In panel C, we break the sample of 884,995 firm ABSR{t-2,t+2} observations into two subsets: 1) the 692,068 for which ABSR(t-1,t+1) is calculable and 2) the subset that have trades on day t-1 and/or t+2 but not on both t-1 and t+1. Showing the possible benefit of bringing in price observations from other days even when ABSR(t-1,t+1) is calculable, the standard deviation is reduced from .8249 to .6628 for the original ABSR(t-1,t+1) sample.

Size and power results for tests based on composite ABSR{t-2,t+2} returns are reported in Table 11 where results for ABSR(t-1, t+1) returns are repeated in the second row of each panel for easier comparison. From the power results in Panels B and C, it is readily apparent that tests based on the composite returns over the four-day window, which utilizes price observations on days t-2 and t+2 as well as t-1 and t+1, are much more powerful than those based on price observations on days t-1 and t+1 alone. For the sign-rank test, the likelihood of type II error is cut approximately in half. Clearly the benefit of additional price observations outweighs the cost of the higher non-event noise in longer returns.

The higher power of tests based on the composite ABSR{t-2, t+2} could be due to the greater number of firm/day observations since some firms without bonds traded on days t-1 and/or t+1 do have traded bonds on t-2 and/or t+2. Alternatively, ABSR{t-2,t+2}'s improved test

---

[17] An alternative to averaging returns would be to average prices, i.e, to average the price observations on days t-2 and t-1 to obtain an average pre-event price and the observations on days t+1 and t+2 for an average post-event price and calculate a single return based on these averages. However, this treats the price observations on all days as equally informative. An advantage of our approach using standardized returns is that it puts more weight on returns over the shorter windows.

power could be higher because for the firms already in the dataset the increased number of price observations either averages out some of the noise in bond prices and/or increases the number of bonds over which the firm ABSR(t-1,t+1) is calculated. To explore this issue, we re-ran the power tests using ABSR{t-2,t+2} returns but only for firm/days for which ABSR(t-1,t+1) exists. In other words, the 192,926 added firm/day returns were treated as unavailable. Results are shown in the last row of Panels B and C. Since powers for this subsample are only slightly lower than those for the full sample (and in one case higher), it is clear that most of the increased test power is due to reducing the noise in returns based on only price observations on days t-1 and t+1. Expanding the firm/day dataset by including firm/days without trades on days t-1 or t+1 accounts for only a minor portion of the power increase.

Aside from the power increase, one advantage of using the ABSR{t-2, t+2} returns is the increased coverage of smaller firms. For the 692,068 firm/day observations for which ABSR(t-1,t+1) is calculable, median total firm assets are $15.1 billion {update} (as of December 2009); for the 192,926 firm/days added due to trades on days t-2 and t+2, median total assets are $6.5 {update} billion. For total sales, the figures are $10.5 billion and $5.2 {update both} billion.[18]

The finding that event studies based on the 4-day event window composite returns ABSR{t-2,t+2} are significantly more powerful than those based on ABSR(t-1,t+1) raises the issue of whether test power could be increased more by further broadening the event window. To explore this issue, we consider 6-, 8-, and 10-day composite returns: ABSR{t-3,t+3}, ABSR{t-4,t+4}, and ABSR{t-5,t+5}. As before these are measured as averages of all ABSR returns within the expanded window. Thus ABSR{t-3, t+3} is an average of ABSR returns over the windows (t-

[18] Total asset and total sales figures were available on Compustat for only 66.5% {update} of the firms in our sample. Some privately held firms issue public debt and some foreign firms issue dollar debt in the US. In addition when firms are acquired, their bonds often continue to trade under the old name and CUSIP. Hence we supplemented Compustat data with internet searches for the remaining firms. We were unable to obtain total asset figures for 12.8% of firms and unable to obtain total sales figures for 29.2%. We collected asset and sales figures as of December 2009. When only available for other dates, we adjusted the figures to a December 2009 basis using the producer price index.

3, t+1), (t-3,t+2), (t-3,t+3), (t-2,t+1), (t-2,t+2), (t-2,t+3), (t-1,t+1), (t-1,t+2), and (t-1,t+3). Again we would point out that since the standard deviation of ABR returns increases with the length of the event window, the use of standardized returns effectively assigns unstandardized returns over the shorter windows a greater weight. Longer windows may raise the power of the test because: 1) the number of event days with calculable returns increases, 2) the larger number of trades before and after day t averages out some of the noise in individual bond trade prices, and/or 3) average firm returns are based on more bonds. On the other hand, test power may be reduced since returns over longer windows include more non-event noise.

