Measuring Abnormal Credit Default Swap Spreads

Christian Andres[1], André Betzer[2] and Markus Doumet[3]
[1]WHU - Otto Beisheim School of Management
[2]BUW - Schumpeter School of Business and Economics
[3]University of Mannheim

January 15, 2013

Abstract

This paper examines the size and power of test statistics designed to detect abnormal changes in credit risk as measured by credit default swap (CDS) spreads. In a spirit similar to that of Brown and Warner (1980, 1985) and Bessembinder et al. (2009), we follow a simulation approach to examine the statistical properties of normal and abnormal CDS spreads and assess the performance of normal return models and test statistics. Using daily CDS data, we find that parametric test statistics are generally inferior to non-parametric tests, with the rank test performing best. Some of the classical normal return models, such as the market model, are found to be poorly specified. A CDS factor model based on factors identified in the empirical literature is generally well specified and more powerful in detecting abnormal performance. If factor information is not available, a simple mean-adjusted approach should be used. Finally, we examine performance in the presence of event-induced variance increases and bootstrapped p-values. Our inferences hold for US and European CDS data and are not affected by reference entity credit quality.

Keywords: CDS spread; event study; Brown-Warner simulation
JEL-Classification: G14

1 Introduction

This paper employs data on credit default swaps (CDSs) to investigate the methodological particularities of event studies. Applying a simulation approach similar to that of Brown and Warner (1985) and Bessembinder et al. (2009), we examine the statistical properties of CDS data and the performance of current methods and models used to test for the presence or absence of abnormal changes in CDS spreads. Furthermore, based on the findings of the literature on the determinants of CDS changes, we test a CDS four-factor model that captures all potential relevant pricing factors identified in the literature.

Measuring the impact of any event on stock prices is among the most common empirical techniques in corporate finance. Brown and Warner (1980, 1985) show that under many conditions the market model and standard parametric tests are well specified in this research context. In addition, an important strand of the literature investigates the influence of firm-specific events on debtholder wealth.[1] In a recent paper, Bessembinder et al. (2009, p. 4256) examine the different methodologies used to detect abnormal bond returns and conclude that "the inferences drawn in previous studies could in fact be incorrect, depending on the sample size, the magnitude of the event... and whether parametric or non-parametric tests were used."

However, since reliable CDS data has become available from vendors such as Bloomberg and Markit, a growing literature measuring the impact of corporate events on debtholder wealth prefers to employ CDS data instead of bond data, among other things, for the following reasons: First, many bond event studies suffer from the fact that firms have a variety of bonds outstanding, with different maturities and credit ratings and differences in liquidity. In such cases, it is not obvious how the different value changes should be aggregated to gauge the total effect of the event. In contrast, only one CDS per firm needs to be valued and contract maturities are typically set to five years[2] (Bessembinder et al., 2009). Second, CDS spreads are a pure measure of credit risk, whereas bond spreads include factors that are unrelated to default risk, such as systematic risk (Longstaff et al., 2005; Callen et al., 2009). Third, empirical studies such as those of Daniels and Jensen (2005) and Zhu (2006) show that price discovery occurs first in the CDS market and subsequently in the bond market. In addition, CDS contracts are more liquid than corporate bonds (Bessembinder et al., 2009). Given these differences between CDS and corporate bonds, it is understandable that researchers might—under certain circumstances—favor CDS event studies over bond event studies.

[Insert Table 1 here]

[1] For an overview, see Bessembinder et al. (2009).

[2] In general, data on CDS contracts are available for maturities between six months and 30 years. However, CDS contracts with a maturity of five years are the most standard and most liquid contracts. Due to the limited availability of spread data for other maturities, we restrict our analyses to five-year contracts.

Table 1 gives an overview of previous CDS event studies. This strand of the literature is relatively young due to the fact that CDS data only became available about 10 years ago. We aim to provide researchers using CDS data with detailed knowledge of the performance of a variety of methods that are used to detect abnormal CDS changes. The evidence presented in this paper should help avoid biased inferences similar to those documented by Bessembinder et al. (2009) in the context of abnormal bond returns.

As already investigated by Brown and Warner (1980, 1985) and Bessembinder et al. (2009) in the case of stock and bond returns, respectively, the performance of the different methods depends essentially on two features: The method under scrutiny should avoid excessive Type I (falsely rejecting the null of no abnormal CDS spread change when it is in fact true) and Type II (not rejecting the null of no abnormal CDS spread change when it is in fact false) errors.

Our descriptive results show that the distributional properties of our CDS dataset are characterized by positive mean spread changes and a positive excess kurtosis independent of the reference entities' ratings (investment or non-investment grade), the region (United States or Europe), and the applied spread change measure (absolute or relative). These observations indicate that the distribution of realized spread changes is not normal. The distributional characteristics of abnormal spread changes similarly show deviations from normality (positive skewness and excess kurtosis). Consequently, parametric test statistics may lead to biased inferences in CDS event studies. Furthermore, our findings, (1) that not just the spread level but also the absolute spread change and (2) the CDS standard deviation are both negatively related with the reference entities' ratings, support the notion that relative and not (!) absolute spread changes should be applied in CDS event studies.

Our main simulation results indicate that only the non-parametric rank test is well specified when assessing the size of the tests (Type I errors). All other test statistics—parametric tests in particular—suffer from disproportionate Type I errors. In addition, similar to the findings of Bessembinder et al. (2009) in the context of abnormal bond returns, non-parametric test statistics are more powerful in detecting abnormal spread changes compared to standard parametric tests. However, in contrast to the work of Bessembinder et al. (2009), our simulations show that the matched portfolio approach used to calculate abnormal spread changes is the least powerful of all tested models. Hence, using this model in CDS event studies may lead to biased inferences. Across all parametric and non-parametric test statistics, the CDS factor model seems to be the best specified and most powerful model since it exhibits the highest detection rates. Surprisingly, the simple mean-adjusted approach also leads to very reasonable results and is recommended if factor information is not available. Our simulation results are very similar for the investment-grade and noninvestment-grade samples. In additional analyses we follow Boehmer, Musumeci and Poulsen (BMP) (1991) and reexamine the performance of the different test statistics when introducing an event-induced variance increase. As expected, only the test of BMP (1991) is consistently well specified under these conditions. However, at least for relatively low variance increases, the generalized sign and the cross-sectional t-test seem to be well specified.

Several previous studies (e.g. Hull et al., 2004) use a bootstrap approach to control for the bias that arises if the empirical distributions of test statistics are skewed. Accordingly, we present simulation results that account for erroneous assumptions about the empirical distribution of employed test statistics. Overall, the bootstrap approach only leads to a slight shift of empirical rejection rates toward the theoretical rates expected under the absence of abnormal performance. The power of the different test statistics is not affected when using bootstrapped p-values. The previous results are mainly confirmed since the new CDS four-factor model still outperforms all other normal return models.

The remainder of the paper is organized as follows. Section 2 describes the CDS data and specifies the methodology following Brown and Warner (1985). Section 3 describes the different spread measures used in previous papers and discusses their suitability. Furthermore, we present the models used in our subsequent simulation study to calculate expected credit default spreads. Section 4 describes the distributional properties of our initial dataset and analyzes the size and power for each combination of test statistic and normal CDS spread change model. Section 5 discusses specific problems in the context of event studies with CDSs, namely, the event-induced variance increase (Harrington and Shrider, 2007) and the skewness of the empirical distributions of the test statistics. Section 6 concludes the paper.