Statistics for 2-, 4-, 6-, 8-, and 10-day returns, as well as 2-day returns, are reported in Table 12. In Panel A, we report statistics for unstandardized, non-composite, bond returns, ABR. As expected, the standard deviation of unstandardized, non-composite returns rises as the event window lengthens but not as much as one would expect if daily returns were independent. Statistics for composite standardized abnormal returns are reported in Panel B. Note that as the event window is broadened: 1) the number of possible event dates for which returns are calculable rises but at a decreasing rate, and 2) the return standard deviation falls, again at a decreasing rate. Both should increase the power of the tests.

Size and power results for tests based on composite ABSRs for expanded event windows are reported in Table 13. As reported in Panel A, while the tests are reasonably well behaved, consistent with the negative means and medians in Panel B of Table 12,[19] the t and sign-rank tests show a slight bias toward finding false evidence of negative events. More importantly, all three tests show further increases in test power as the event window is broadened, albeit at a decreasing rate. For a negative 15 basis point event, the likelihood of Type II error for the sign-rank test falls from 21.95% for ABSR(t-1,t+1) to 6.97% for ABSR{t-3,t+3}, to 4.94% for ABSR{t-5,t+5}. For

[19] As shown in Panel A mean and median bond returns are close to zero but, as shown in Panel B, means and medians fall as the event window lengthens because negative (positive) abnormal returns tend occur on bonds and event windows with low (high) volatility.

a positive 15 bp event, the likelihood of Type II error for the sign-rank test falls from 20.69% for ABSR(t-1,t+1) to 9.05% for ABSR{t-3,t+3}, to 7.96% for ABSR{t-5,t+5}. Of course, the increase in power as the event window is broadened must be balanced against other considerations. For instance, if one is interested in testing for information leakage before the event or how quickly and efficiently the information is impounded, a 10-day event window is likely too long.

8. Summary and conclusions

Extant bond market event studies have employed tests with relatively low power to detect bond price reactions to announcements and events. Extending BKMX(2009), we show that much more powerful tests can be obtained by two major and one minor refinements to the procedures they suggest. First, because bond volatility differs substantially by term-to-maturity, rating, and other characteristics, it is important to remove this heteroskedasticity by standardizing bond returns by their estimated volatility. Return standardization leads to a considerable increase in the powers of the t-test and sign-rank test and slight for the sign test. This applies to both transaction and quote databases. Second, due to very thin trading and fairly large bid-ask spreads, basing bond and firm returns on trades on single days before and after an event leads to: 1) many event dates without a calculable return measure, and 2) noisy returns for the events with observed returns. Broadening the event window to 4, 6, 8, or 10 days and calculating composite returns based on price observations on several days before and after the event while putting more weight on returns based on trades on days closest to the event both increases the sample size and reduces the noise in calculated returns. We further find that: 1) weighting trade prices by the square root of trade size increases test power slightly, 2) restricting the sample to institution size trades has little impact on test power, and 3) restricting the sample to trades between dealers reduces test power.

The resulting improvement in test power is dramatic. For instance, for negative 15 basis point events, the power of the t-test rises from 19.89% for tests based on unstandardized abnormal bond returns from t-1 to t+1 for a sample of 300 possible firm/event days (including firm/event days without calculable bond returns)(Table 4) to 63.87% for tests based on standardized (t-1,t+1) returns (Table 4) to 84.53% for tests based on standardized composite returns over the t-3 to t+3 window (Table 13). For positive 15 bp events, the power of the t-test rises from 20.88% to 65.50% to 81.03%. For negative 15 bp events, the power of the sign-rank test rises from 47.71% for tests based on unstandardized abnormal bond returns from t-1 to t+1 to 75.69% for tests based on ABSR(t-1,t+1) returns to 93.03% for tests based on composite ABSR returns over the t-3 to t+3 window. For positive 15 bp events, the power of the sign-rank test rises from 51.85% to 77.40% to 90.95%. Expanding the event window also increases coverage of smaller firms.

We hope the procedures we outline prove helpful to bond market researchers. One issue left for further study is the impact of event clustering on the probability of type I error in bond market studies.

References

Barber, B., and J. Lyon. 1997. Detecting long-run abnormal returns: The empirical power and specification of test statistics. *Journal of Financial Economics* 43:341–72.

Bessembinder, H., K.M. Kahle, W. F. Maxwell, and D. Xu. 2009. Measuring abnormal bond performance. *Review of Financial Studies* 22, 4219-4258.

Bessembinder, H., W. Maxwell, and K. Venkataraman. 2006. Market transparency, liquidity externalities, and institutional trading costs in corporate bonds. *Journal of Financial Economics* 82:251–88.

Billett, M., D. King, and D. Mauer. 2004. Bondholders wealth effects in mergers and acquisitions: New evidence from the 1980s and 1990s. *Journal of Finance* 59:107–35.

Boehmer, E., J. Musumeci, and A. B. Poulsen. 1991. Event study methodology under conditions of event induced variance. *Journal of Financial Economics* 30:253–72.

Brown, S., and J. Warner. 1980. Measuring security price performance. *Journal of Financial Economics* 8:205– 58.

Brown, S., and J. Warner. 1985. Using daily stock returns: The case of event studies. *Journal of Financial Economics* 14:3–31.

Charest, G. 1978. Dividend information, stock returns, and market efficiency. *Journal of Financial Economics* 6:265-96.

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

Cook, D. and J. Easterwood. 1994. Poison put bonds: An analysis of their economic role. *Journal of Finance* 49: 1905-1920.

Ederington, L., and W. Guan. 2011. Dealer-client and interdealer spreads in the secondary corporate bond market, working paper

Edwards, A., L. Harris, and M. Piwowar. 2007. Corporate bond market transparency and transactions costs. *Journal of Finance* 62:1421–51.

Hand, J., R. Holthausen, and R. Leftwich. 1992. The effect of bond rating agency announcements on bond and stock prices. *Journal of Finance* 47:733–52.

Jaffe, J. 1974. Special information and insider trading. *Journal of Business* 47 :410-428

Kolari, J., and S. Pynnönen. 2010. Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies* 23: 3996-4025.

Kothari, S., and J. Warner. 2006. Econometrics of event studies. Chapter 1 in *Handbook of Corporate Finance: Vol 1 Empirical Corporate Finance*, E. Eckbo (ed.), Elsevier, North Holland.

Lyon, J. D., B. Barber, and C. Tsai. 1999. Improved methods for tests of long-run abnormal stock returns. *Journal of Finance* 53: 165-201.

Mandelker, G. 1974. Risk and return the case of merging firms. *Journal of Financial Economics* 1: 303-335.

Marais, L., K. Schipper, and A. Smith. 1989. Wealth Effects of Going Private for Senior Securities. *Journal of Financial Economics* 23:155–91.

Narayanan, J., and K. Shastri. 1988. The valuation impacts of specially designated dividends. *Journal of Financial and Quantitative Analysis* 23:301–12.

Patell, J. M. 1976. Corporate forecasts of earnings per share and stock price behavior: Empirical tests. *Journal of Accounting Research* 14:246-76.

Penman, S. 1982. Insider trading and the dissemination of firms' forecast information. *Journal of Business* 55:479–503.

Warga, A., and I. Welch. 1993. Bondholder losses in leveraged buyouts. *Review of Financial Studies* 6:959–82.

Table 1 - Distribution of 2-day bond and firm returns

Statistics are reported for 2-day returns from day t-1 to day t+1. Bond returns are calculated from average bond prices each day weighting each trade price by trade size. Abnormal 2-day bond returns are calculated as the percentage or log return from t-1 to t+1 minus the mean 2-day return for all bonds in the same rating/maturity class. Abnormal firm returns are calculated as an equally weighted average of abnormal returns on all the firm's bonds. Standardized abnormal returns in Panel C are calculated by standardizing 2-day bond returns by the standard deviation of the bond's 2-day returns from t-25 through t+25 excepting the t-1 to t+1 return. Abnormal standardized returns (ABSR) are calculated by standardizing before calculating abnormal returns; standardized abnormal returns (SABR) by standardizing after calculating abnormal returns. Returns are calculated from bond trades reported on TRACE from 1/1/2005 through 12/31/2011 for rated industrial bonds and notes maturing in 1 to 50 years excluding very rarely traded bonds.