2 Data and Simulation Methodology

2.1 Data

Credit default swaps are contractual agreements between two parties that protect against the default risk of a reference entity (in most cases a company). The protection buyer pays a periodic fee, the spread, to the protection seller. In return, the protection seller is obligated to compensate the protection buyer if a predefined credit event (e.g., default or bankruptcy) happens to the reference entity. Generally, CDS contracts are standardized under the rules set by the International Swaps and Derivatives Association (ISDA) and are traded over the counter. They therefore lack the legal reporting and disclosure requirements of regulated exchanges. Consequently, reliable and complete information on spreads is not directly observable. This difficulty is compounded by the fact that neither transnational legal reporting requirements nor international clearing and settlement standards exist.[3] In contrast to stock event studies, a critical examination of the employed spread data is therefore an inevitable step prior to measuring abnormal CDS spreads.

Empirical studies on CDS data are based on either a limited set of transaction and quote data or use composite spreads derived from periodical surveys of major market participants. Transaction spreads are provided by interdealer platforms (e.g., GFI) or directly out of the book of a single market participant (e.g. Norden and Weber, 2004) but generally suffer from low reporting frequencies. While it is obvious that spread data from a single market maker

[3]Regulatory agencies in the United States and Europe started discussions about the introduction and/or information sharing of one (or more) central CDS counterparties in 2009.

are not representative of the overall CDS market, Mayordomo et al. (2010) argue that the same may apply to interdealer platform data. Their major concern is that interdealer data are incomplete, since CDSs are commonly negotiated via voice transactions. Due to this lack of reliable transaction and quote data, most empirical studies on CDSs employ composite spreads provided by specialized data vendors (see column 7 of Table 1). The general approach of these data vendors is based on a collection of book of record data from numerous major market participants. After cleaning these data for outliers and stale spreads, the remaining data are aggregated into composite spreads. By construction, these spreads are neither pure transaction nor quoted spreads. Nevertheless, we believe that composite spreads are a more representative valuation of the overall CDS market than spreads derived from interdealer platforms or a single market participant.

Even though most data vendors use a similar approach to derive their composite spreads, Mayordomo et al. (2010) show that composite spreads from different databases systematically diverge from common trends. Additionally, spreads reported by different vendors seem to adjust to new information at different speeds. According to Mayordomo et al. (2010), data from Markit Ltd. and CMA Datavision lead all other databases in terms of price discovery for European entities. The authors report comparable results for US entities, except that price data from CMA Datavision also lead data from Markit Ltd. While differences in price discovery are of minor importance in a simulation study with artificial abnormal effects, the database selection can have a substantial influence on the results of an applied event study. Since Markit Ltd. provides supplementary information for each contract, such as the contractual standards, the firm's country of domicile, and the average rating of both Standard and Poor's and Moody's, we use their daily CDS database as our main data source. Markit Ltd. employs the aforementioned collection and data cleaning processes to construct their composite spreads. As an additional restriction, they only report composite spreads whenever the pricing information is based on at least three different market participants.

Within their daily CDS database, Markit Ltd. provides composite spreads over the period from January 2001 to July 2011. However, since the ISDA published its refined contractual standards for CDSs on February 11, 2003, we do not consider CDS quotes prior to that date. Within their contractual standards for CDSs, the ISDA provides alternative definitions of feasible restructuring credit events. While most CDSs on North American entities follow the so-called modified (frequently referred to as Mod R) restructuring, CDSs on European entities follow the modifiedmodified (Mod Mod R) restructuring convention. Packer and Zhu (2005) show that differences in the restructuring clause of CDSs can non-negligibly affect their pricing. We therefore split our sample into two regional subsamples: North America and Europe. Within each subsample, we only consider corporate single-name contracts with a maturity of five years, since these are the most actively traded. We further restrict our sample to contracts written on senior unsecured debt to avoid any bias due to differences in seniority.

On April 8, 2009, the ISDA issued a supplement to the 2003 ISDA contractual standards that primarily led to a refinement of the credit event definitions and the terms of settlement. Two of these changes potentially affect our simulation results: (i) CDS contracts on North American entities do not include debt restructuring as a credit event. We therefore consider ex restructuring (XR) contracts for the North American subsample from the introduction of the supplement onwards. We further exclude contracts with an event window containing spread changes based on mixed restructuring clauses. (ii) CDSs are traded with a fixed coupon. For instance, CDSs on North American entities can be traded with a fixed coupon of 100 or 500 basis points. Additionally, a variable upfront payment can be used to account for the specific credit risk of a certain reference entity. However, fixed coupons and upfront payments can easily be converted into the previously used variable par spread and should therefore not affect our results. The final North American subsample consists of 2,146,519 daily quotes denominated in US dollars. The European subsample consists of 1,186,072 daily quotes denominated in Euros.

Additional data on stock market indices, volatility indices, and swap zero curves are obtained from Bloomberg. As stock market indices we use the Standard & Poor's (S&P) 500 for North America and the DJ Euro Stoxx for Europe. Stock market volatility is measured by the VIX Index for the North American subsample and the VSTOXX for the European subsample. As a proxy for the overall level of the zero curves we use five-year swap rates denominated in US dollars and Euros. The slope of each curve is estimated by the difference between the 10-year and one-year swap rates.

2.2 Simulation Methodology

We apply Brown and Warner's (1985) simulation approach to investigate the statistical properties of different test statistics and normal CDS spread change models. Based on 5,000 randomly drawn samples of realized spread changes, we examine the size and power for each combination of test statistic and normal CDS spread change model. We further assess the robustness of our results by varying the sample size and level of abnormal spread changes. In additional tests we examine how well the different test statistics behave in the face of event-induced variance and determine the suitability of bootstrapped p-values as a remedy for the non-normality of spread changes.

In a first step, we generate 5,000 samples consisting of 200 CDS contracts each. The contracts are randomly selected from our initial dataset. We further assign a random event date to each contract. There is no general rule as to whether the asset and event date should be drawn sequentially (Barber and Lyon, 1997; Brown and Warner, 1980, 1985) or simultaneously (Bessembinder et al., 2009). Moreover, related studies applying the sequential approach also differ in the order of selecting the asset and event date. Bessembinder et al. (2009) argue that the optimal random sampling procedure depends on the specific structure of the underlying database. For the case of CDSs, the number of tradable contracts as well as the quote frequency increase over time. We therefore follow the approach suggested by Bessembinder et al. (2009) and simultaneously select contract–event date combinations. This procedure leads to samples that are weighted toward later dates and firms that exhibit longer spread histories.

In unreported results available upon request, we conduct all simulations applying both sequential methods, but the results are virtually unchanged. We also control for contracts with insufficient spread data during the event and estimation period. Each combination of contract and event date must meet the following additional requirements to remain in a sample: (i) Spread changes must be observable for each day of the event window, which includes the 41 trading days, from -20 to 20; (ii) spread changes must be observable for at least 50% of the estimation window, which includes the 150 trading days, from -170 to -21; and (iii) the percentage of zero spread changes should not exceed 10%. If a combination does not meet all the criteria, the observation is dropped and a new random combination is drawn. We repeat this procedure for each sample until we reach the final sample size of 200 contracts.

3 Measuring Abnormal CDS Spreads

In this section, we discuss the methodological particularities of event studies employing CDSs. Column 5 of Table 1 shows that different spread change measures are used in the literature, namely, relative and absolute spread changes. Therefore, we discuss the suitability of both measures and provide a rationale for the use of relative spreads. We also briefly review normal return models applied in the classical event study framework and discuss their adaptation to CDSs. Column 4 of Table 1 describes the various models used to calculate expected spread changes in the prior literature. In addition, we go beyond these classical models and introduce a (four-) factor model for CDSs. In our view, it is appropriate to specify a unique model for expected (i.e., normal) CDS spread changes, since the recent empirical literature documents the importance of several determinants not accounted for in classical models. Accordingly, we build on the factors identified in the empirical literature on specifying our factor model. Finally, column 6 of Table 1 displays the different test statistics employed in previous studies. The most prevalent parametric and non-parametric test statistics used to assess the presence of abnormal spread changes are discussed in the Appendix B.