	Mean	Median	Standard Deviation	Skewness	Excess Kurtosis	% positive	Observa-tions	% days traded
Panel A - bond returns -								
Raw percentage returns	.000518	.000226	.01701	0.177	18.572	51.45%	1,672,765	31.34%
Abnormal percentage returns	.000000	-.000078	.01570	0.140	19.352	49.41%	1,672,765	31.34%
Abnormal log returns	.000000	-.000029	.01568	-0.085	19.203	49.79%	1,672,740	31.34%
Panel B - abnormal firm returns -								
Abnormal percentage returns	.000057	-.000061	.01567	0.133	21.114	49.54%	752,204	38.61%
Abnormal log returns	.000048	.000012	.01565	-0.108	20.895	49.91%	752,188	38.60%
Panel C - standardized abnormal firm log returns								
Standardized abnormal. ret., SABR	-.002207	-.006070	.90127	0.036	1.830	49.64%	692,080	35.52%
Abnormal standardized ret., ABSR	.000478	.003908	.82671	0.045	2.025	49.74%	692,080	35.52%

Table 2 - Size and Power tests for percentage and log returns

In panel A, the percentage of times the tests incorrectly reject the no event null at the 5% level are reported. For 10,000 random samples, 300 firm/days are chosen at random from all days a firm had bonds outstanding and the tests are applied to those observations for which 2-day returns from t-1 to t+1 could be calculated. We report separately the percentage of times the no-event null is rejected in favor of finding evidence of a negative event (the 2.5% tail) and the percentage of time it is rejected in favor of a positive event (the 97.5% tail). In panel B, an artificial event is simulated by increasing or decreasing the (t-1,t+1) return by 15 basis points and the percentage of times a test correctly detects the event based on a 5% significance level is reported.

	t-test		Sign-rank test		Sign test	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Panel A - Size tests - no event						
Percentage returns - all trades	1.95%	2.61%	2.69%	2.52%	2.60%	1.68%
Log returns - all trades	1.84%	2.63%	2.21%	2.95%	2.00%	2.02%
Panel B - Power tests for 15 basis point return shocks						
Percentage returns - positive shock	21.06%		49.67%		56.19%	
Percentage returns - negative shock	19.30%		50.95%		61.13%	
Log returns - positive shock	21.31%		52.42%		59.22%	
Log returns - negative shock	18.90%		48.52%		58.03%	

Table 3 - Abnormal return standard deviations by rating and maturity

Standard deviations of 2-day abnormal bond log returns are reported stratified by rating and term-to-maturity. The bonds are classified by Moody's rating if available; if not, by Standard and Poors rating. Returns are based on average daily prices in which each trade price is weighted by the size of the trade. Abnormal returns are calculated as the difference between the log percentage return on the bonds and the average for bonds of the same rating and maturity grouping over the same two-day window.

Rating	Bond term-to-maturity			
	1 to 3 years	3+ to 5 years	5+ to 10 years	Over 10 years
Aaa and Aa	0.421%	0.527%	0.758%	1.574%
A	0.690%	0.835%	1.017%	1.766%
Baa	1.037%	1.332%	1.404%	2.004%
Ba	1.240%	1.394%	1.395%	2.146%
B	1.808%	2.104%	1.834%	2.186%
Below B	2.730%	2.815%	2.522%	3.094%

Table 4 - The size and power of test based on standardized and unstandardized abnormal returns

The size and power of tests based on standardized and unstandardized abnormal 2-day returns are compared. ABR is the unstandardized 2-day abnormal return calculated as the difference between the average return on the firm's bonds from day t-1 to t+1 minus the average return for bonds in the same maturity/rating category. SABR is the standardized return calculating by dividing ABR by its standard deviation over the 50 trading day period before and after the event date. ABSR is the difference between the standardized return and the average standardized return for bonds in the same maturity/rating category. The size and power tests are conducted as described in Table 2 with a 5% significance level.

Panel A - Size tests	t-test		Sign-rank test		Sign test	
Significance level	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
Unstandardized abnormal return, ABR	2.01%	2.65%	2.11%	3.21%	1.86%	2.29%
Standardized abnormal return, SABR	2.32%	3.02%	2.37%	3.13%	2.04%	2.49%
Abnormal standardized return, ABSR	2.27%	2.92%	2.37%	3.06%	1.93%	2.26%
Panel B - Power tests	Negative Event (-15 bp)			Positive (+15 bp)		
	t-test	sign-rank	sign	t-test	sign-rank	sign
Unstandardized abnormal return, ABR	19.89%	47.71%	56.47%	20.88%	51.85%	59.31%
Standardized abnormal return, SABR	66.01%	75.43%	66.55%	66.64%	76.96%	68.62%
Abnormal standardized return, ABSR	63.87%	75.69%	68.80%	65.50%	77.40%	69.61%