3.1 Spread Change Measures

In stock or bond event studies, abnormal returns are calculated to measure the impact of any event on security prices. In contrast, so-called abnormal spread changes are calculated whenever researchers quantify the effect of a firm-specific event on a firm's credit risk with the help of CDS data. A spread change measures the change in the premium of newly issued default swap contracts with constant maturity.[4] Both absolute (Hull et al., 2004; Norden and Weber, 2004; Galil and Soffer, 2011) and relative spread changes (Callen et al., 2009; Shivakumar et al., 2011) are used in the literature. The daily absolute spread change is simply the daily difference in the CDS spread:

$$\Delta_{abs} S_{i,\tau} = S_{i,\tau} - S_{i,\tau-1} \tag{1}$$

The daily relative CDS spread is the percentage change of daily spreads. We apply continuous compounding by calculating the difference in the logarithm of daily spreads:

$$\Delta_{rel} S_{i,\tau} = ln\left(S_{i,\tau}\right) - ln\left(S_{i,\tau-1}\right) \tag{2}$$

Even though relative spread changes are closer to the concept of returns, early CDS event studies in particular use absolute spread changes (Hull et al., 2004; Norden and Weber, 2004). In principle, the application of both calculation methods is correct. However, the distributional properties using one or the other may be quite different. This can potentially affect the size and power of the test in the presence of abnormal spread changes. Furthermore, the economic interpretations of absolute and relative average abnormal spread changes are different. Therefore, the choice of the right measure also depends on the expected effect on the firm's credit risk. For most corporate events it seems more plausible to expect an event-induced effect on credit risk that is proportional to the firm's initial default probability and its loss, given default. In line with this argument, we recommend the use of relative spread changes as an appropriate measure in most applications.

Independent of the choice of measurement, an abnormal spread change can be defined as the realized spread change minus the normal spread change:

$$\Delta AS_{i,\tau} = \Delta S_{i,\tau} - E\left[\Delta S_{i,\tau} | \Omega_{\tau}\right] \tag{3}$$

where $\Delta AS_{i,\tau}$, $\Delta S_{i,\tau}$, and $E\left[\Delta S_{i,\tau} | \Omega_{\tau}\right]$ are the abnormal, realized, and normal spread changes, respectively, for contract i at event date τ. The normal spread change is the expected spread change conditional on the information set Ω_{τ} (e.g., past spread changes). In the following sections we discuss different models for normal spread changes.

3.2 Naïve Models

In this section we define two naïve models based on models employed in current stock and bond event studies.

[4] Spreads are only reported for newly issued contracts. However, a time series with transaction prices on a specific contract is necessary to calculate CDS returns. Furthermore, CDS returns differ for protection buyers and sellers. For a detailed discussion on the calculation of CDSs, see Berndt and Obreja (2010).

Mean-Adjusted Spreads

The mean-adjusted model is possibly the simplest normal return model. Brown and Warner (1985) discuss this model in the context of stock returns. The normal return is the arithmetic average of realized returns in the estimation period. Accordingly, the abnormal return of stock i at event date τ is the difference between the realized return and the estimated mean return. Handjinicolaou and Kalay (1984) use a similar model for bonds. Instead of realized bond returns, they calculate returns in excess of matched Treasury securities. Applying the event study methodology to CDS data, we estimate the normal return calculating the sample mean spread change of firm i in the estimation period:

$$\overline{\Delta S_i} = \frac{1}{150} \sum_{t=-170}^{-21} \Delta S_{i,t} \tag{4}$$

Assuming that the normal spread can differ by firm but is constant over time, the abnormal spread change of firm i at event date τ is

$$\Delta AS_{i,\tau} = \Delta S_{i,\tau} - \overline{\Delta S_i} \tag{5}$$

Brown and Warner (1985) show that short-term stock event studies based on the mean adjusted model yield similar results to event studies based on more sophisticated models such as the market model or the capital asset pricing model (CAPM). However, Bessembinder et al. (2009) cannot confirm this finding for bond event studies. They document that the mean-adjusted model is the least powerful of all examined approaches. Therefore, the performance of this model in the context of CDS event studies is an open question.

Market Model

The market model is the workhorse of stock event studies. This model is a single-factor asset pricing model based on the assumption of a stable linear relation between individual stock returns and the return of a broad market index. We adapt this model to the case of CDS spreads and define the model equation

$$\begin{aligned} \Delta S_{i,\tau} &= \alpha_i + \beta_i \Delta S_{index,\tau} + \epsilon_{i,\tau} \\ E\left[\epsilon_{i,\tau}\right] = 0 &\quad VAR[[\epsilon_{i,\tau}] = \sigma^2_{\epsilon_i} \end{aligned} \tag{6}$$

where $\Delta S_{index,\tau}$ is the spread change of the CDS index and $\epsilon_{i,\tau}$ is the zero mean disturbance term. The daily index spread at time τ is equal to the mean credit default spreads of all firms in our dataset at time τ.[5] The parameters of the market model are α_i, β_i, and $\sigma^2_{\epsilon_i}$. We estimate

[5]We calculate separate indices for North America and Europe because of the different restructuring clauses. We additionally construct an equally weighted index of all (non)investment-grade contracts that is used for robustness tests based on the (non)investment-grade subsample.

the model parameters for each firm based on the spread changes in the estimation period. The abnormal spread change of firm i at event date τ is, accordingly,

$$\Delta AS_{i,\tau} = \Delta S_{i,\tau} - \widehat{\alpha}_i + \widehat{\beta}_i \Delta S_{index,\tau} \tag{7}$$

MacKinlay (1997) points out that the ability to detect event effects increases with the R^2 of the market model regressions. This implies that the market model dominates the mean-adjusted model in terms of size and power if it also exhibits greater explanatory power for spread changes.

3.3 Matching Portfolios

Another technique often applied in bond and long-term stock event studies to generate normal returns is the matched portfolio approach. The abnormal return is the difference between the respective firm's realized return and the return of a reference portfolio. The reference portfolio contains firms that resemble the event firm in certain risk characteristics but are assumed to be unaffected by the particular event. While matching on firm characteristics such as size or the market-to-book ratio is common for long-term stock event studies (Ritter, 1991; Loughran and Ritter, 1995; Barber and Lyon, 1997), the bond rating is the most important matching criterion in bond event studies (Asquith and Kim, 1982; Bessembinder et al., 2009). Since CDS spreads should depend on the expected default probability and the expected loss of the reference obligation, we adapt the matching to ratings. Therefore, we define the abnormal spread change as the difference between the realized spread change and the spread change of a rating-equivalent reference portfolio:

$$\Delta AS_{i,\tau} = \Delta S_{i,\tau} - \Delta S_{RE,\tau} \tag{8}$$

where $\Delta S_{RE,\tau}$ is the spread change of the rating-equivalent portfolio. For each rating letter we calculate daily spread changes as the average spread change of all available contracts with the corresponding rating while we exclude event firms for their entire event period. The rating for each contract is the average issuer rating of S&P and Moody's as reported by Markit Ltd.