Table 5 - Test power stratified by rating and maturity for a 15 basis point return shock				
The percentage of times the sign-rank test based on abnormal standardized returns, ABSR, correctly finds evidence of an event which shifts 2-day bond returns by 15 basis points are reported by rating and maturity. Percentages for positive and negative return shocks are averaged. Results are based on 10,000 simulations at a 5% significance level for each maturity/rating category as described in Table 2. The tests are conducted at the bond level. The bonds are classified by Moody's rating if available; if not, by Standard and Poors rating.				
Rating	Bond term-to-maturity			
	1 to 3 years	3+ to 5 years	5+ to 10 years	Over 10 years
Aaa and Aa	100.00%	99.78%	87.01%	29.45%
A	99.94%	96.77%	78.02%	26.89%
Baa	95.73%	76.49%	60.03%	27.91%
Ba	64.85%	55.11%	49.04%	21.34%
B	43.80%	33.96%	38.76%	20.30%
Below B	25.08%	20.96%	23.24%	12.29%

Table 6 - Comparing the power and size of tests based on large trades versus all trades

The size and power of test statistics based on firm ABSR returns calculated solely from trades of 100 or more bonds are compared with test statistics based on all trades. Size results (the percentage of times the test found evidence of a false event) are reported Panel A and power results (the percentage of times the tests correctly found evidence of an event at a 5% significance level) are reported in panel B. All are based on 10,000 simulations where 300 firm/days are chosen at random from all possible firm/days 2005-2011 when the firm had bonds outstanding. In the power tests, results are reported both for an across-the-board shock in bond returns of 15 basis points and for proportional shifts where the shock for bonds of maturity m and rating r is proportional to the average 2-day return standard deviation for bonds of the same rating and maturity relative to all bonds.

Panel A - Size tests	t-test		Sign-rank test		Sign test	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
All trades	2.27%	2.92%	2.37%	3.06%	1.93%	2.26%
Trades of 100+ bonds only	2.71%	2.61%	2.70%	2.69%	1.97%	1.84%

Panel B - Power tests	Negative Event			Positive event		
	t-test	sign-rank	sign	t-test	sign-rank	sign
Across-the-board 15 bp shocks						
All trades	63.87%	75.69%	68.80%	65.50%	77.40%	69.61%
Trades of 100+ bonds only	66.85%	75.87%	64.60%	67.14%	75.71%	62.93%
Proportional shocks						
All trades	50.34%	61.53%	53.44%	52.01%	63.01%	54.41%
Trades of 100+ bonds only	53.34%	62.06%	50.23%	53.42%	61.94%	50.50%

Table 7 - The power and size of tests based on interdealer trades only

The size and power of test statistics based solely on interdealer trades are compared with test statistics based on all trades. Size results (the percentage of times the test found evidence of a false event) are reported Panel A and power results (the percentage of times the tests correctly found evidence of an event at a 5% significance level) are reported in panel B. All are based on 10,000 simulations where 300 firm/days are chosen at random from all firm/days from November 2008 (when TRACE began identifying interdealer trades) through December 2011 when the firm had bonds outstanding. In the power tests, results are reported both for an across-the-board shock in bond returns of 15 basis points and for proportional shifts where the shock for bonds of maturity m and rating r is proportional to the average 2-day return standard deviation for bonds of the same rating and maturity relative to all bonds.

Panel A - Size tests	t-test		Sign-rank test		Sign test	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
All trades	2.67%	2.19%	2.75%	2.21%	2.15%	2.04%
Interdealer trades only	2.61%	2.47%	2.93%	2.38%	2.12%	1.66%

Panel B - Power tests	Negative Event			Positive event		
	t-test	sign-rank	sign	t-test	sign-rank	sign
Across-the-board 15 bp shocks						
All trades	73.00%	84.41%	78.50%	72.06%	83.21%	76.75%
Interdealer trades only	71.20%	81.98%	74.36%	67.91%	79.14%	71.01%
Proportional shocks						
All trades	58.98%	70.39%	61.91%	57.68%	68.94%	60.34%
Interdealer trades only	54.43%	65.32%	56.53%	50.81%	61.28%	52.31%

Table 8 - Comparing the power of different measures of average daily prices

The powers of test statistics based on abnormal standardized returns, ABSR, are compared when average bond prices each day are calculated from: 1) an unweighted average of all transaction prices, 2) transaction prices weighted by trade size, and 3) transaction prices weighted by the square root of trade size. In panel A, the percentage of times the tests correctly found evidence of an event at the 5% significance level is reported where event day returns were shocked up or down 15 basis points. The tests in Panel B are identical to those in panel A except the event day shock for bonds of maturity m and rating r is proportional to the average 2-day return standard deviation for bond of the same rating and maturity relative to all bonds. Firm returns are an unweighted average of bond returns. All are based on 10,000 simulations where 300 firm/days are chosen at random from all possible firm/days.

Daily bond prices calculated from:	Negative Event			Positive event		
	t-test	sign-rank	sign	t-test	sign-rank	sign
Panel A - Power tests - 15 bp across-the-board shock						
Transaction prices weighted by trade size	63.87%	75.69%	68.80%	65.50%	77.40%	69.61%
Prices weighted by square root of trade size	66.87%	78.05%	71.17%	67.78%	79.31%	71.82%
Unweighted transaction price average	63.45%	74.65%	67.49%	64.71%	75.23%	67.36%
Panel B - Power tests - proportional shocks						
Trade prices weighted by trade size	50.34%	61.53%	53.44%	52.01%	63.01%	54.41%
Prices weighted by square root of trade size	52.90%	63.98%	55.50%	54.17%	65.20%	56.79%
Unweighted trade price average	49.88%	60.80%	51.83%	51.83%	61.76%	52.09%

Table 9 - Comparing the power of firm return statistics based on different weightings of bond returns

The powers of test statistics based on abnormal standardized returns, ABSR, are compared when 2-day firm returns are calculated as: 1) an unweighted average of bond returns, 2) bond returns weighted by bonds outstanding, 3) bond returns weighted by the square root of the number of trades, and 4) bond returns weighted by the square root of trade volume. In panel A, the percentage of times the tests correctly found evidence of an event at the 5% significance level is reported where event day returns were shifted up or down 15 basis points. The tests in Panel B are identical to those in panel A except the event day shocks are higher for low rated and/or long maturity bonds than for high rated short maturity bonds as described in the text. All are based on 10,000 simulations where 300 firm/days are chosen at random from all possible firm/days.

	Negative Event			Positive event		
	t-test	sign-rank	sign	t-test	sign-rank	sign
Panel A - Power tests - 15 bp across-the-board shock						
Unweighted average of standardized bond returns.	66.87%	78.05%	71.17%	67.78%	79.31%	71.82%
Standardized bond returns weighted by bonds outstanding	66.86%	78.26%	71.14%	68.10%	79.35%	72.23%
Standardized bond returns weighted by square root of trades	67.24%	78.51%	72.29%	69.40%	81.22%	74.19%
Standardized bond returns weighted by square root of trade volume	66.32%	77.41%	70.78%	68.34%	79.98%	72.95%
Panel B - Power tests - proportional shocks						
Unweighted average of standardized bond returns.	52.90%	63.98%	55.50%	54.17%	65.20%	56.79%
Standardized bond returns weighted by bonds outstanding	53.38%	64.05%	55.73%	54.48%	65.32%	56.94%
Standardized bond returns weighted by square root of trades	53.00%	64.47%	55.92%	55.80%	67.13%	58.80%
Standardized bond returns weighted by square root of trade volume	52.86%	63.40%	55.83%	54.75%	66.57%	58.21%

Table 10 - Abnormal return statistics for 2-day, 3-day, and 4-day event windows

Statistics are reported for two, three, and four day abnormal firm returns surrounding day t: (t-1, t+1), (t-1, t+2), (t-2, t+1), and (t-2, t+2). Statistics for unstandardized and standardized abnormal firm returns are reported in Panels A and B respectively. Statistics are also reported for a composite ABSR return, ABSR{t-2,t+2} which is an average of the (t-1, t+1), (t-1, t+2), (t-2, t+1), and (t-2, t+2) ABSRs (or as many of the four as are calculable due to thin trading). In Panel C, statistics on standardized abnormal returns are reported separately for two subsamples: 1) days t for which trades occur on both days t-1 and t+1 so that the (t-1, t+1) return is calculable, and 2) days t for which returns (t-1, t+2), (t-2, t+1) and/or (t-2, t+2), but not (t-1, t+1), are calculable.