Bessembinder et al. (2009) recommend applying a value-weighted matched portfolio approach for bond event studies. Based on their simulation results, they conclude that this approach combined with non-parametric test statistics is well specified and dominates all other approaches in terms of power. The reluctance of rating agencies to make timely rating adjustments (also referred to as rating stickiness), as documented, for example, by Posch (2011), could be seen as a potential disadvantage of adopting the portfolio approach to CDS spreads.

3.4 The CDS Factor Model

The three aforementioned approaches are an adaptation of the classical event study methodology. In them, we implicitly assume that spread changes are sufficiently explained by common pricing factors. However, researchers have identified additional important determinants of credit spread changes. Collin-Dufresne et al. (2001) investigate the impact of possible determinants of credit risk on bond spread changes. They derive theoretical determinants of spread changes from structural models of default. Their findings imply that aggregate factors exhibit a higher explanatory power than firm-specific factors. Ericsson et al. (2009) conduct a similar regression analysis with CDSs. They identify leverage, equity implied volatility, and the level of the Treasury yield curve as major determinants of credit spread changes. Alexander and Kaeck (2008) apply a regime-switching model to identify the regime-dependent determinants of CDS index spread changes. In addition to the previous findings, they document a statistically significant relation between the slope of the risk-free yield curve and spread changes. Furthermore, they provide evidence that the factor loadings of major determinants significantly differ in regimes of high-/low-volatility CDS markets.[6]

Based on these findings, we consider the following market-wide factors as potential explanatory variables in our factor model: (i) the level of the risk-free yield curve, (ii) the slope of the risk free yield curve, (iii) the equity implied volatility, and (iv) stock market performance. We use five-year swap rates as a proxy for the level of the risk-free yield curve. The difference between 10- and one-year swap rates serves as a proxy for the slope. We use swap zero curves instead of Treasury zero curves since the results of Hull et al. (2004), Blanco et al. (2005), and Houweling and Vorst (2005) indicate that the swap zero curve seems to be the relevant risk-free rate on credit derivative markets. Following the literature on the determinants of spread changes (e.g. Collin-Dufresne et al., 2001; Ericsson et al., 2009), we measure the equity implied volatility by the VIX index for the North American subsample and the VSTOXX for the European subsample. The stock market index for the North American subsample is the S&P 500. The stock market index for the European subsample is the Dow Jones Euro Stoxx Index.

An important step in the derivation process of our final factor model is the identification of potentially redundant variables. For this purpose, we calculate the pairwise correlation coefficients of all the determinants of credit spread changes identified in the literature and test for multicollinearity using variance inflation factors. Table 2 provides an overview of the pairwise correlation coefficients for relative spread changes. We find that stock market returns and equity implied volatility are highly correlated, with a correlation exceeding -70% in both subsamples. Since our tests for multicollinearity based on all variables also indicate a problem with the variable stock market returns, we drop this variable.

[Insert Table 2 here]

Table 3 provides the variance inflation factors based on the remaining variables. Since all variance inflation factors are close to one, we find no signs of serious multicollinearity.

[Insert Table 3 here]

[6] We estimate individual model parameters for each firm and consider short time periods. We therefore do not employ a regime-switching approach. However, researchers conducting long-term event studies should allow for time-varying model parameters.

Based on the remaining variables, we define the abnormal spread change of firm i at event date τ as

$$\Delta AS_{i,\tau} = \Delta S_{i,\tau} - \widehat{\alpha}_i - \widehat{\beta}_{1,i}\Delta S_{index,\tau} - \widehat{\beta}_{2,i}\Delta Level_\tau - \widehat{\beta}_{3,i}\Delta Slope_\tau - \widehat{\beta}_{4,i}\Delta Vola_\tau \qquad (9)$$

where $\Delta Level_\tau$ and $\Delta Slope_\tau$ are the proxies for the level and slope of the risk-free yield curve and $\Delta Vola_\tau$ is the proxy of the equity implied volatility.

4 Results

4.1 Distributional Properties of CDS Spreads

Before we analyze the size and power for each combination of test statistic and normal CDS spread change model, we provide a basic description of the distributional properties of our initial dataset. We examine the distributional characteristics of realized CDS spread changes and abnormal CDS spread changes to obtain a first impression about which test statistics might be best suited for the detection of any firm-specific event's effect on credit risk.

[Insert Table 4 here]

Panel A of Table 4 shows descriptive statistics for the entire sample based on relative spread changes. Statistics are further broken down into two subsamples: investment grade and non-investment-grade reference entities. As a striking first result, we observe positive mean spread changes for all subsamples. This effect is independent of the references entities rating, region, and applied spread change measure, indicating a general widening of spreads over our observation period. Because all median spreads are equal to zero, the resulting distributions exhibit positive skewness. Furthermore, we observe a positive excess kurtosis similar to daily stock returns (e.g. Cont, 2001). These features suggest a non-Gaussian character of the distribution of realized spread changes. As can be seen in Panel B of Table 4, the aforementioned distributional properties largely apply to absolute spread changes, too. However, skewness and excess kurtosis are much larger when considering absolute compared to relative spread changes.

As discussed in Section 2.1, the selection of the correct spread change measure depends on its distributional properties and the assumed event effect. While mean relative spread changes do not change with the rating, we observe an increase in mean absolute spread changes as the rating decreases from investment grade to non-investment grade. This suggests that not just the spread level but also the absolute spread change is statistically negatively related to the reference entities' rating. We also find a negative statistical relation between the rating and the standard deviation. As already pointed out in Section 2.1, these findings support the notion that relative spread changes should be applied in CDS event studies. This does not necessarily imply that abnormal CDS spread changes and the related test statistics also suffer from non-Gaussian distributions. Table 5 displays the distributional properties of abnormal spread changes obtained by using the normal return models described above.

[Insert Table 5 here]

At this point, no abnormal performance has been introduced. Nevertheless, the results in Table 5 show deviations from zero, both for mean and median daily abnormal spread changes. Compared to the spread changes discussed above, departures from normality are a little less pronounced for relative abnormal spread changes (Panel A). However, daily abnormal spread changes still exhibit positive skewness and excess kurtosis. This indicates that parametric test statistics may yield biased results. In addition, the distribution of abnormal CDS spread changes seems to differ substantially across the different normal return models. The CDS factor model seems to produce the lowest level of skewness and excess kurtosis. In contrast, the matching portfolio approach results in abnormal spreads that exhibit the highest levels of skewness and excess kurtosis. Concerning absolute CDS spread changes (Panel B), all models except the CDS market model show comparatively large positive abnormal spread changes. Skewness and excess kurtosis remain very high for absolute CDS spread changes.

As discussed above, our main analysis focuses on relative CDS spread changes.[7] Moreover, all simulations yield very similar results for the subsamples of investment-grade and non-investment-grade reference entities. Accordingly, we do not present detailed results for these subsamples. They are available upon request.

4.2 Size of Tests

We start our main analysis by estimating the empirical size of the different test statistics. Tables 6 and 7 contain results for our simulation of 5,000 randomly drawn samples of 200 abnormal CDS returns and document the probabilities with which the null hypothesis of no abnormal performance is rejected. Probabilities are based on a standard t-test (both time series and cross-sectional), the test of BMP (1991), a rank test, and a generalized sign test. If a test statistic is well specified, the empirical rejection rate should not deviate significantly from the assumed theoretical significance level. Considering a two-sided test at the 5% level of significance, this corresponds to a rejection rate of 2.50% in the lower tail and 2.50% in the upper tail. We follow the common assumption of normality of the underlying Bernoulli process (Bessembinder et al., 2009; Campbell et al., 2010) to test whether the empirical rejection rates deviate significantly from 2.50%. On the basis of 5,000 random samples, the test statistic should be between 2.07% and 2.93% in 95% of cases.[8] As Bessembinder et al. (2009), we are primarily concerned with

[7]We report results for the empirical size and power of absolute CDS spread changes in Appendix A.