Return Window	Obs.	Mean	Median	Standard Deviation	Skewness	Excess Kurtosis
Panel A - Unstandardized abnormal firm/day returns						
(t-1, t+1)	692,068	0.00001	0.00000	0.01463	-0.170	22.056
(t-1, t+2)	682,461	-0.00001	0.00001	0.01568	-0.166	20.707
(t-2, t+1)	682,665	-0.00001	0.00000	0.01568	-0.170	20.723
(t-2, t+2)	677,307	0.00001	0.00001	0.01663	-0.225	19.415
ABR average	884,994	-0.00004	-0.00000	0.01498	-0.248	22.378
Panel B - Standardized abnormal firm/day returns						
(t-1, t+1)	692,068	0.00032	-0.00013	0.8249	-0.016	2.021
(t-1, t+2)	682,461	-0.00077	-0.00134	0.8114	-0.000	1.898
(t-2, t+1)	682,665	-0.00071	-0.00096	0.8106	-0.003	1.896
(t-2, t+2)	677,307	-0.00062	-0.00085	0.8019	-0.003	1.864
Composite {t-2, t+2}	884,994	-0.00291	-0.00294	0.7259	-0.004	2.303
Panel C - Standardized abnormal composite {t-2, t+2} returns for subsamples with and without (t-1, t+1) returns						
With (t-1, t+1) return	692,068	0.00310	0.0008	0.6628	0.020	2.501
Without (t-1, t+1) return	192,926	-0.0245	-0.0022	0.9169	-0.002	1.170

Table 11 - Comparing the size and power of tests based on two-day and composite four-day event windows

The size and power of test statistics based on abnormal standardardized returns over the (t-1, t+1) window, ABSR(t-1,t+1), and composite abnormal standardized returns over the {t-2, t+2} event window, ABSR{t-2,t+2}, are compared. The ABSR{t-2, t+2} return is an average of the ABSR(t-1, t+1), ABSR(t-2, t+1), ABSR(t-1, t+2), and ABSR(t-2, t+2) returns. Results for the ABSR{t-2,t+2} test are reported for both tests based on both the full sample of ABSR{t-2,t+2} returns and the subsample for which ABSR(t-1,t+1) returns are calculable. In panel A the percentage of times the tests incorrectly found evidence of an event in 10,000 simulations in which 300 non-event firm/days were chosen at random are reported. In panel B, the percentage of times the tests correctly found evidence of a 15 basis point event at the 5% level is reported. The tests in Panel C are identical to those in panel B except the event day shocks are higher for low-rated and long-maturity bonds than for high-rated and short-maturity bonds.

Panel A - Size tests	t-test		Sign-rank test		Sign test	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
ABSR{t-2,t+2} over full sample	2.93%	2.48%	2.75%	2.42%	2.16%	1.97%
ABSR(t-1,t+1)	2.35%	2.71%	2.33%	2.89%	2.09%	2.25%
ABSR{t-2,t+2} over ABSR(t-1,t+1) subsample	2.18%	2.86%	2.17%	2.70%	1.83%	2.17%
Panel B - Power tests - 15 bp across-the board shocks	Negative Event			Positive event		
	t-test	sign-rank	sign	t-test	sign-rank	sign
ABSR{t-2,t+2} over full sample	80.06%	89.63%	85.01%	77.69%	88.36%	83.69%
ABSR(t-1,t+1)	66.87%	78.05%	71.17%	67.78%	79.31%	71.82%
ABSR{t-2,t+2} over ABSR(t-1,t+1) subsample	77.01%	87.20%	80.98%	80.26%	89.18%	83.35%
Panel C - Power tests - proportional shocks						
ABSR{t-2,t+2} over full sample	68.81%	79.85%	72.99%	65.82%	77.50%	70.48%
ABSR(t-1,t+1)	52.90%	63.98%	55.50%	54.17%	65.20%	56.79%
ABSR{t-2,t+2} over ABSR(t-1,t+1) subsample	63.14%	74.45%	66.23%	67.68%	78.06%	68.98%

Table 12 - Abnormal return statistics for 2-day, 4-day, 6-day, 8-day and 10-day event windows

In panel A statistics are reported for unstandardized abnormal bond returns, ABR, over 2, 4, 6, 8 and 10 day event windows. In panel B statistics are reported for composite standardized abnormal firm returns, ABSR. Composite returns are calculated as an average of standardized returns for all possible return windows with data within the wider window. For instance, ABSR{t-2, t+2} is an average of ABSR(t-1,t+1), ABSR(t-2,t+1), ABSR(t-1,t+2), and ABSR(t-2,t+2). Firm ABSRs are averages of the ABSRs for the firm's bonds.

Return Window	Obs.	Mean	Median	Standard Deviation	Skewness	Excess Kurtosis
Panel A - Unstandardized abnormal bond returns						
ABR(t-1, t+1)	1,565,924	-0.00001	0.00000	0.01461	-0.141	20.514
ABR(t-2, t+2)	1,535,135	-0.00000	0.00001	0.01627	-0.191	18.320
ABR(t-3, t+3)	1,502,344	-0.00001	0.00002	0.01762	-0.215	16.842
ABR(t-4, t+4)	1,469,234	0.00000	0.00003	0.01859	-0.215	15.544
ABR(t-5, t+5)	1,438,857	0.00000	0.00004	0.01948	-0.255	14.422
Panel B - Standardized abnormal composite firm returns						
ABSR(t-1, t+1)	692,068	0.00032	-0.00013	0.8249	-0.016	2.021
ABSR{t-2, t+2}	884,994	-0.00291	-0.00294	0.7259	-0.004	2.306
ABSR{t-3, t+3}	966,771	-0.00447	-0.00413	0.6751	0.003	2.503
ABSR{t-4, t+4}	1,006,043	-0.00512	-0.00459	0.6426	0.007	2.623
ABSR{t-5, t+5}	1,025,511	-0.00586	-0.00447	0.6216	-0.002	2.711

Table 13 - Comparing the size and power of tests based on 2-day, and composite 4-day, 6-day, 8-day and 10-day event windows

The size and power of test statistics based on abnormal standardardized returns over the (t-1, t+1) window, ABSR(t-1,t+1), and composite abnormal standardized returns over the {t-2, t+2}, {t-3,t+3}, {t-4,t+4}, and {t-5,t+5} event windows are compared. The composite returns are averages of all ABSR returns surrounding day t within the event window. In panel A the percentage of times the tests incorrectly found evidence of an event in 10,000 simulations in which 300 non-event firm/days were chosen at random are reported. In panel B, the percentage of times the tests correctly found evidence of a 15 basis point event at the 5% level is reported. The tests in Panel C are identical to those in panel B except the event day shocks are higher for low-rated and long-maturity bonds than for high-rated and short-maturity bonds.

Panel A - Size tests	t-test		Sign-rank test		Sign test	
ABSR(t-1, t+1)	2.35%	2.71%	2.33%	2.89%	2.09%	2.25%
ABSR{t-2, t+2}	2.93%	2.48%	2.75%	2.42%	2.16%	1.97%
ABSR{t-3, t+3}	2.93%	2.40%	2.96%	2.29%	2.40%	1.67%
ABSR{t-4, t+4}	2.95%	2.20%	3.05%	1.99%	2.60%	1.44%
ABSR{t-5, t+5}	3.15%	2.07%	3.25%	1.86%	2.80%	1.64
Panel B - Power tests - 15 bp across-the board shocks	Negative Event			Positive event		
	t-test	sign-rank	sign	t-test	sign-rank	sign
ABSR(t-1, t+1)	66.87%	78.05%	71.17%	67.78%	79.31%	71.82%
ABSR{t-2, t+2}	80.06%	89.63%	85.01%	77.69%	88.36%	83.69%
ABSR{t-3, t+3}	84.53%	93.03%	89.04%	81.03%	90.95%	86.87%
ABSR{t-4, t+4}	86.70%	94.42%	91.38%	82.07%	91.51%	88.73%
ABSR{t-5, t+5}	87.30%	95.06%	92.00%	82.33%	92.04%	89.12%
Panel C - Power tests - proportional shocks						
ABSR(t-1, t+1)	52.90%	63.98%	55.50%	54.17%	65.20%	56.79%
ABSR{t-2, t+2}	68.81%	79.85%	72.99%	65.82%	77.50%	70.48%
ABSR{t-3, t+3}	75.30%	85.30%	78.76%	69.78%	81.35%	74.67%
ABSR{t-4, t+4}	77.46%	87.44%	81.57%	71.82%	82.86%	76.99%
ABSR{t-5, t+5}	79.13%	88.49%	82.12%	71.27%	83.31%	77.60%