[8]Under the assumption that the outcomes of each test of the 5,000 trials are independent, the underlying Bernoulli process implies a mean rejection rate of 0.025 (lower or higher tail of 2.50%), with a standard deviation of 0.0022 ($=\sqrt{0.025 \times 0.975}/\sqrt{5,000}$). The proportion of rejections should hence be between $0.025 \pm (1.96 \times 0.0022) = 2.07\%$ and 2.93% in 95% of the cases for a significance test at the 5% level.

rejection rates that are too high in the absence of abnormal spread changes, that is, Type I errors, or overrejections. All combinations of normal return models and test statistics that yield rejection rates that are significantly higher than 2.93% erroneously show a significant effect of a (non-existent) event on spread changes and should therefore not be used in CDS event studies. Even though too low a rejection rate is not directly a problem for the specification of the test statistics, we also mark significant underrejection in the tables. A rejection rate that is too low under validity of the null hypothesis of no abnormal return may indicate low power. The results are reported by region in Table 6, with North American spreads in Panel A and European data in Panel B.

[Insert Table 6 here]

The result in both panels of Table 6 show that only the non-parametric rank test is well specified for most models. All other test statistics show rejection rates that are strongly asymmetric and lie outside of the expected range. In the lower tail, rejection rates are generally too low. On the other hand, most test statistics reject the null hypothesis too frequently in the upper tail. With the exception of the test of BMP (1991), however, rejection rates in the lower and upper tails frequently sum to 5%. The test of BMP (1991) also reveals an asymmetric distribution but leads to rejection rates that are often too low in both tails. As mentioned above, this is not directly a problem for the specification of this test statistic but it may point to a lower power. These findings are robust across regions and credit quality.

With respect to the normal return models, we find that the market model performs worst in both subsamples, regardless of the test statistic used. In the lower tail, all test statistics show significant deviations from the theoretical significance level. The rating portfolio approach performs slightly better. Together with the CDS factor model and the mean-adjusted model, it shows rejection rates that are consistently within the expected range for the non-parametric rank test. Again, our results do not show qualitative differences across the regional subsamples and do not seem to be affected by credit quality (not reported here but available upon request).

Summarizing the results of the size tests, we find that the non-parametric rank test is the only test statistic that is well specified across all models and should hence be used for event studies based on CDS data. The test of BMP (1991) also does not lead to excessive Type I errors, but we expect low power due to very low rejection rates. In terms of normal return models, researchers should either use a CDS factor model or, alternatively, when factor data are not available, the simple mean-adjusted model.

4.3 Power of Tests

In this section, we examine the performance of the different models and test statistics with regard to potential Type II errors (not rejecting the null of no abnormal CDS spread change when it is false). By introducing positive and negative relative spread change shocks on day zero of +0.5% and -0.5%, respectively, we observe how frequently the null hypothesis of no abnormal CDS spread change is correctly rejected.[9]

Based on the results displayed in Table 7, we document the same empirical pattern across different regions. Surprisingly and in contrast to the findings of Bessembinder et al. (2009) in the context of abnormal bond returns, our simulations show that the matched portfolio approach used to calculate abnormal spread changes is the least powerful of all models. Hence, using this model in CDS event studies may lead to biased inferences. Independent of the use of parametric and non-parametric test statistics, the CDS factor model seems to be the most powerful model since it exhibits the highest detection rates overall. Similar to the findings of Bessembinder et al. (2009) in the context of abnormal bond returns, non-parametric test statistics are more powerful in detecting abnormal spread changes compared to the standard parametric tests. Given shocks of +0.5% and -0.5%, respectively, the non-parametric rank test rejects the null hypothesis in at least 98% of the cases across all models. The power of the generalized sign test is qualitatively similar but performs slightly worse when positive shocks are introduced. The performance of the simple t-statistics is very volatile since rejection rates range between 25% and 70%, depending on the CDS model used. Hence, these are the least powerful test statistics in CDS event studies. The power of the non-parametric statistic of BMP (1991) is in between the simple test statistics and the non-parametric test statistics. Its rejection rates range between 55% and 82%.

[Insert Table 7 here]

In sum, we conclude that relying on the CDS factor model and using the non-parametric generalized sign or rank test is the best method to detect abnormal CDS spread changes that are in fact true. Alternatively, if, for example, the data to construct the CDS factor model are not available, researchers can also rely on the mean-adjusted or market model since these models perform only slightly worse than the CDS factor model if the non-parametric tests are used.

Given the above conclusions, we now take into account the fact that the sample size and level of shocks in CDS event study applications may differ with regard to the specific event under scrutiny. Therefore, similar to Bessembinder et al. (2009), we run simulations where the number of observations varies from 50 to 200 and the abnormal shock varies from -1% to +1%. We apply the CDS factor model, which has been shown to be the best-performing model in the

[9] We think that introducing a shock of +/- 0.5% is, on the one hand, reasonable and, on the other hand, "conservative", since this "artificial" shock accounts for less than one-third of the daily realized unsigned spread changes (the average unsigned relative spread change for our European sample is 1.84% and that for our North American sample is 1.66%).

context of CDS event studies in our previous analyses. Our goal is to evaluate the power of the different test statistics under the changing parameters (number of observations and level of shock). Due to the complexity of the results, they are better presented in graphical form, as seen in parts (a) and (b) of Figure 1. The most important finding is that the conclusions we draw on the basis of +/- 0.5% shocks hold true for different sample sizes and different levels of shocks. As Figure 1 shows, we observe that the non-parametric rank test performs substantially better than the parametric test of BMP (1991) along all different dimensions. We choose the test of BMP (1991) among the parametric tests and the rank test among the non-parametric tests since both tests perform best in their respective groups.

[Insert Figure 1 here]

5 Robustness and Extensions

5.1 Event-Induced Variance Increases

Higgins and Peterson (1998) and Harrington and Shrider (2007) argue that almost any event will induce an increase in cross-sectional variance. This implies that the variance of abnormal returns is higher in the event window (compared to the estimation window). It has been shown that such an event-induced increase in variance can lead to a severe bias of classic test statistics (e.g. Brown and Warner, 1985; Corrado, 1989; Boehmer et al., 1991). This effect is exacerbated if the estimate of the variance of abnormal returns is based only on estimation window returns. For the null of no abnormal return, the event-induced increase in variance will lead to excess rejection rates. Accordingly, we expect the simple time series t-test in particular to be misspecified in the presence of event-induced variance increases.

On the other hand, the importance of this potential bias is disputed in the literature. First, it is not a priori obvious that all events necessarily increase the cross-sectional variance. Depending on the research setting, the event of interest may even lower uncertainty, thus leading to lower variance. Brown and Warner (1985, p. 22) speak of "*some* types of events" around which returns increase, which also implies that the type of event matters. Second, even in the presence of event-induced volatility, the impact on the validity of conventional test statistics may be very limited. As pointed out by Corrado (2011), the importance of the bias depends critically on whether interest is in the sample per se or an extraneous population of similar events.[10] In many cases, the sample can be very close to or even *be* the population itself (e.g., in the case of historical events). In these cases, where interest is in the mean event-induced return, variance increases are not relevant by definition. Test statistics that account for event-induced variance increases only become important if interest is in the population, that is, when inferences beyond the sample mean to the population mean are required. Corrado (2011, pp. 218-219) concludes,

[10]See Corrado (2011, p.216) for an extensive discussion.

> "A perusal of the finance literature suggests that many event studies limit themselves to statistical inferences about the mean event-induced return within the sample ... without projecting inferences onto the mean return of a parent population. This may suggest a conservative bias; however, this bias is diminished by the inevitable follow-up studies with extended data sets that form ongoing streams of research into interesting and important topics. Nevertheless, projecting inferences onto a population larger than the sample can often be instructive. In forming these inferences, cross-sectional variance adjustment procedures advanced in BMP (1991), Sanders and Robins (1991), or Corrado and Zivney (1992) are aptly recommended."

It is thus up to the researcher to decide whether the issue is important in a specific research setting. As a next step, we therefore model an event-induced variance increase and reexamine the performance of the different test statistics. Following the literature (e.g. Brown and Warner, 1985; Boehmer et al., 1991), we assume that the variance increases proportionally to the variance of abnormal returns in the estimation window. In most applications, the variance estimator of the individual time series is used. BMP (1991), however, additionally consider the average variance across all observations in the estimation window. In our simulation, we apply both methods. A constant shock μ as well as a normally distributed random variable with mean zero and a variance that is proportional to the variance of the estimation window (σ^2) are added to the realized abnormal spread change:

$$\mu + x \quad with \quad x \sim N(0, k\sigma^2) \tag{10}$$

with k standing for the proportionality factor. BMP (1991) derive economically plausible values for this factor based on several empirical papers (Charest, 1978; Mikkelson, 1981; Penman, 1982; Rosenstein and Wyatt, 1990) and conclude that values for k should lie in a range between 0.44 und 1.25. Accordingly, we use values of 0, 0.5, 1, and 1.5, with $k = 0$ standing for no increase in event-induced variance.

In line with our expectations, the findings in Table 8 indicate that the test of BMP (1991) is the only test statistic that is consistently well specified for a variance increase within an economically plausible range. All other test statistics reject the null of no abnormal return too frequently. As an exception, the generalized sign test and the cross-sectional t-test seem well specified, at least for comparatively low variance increases. The results of the power test in Table 9 further show that the power of the different test statistics decreases significantly with increases of the proportionality factor k. Given the comparatively low magnitude of the shock (+/- 0.50%), the power of all tests seems to be very low for values of k that exceed 0.5. Rejection rates for both the test of BMP (1991) and the non-parametric tests only become reliable for shocks larger than +/- 1% (results not tabulated here but available upon request). To summarize the size and power tests in Tables 8 and 9, the test of BMP (1991) seems to be the only test that produces reliable results in the presence of event-induced variance increases.

[Insert Tables 8 and 9 here]

5.2 Bootstrapped p-Values

If we take our simulation of event-induced variance increases into account, no single test statistic seems to be consistently well specified. In addition, parametric tests suffer from low power in detecting abnormal spread changes. This is mainly due to the fact that the empirical distributions of our test statistics are skewed, which contradicts the assumption of the theoretical normal distribution. Hull et al. (2004), without providing evidence, suggest a bootstrap approach to control for this bias. In a similar setting, Barber et al. (1999) show that even a skewness-adjusted t-test of abnormal buy-and-hold returns deviates from its theoretical distribution under validity of the null hypothesis. The authors argue that a simple bootstrap method will lead to much improved results. We adopt this approach for the case of CDSs. To do so, all abnormal announcement day spread changes are adjusted by their corresponding means. In the next step, 1,000 random samples of size $n/2$ are generated from the original test statistic.[11] This procedure results in an empirical approximation of the null distribution.

[Insert Tables 10 and 11 here]

The results in Table 10 show that bootstrapped p-values consistently lead to rejection rates that are within the theoretically expected range. This holds across all normal return models. The asymmetry documented in Section 4.2 is no longer present, irrespective of regional subsamples. While bootstrapping apparently leads to large improvements in model specification, the power of the different test statistics is not positively affected. As Table 11 shows, the rejection rates remain comparatively low. However, in contrast to the results presented in Section 4.3, the rejection rates are roughly similar for positive and negative shocks of the same magnitude. Again, the CDS factor model slightly outperforms other normal return models.

6 Conclusion

This paper extends the findings of Brown and Warner (1980, 1985) and Bessembinder et al. (2009) by applying their simulation approach in the context of stock and bond returns to CDS spread changes. We provide evidence as to which models and test statistics are best suited for empirical applications that investigate the impact of firm-specific or macroeconomic events on firm credit risk. We measure credit risk by examining the change in value of firm CDSs.

Our main finding is that when employing daily CDS data, the non-parametric rank test is the only test statistic that performs well across all models with regard to the avoidance of excessive Type I and II errors. Some of the classical normal return models such as the market model or the matching portfolio approach are only poorly specified. A CDS four-factor model based on the findings of the previous literature on CDS spreads is generally well specified and performs best in detecting abnormal CDS spreads. Surprisingly, the simple mean-adjusted approach also

[11] As a robustness test, we use a random sample of size $n/4$. All results are virtually identical.

leads to very reasonable results and is recommended if data on the different factors are not available.

In additional analyses we follow BMP (1991) and reexamine the performance of the different test statistics when introducing an event-induced variance increase. As expected, only the test of BMP (1991) is consistently well specified under these conditions. However, at least for relatively low variance increases, the generalized sign and cross-sectional t-test seem to be well specified.

Since several previous studies (e.g. Hull et al., 2004) use a bootstrap approach, we also present results for simulations accounting for the fact that the results may be biased because of inappropriate assumptions about the empirical distribution of the test statistics employed. Overall, the power of the different test statistics is not affected when using bootstrapped p-values. The previous results are mainly confirmed as the CDS four-factor model outperforms all other normal return models.

References

Alexander, C. and A. Kaeck (2008). Regime dependent determinants of credit default swap spreads. *Journal of Banking and Finance 32*, 1209–1228.

Asquith, P. and E. Kim (1982). The impact of merger bids on the participating firms security holders. *Journal of Finance 37*, 1209–1228.

Barber, B. M. and J. D. Lyon (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics 43*, 341–372.

Barber, B. M., J. D. Lyon, and C. L. Tsai (1999). Improved methods of long-run abnormal stock returns. *Journal of Finance 54*, 165–201.

Berndt, A. and I. Obreja (2010). Decomposing european cds return. *Review of Financial Studies 14*, 189–233.

Bessembinder, H., K. Kahle, W. Maxwell, and D. Xu (2009). Measuring abnormal bond performance. *Review of Financial Studies 22 (10)*, 4219–4258.

Blanco, R., S. Brennan, and I. W. March (2005). An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. *Journal of Finance 60 (5)*, 2255–2281.

Boehmer, E., J. Musumeci, and A. B. Poulsen (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics 30 (2)*, 253–272.

Brown, S. J. and J. B. Warner (1980). Measuring security price performance. *Journal of Financial Economics 8 (3)*, 205–258.

Brown, S. J. and J. B. Warner (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics 14 (1)*, 3–31.

Callen, J. L., J. Livant, and D. Segal (2009). The impact of earnings on the pricing of credit default swaps. *The Accounting Review 84 (5)*, 1363–1394.

Campbell, C. J., A. R. Cowan, and V. Salotti (2010). Multi-country event study methods. *Journal of Banking and Finance 34*, 3078–3090.

Charest, G. (1978). Dividend information, stock return, and market efficiency – ii. *Journal of Financial Economics 6*, 297–330.

Collin-Dufresne, P., R. S. Goldstein, and J. S. Martin (2001). The determinants of credit spread changes. *Journal of Finance 56 (6)*, 2177–2207.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance 1*, 223–236.

Corrado, C. J. (1989). A nonparametric test for abnormal security-price performance in event studies. *Journal of Financial Economics 23 (2)*, 385–396.

Corrado, C. J. (2011). Event studies: A methodology review. *Accounting and Finance 51*, 207–234.

Corrado, C. J. and T. L. Zivney (1992). The specification and power of the sign test in event study hypothesis tests using daily stock returns. *Journal of Financial and Quantitative Analysis 27 (3)*, 465–478.

Cowan, A. R. (1992). Nonparametric event study tests. *Review of Quantitative Finance and Accounting 2*, 343–358.

Daniels, K. N. and M. S. Jensen (2005). The effect of credit ratings on credit default swap spreads and credit spreads. *Journal of Fixed Income 15 (3)*, 16–33.

Dittmann, I., L. Norden, and G. Zhu (2012). Cds spreads, executives preferences, and risk taking: Evidence from ceo stock grants. *Working Paper*.

Ericsson, J., K. Jacobs, and R. Oviedo (2009). The determinants of credit spread changes. *Journal of Financial and Quantitative Analysis 44 (1)*, 109–132.

Galil, K. and G. Soffer (2011). Good news, bad news and rating announcements: An empirical investigation. *Journal of Banking and Finance 35 (11)*, 3101–3119.

Greatrex, C. A. (2008). The credit default swap market's reaction to earnings announcements. *Working Paper*.

Handjinicolaou, G. and A. Kalay (1984). Wealth redistributions or changes in firm value: An analysis of returns to bondholders and stockholders around dividend announcements. *Journal of Financial Economics 13*, 35–63.

Harrington, S. E. and D. G. Shrider (2007). All events induce variance: Analyzing abnormal returns when effects vary across firms. *Journal of Financial and Quantitative Analysis 42 (1)*, 229–256.

Higgins, E. J. and D. R. Peterson (1998). The power of one and two sample t-statistics given event-induced variance increases and nonnormal stock returns: A comparative study. *Quarterly Journal of Business and Economics 37 (1)*, 27–49.

Houweling, P. and T. Vorst (2005). Pricing default swaps: Empirical evidence. *Journal of International Money and Finance 24*, 1200–1225.

Hull, J., M. Predescu, and A. White (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking and Finance 28 (11)*, 2789–2811.

Jorion, P. and G. Zhang (2007). Good and bad credit contagion: Evidence from credit default swaps. *Journal of Financial Economics 84*, 860–883.

King, M. R. (2009). Time to buy or just buying time? the market reaction to bank rescue packages. *BIS Working Paper No. 288*, 1–34.

Lehnert, T. and F. Neske (2006). On the relationship between credit rating announcements and credit default swap spreads for european reference entities. *Journal of Credit Risk 2 (2)*, 83–90.

Longstaff, F. A., S. Mithal, and E. Neis (2005). Corporate yield spreads: Default risk or liquidity? new evidence from the credit default swap market. *Journal of Finance 60 (1)*, 2213–2253.

Loughran, T. and J. R. Ritter (1995). The new issue puzzle. *Journal of Finance 50*, 23–51.

MacKinlay, C. A. (1997). Event studies in economics and finance. *Journal of Economic Literature 35*, 13–39.

Mayordomo, S., J. I. Pena, and E. S. Schwartz (2010). Are all credit default swap databases equal? *NBER Working Paper*.

Micu, M., E. M. Remolona, and P. D. Wooldridge (2004). The price impact of rating announcements: evidence from the credit default swap market. *BIS Quarterly Review June*, 55–65.

Mikkelson, W. (1981). Convertible security calls and security returns. *Journal of Financial Economics 9*, 237–264.

Norden, L. and M. Weber (2004). Informational efficiency of credit default swap and stock markets: The impact of credit rating announcements. *Journal of Banking and Finance 28*, 2813 –2843.

Packer, F. and H. Zhu (2005). Contractual terms and cds pricing. *BIS Quarterly Review March*, 89–100.

Patell, J. (1976). Corporate forecasts of earnings per share and stock price behavior: Empirical tests. *Journal of Accounting Research 14 (2)*, 246–276.

Penman, S. (1982). Insider trading and the dissemination of firm forecast information. *Journal of Business 55*, 479–503.

Pop, A. and D. Pop (2009). Requiem for market discipline and the specter of tbtf in japanese banking. *Quarterly Review of Economics and Finance 49*, 1429–1459.

Posch, P. N. (2011). Time to change. rating changes and policy implications. *Journal of Economic Behavior & Organization 80 (3)*, 641–656.

Ritter, J. R. (1991). The long-run performance of initial public offerings. *Journal of Finance 46*, 3–27.

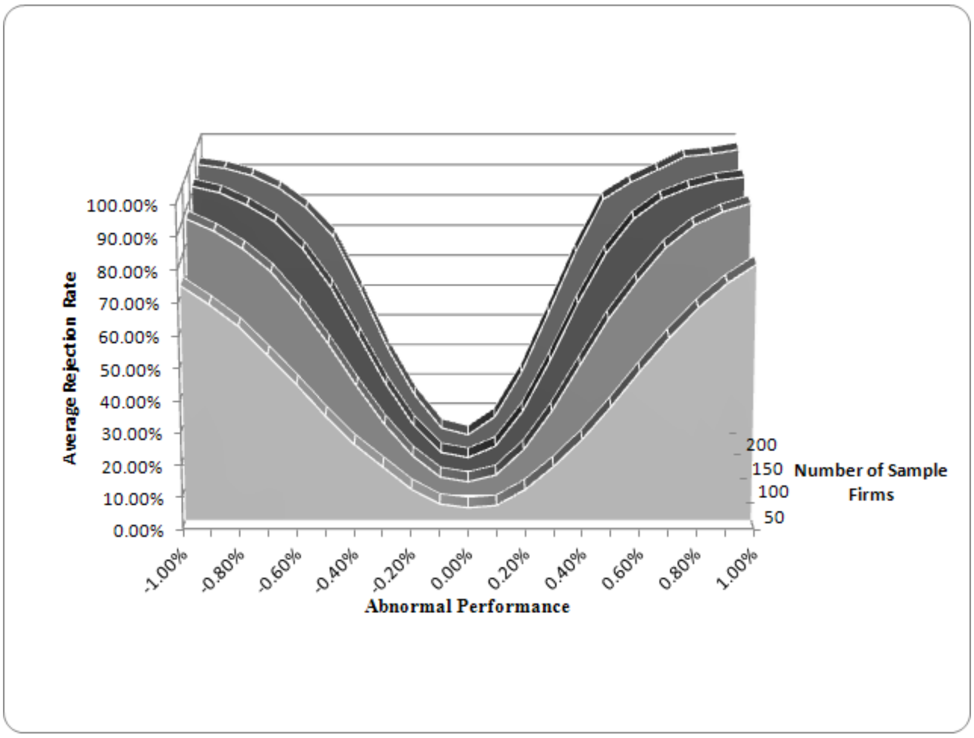
Rosenstein, J. and S. Wyatt (1990). Outside directors, board independence and shareholder wealth. *Journal of Financial Economics 26*, 175–192.

Sanders, R. W. and R. Robins (1991). Discriminating between wealth and information effects in event studies. *Review of Quantitative Finance and Accounting 1*, 307–330.

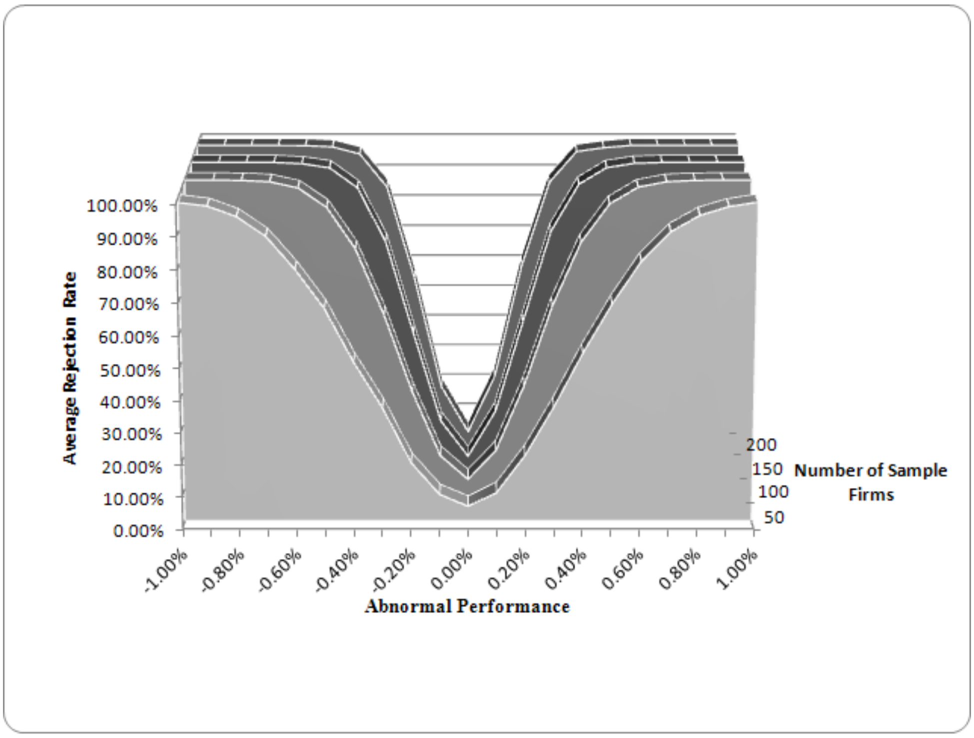
Shivakumar, L., O. Urcan, F. P. Vasvari, and L. Zhang (2011). The debt market relevance of management earnings forecasts: Evidence from before and during the credit crisis. *Review of Accounting Studies 16 (3)*, 464–486.

Zhu, H. (2006). An empirical comparison between the bond market and the credit default swap market. *Journal of Financial Services Research 29 (3)*, 211–265.

Figure 1: The power of tests for different sample sizes.



(a) The BMP (1991) test



(b) The rank test

Table 1: CDS Event Studies in the Literature

Study authors, year	No. of spread observations	Time period	Benchmark model	Spread change calculation	Test statistic	Data source
Hull et al. (2004)	29,032	Oct. 1998-May 2002	Index based on average CDS spread in rating category	Absolute	Bootstraped t-Test	GFI
Micu et al. (2004)	694 financial and non-financial firms	Jan. 2001-Dec. 2003	Index based on average CDS spread in rating category	Absolute	t-Test, sign test	Markit
Norden and Weber (2004)	60,827	July 1998-Dec. 2002	Index based on CDS spread in rating category	Absolute	Wilcoxon sign(rank) test	Unknown European bank
Daniels and Jensen (2005)	72 firms	Jan. 2002-Dec. 2002	Constant mean model	Absolute	t-Test	JP Morgan Chase
Lehnert and Neske (2006)	100 major European firm	Aug. 2000-Aug. 2003	Average daily percentage change of CDS spreads within a particular investment-grade category	Absolute	t-Test (not clearly stated)	Trac-X Europe index
Jorion and Zhang (2007)	512,292	Jan. 2003-Dec. 2004	Rating-adjusted CDS spread	Absolute	t-Test	Markit
Greatrex (2008)	413,844	Jan. 2001Apr. 2006	Index based on CDS spread in rating category Market (index) model	Absolute	Cross-sectional t-test	Markit
Callen et al. (2009)	20,328	2002-2005	None	Relative	None	Lombard Risk
King (2009)	28 banks	Jan. 2008–Jan. 2009	Multifactor model Market Model	Absolute	none	Markit
Pop and Pop (2009)	31 financial institutions	Jan. 2003-Jan. 2004	Constant mean model	Absolute	t-Test	Markit and Credit Market Analysis
Galil and Soffer (2011)	1,176,640	Jan. 2002-June 2006	Index based on average CDS spread in rating category	Absolute	t-Test Wilcoxon sign (rank) test	Markit
Shivakumar et al. (2011)	846,261	2001-2008	Matched basket of CDS contracts, selecting the CDS contracts with the same credit rating category	Relative	None	Markit
Dittmann et al. (2012)	108 firms	2000-2010	Index based on average CDS spread in rating category	Absolute	t-Test, sign test Wilcoxon sign (rank) test	Markit

Table 2: Factor Correlation

Panel A: North America

	ΔS_{Index}	$\Delta Level$	$\Delta Slope$	$\Delta Vola$	$\Delta Stock$
ΔS_{Index}	1.00				
$\Delta Level$	-0.14	1.00			
$\Delta Slope$	-0.01	0.10	1.00		
$\Delta Vola$	0.26	-0.25	0.01	1.00	
$\Delta Stock$	-0.29	0.32	0.01	-0.74	1.00

Panel B: Europe

	ΔS_{Index}	$\Delta Level$	$\Delta Slope$	$\Delta Vola$	$\Delta Stock$
ΔS_{Index}	1.00				
$\Delta Level$	-0.17	1.00			
$\Delta Slope$	0.03	0.16	1.00		
$\Delta Vola$	0.36	-0.25	-0.02	1.00	
$\Delta Stock$	-0.41	0.36	0.02	-0.75	1.00

Table 3: Variance Inflation Factors

Panel A: All Variables

	ΔS_{Index}	$\Delta Level$	$\Delta Slope$	$\Delta Vola$	$\Delta Stock$
North America	3.10	1.14	1.01	2.23	4.01
Europe	3.22	1.18	1.03	2.33	5.04

Panel B: Factor Model Variables

	ΔS_{Index}	$\Delta Level$	$\Delta Slope$	$\Delta Vola$
North America	1.08	1.09	1.01	1.14
Europe	1.17	1.11	1.03	1.20

Table 4: Descriptive Statistics (Realized Spread Changes)

Panel A: ***Statistics of Relative CDS Spread Changes (%), 0.01 = 1%***

North America

	Number of observations	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads	Zero spreads
All	2,091,244	0.0002	0.0000	0.0466	1.1860	324	38.4%	22.9%
Investment grade	1,352,614	0.0002	0.0000	0.0158	1.5441	219	39.4%	21.8%
Non-investment grade	738,630	0.0003	0.0000	0.0480	0.6150	480	37.8%	24.9%

Europe

	Number of observations	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads	Zero spreads
All	1,083,294	0.0004	0.0000	0.0461	1.8442	500	39.3%	19.4%
Investment grade	922,296	0.0004	0.0000	0.0454	1.7884	183	39.3%	19.7%
Non-investment grade	160.998	0.0003	0.0000	0.0495	2.0791	1,786	39.0%	18.0%

Panel B: ***Descriptive Statistics of Absolute CDS Spread Changes (basis points (bps)), 0.01 = 1 bps***

North America

	Number of observations	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads	Zero spreads
All	2,091,244	0.0032	0.0000	0.9520	25.4	18,116	38.4%	22.9%
Investment grade	1,352,614	0.0002	0.0000	0.1570	6.4	5,621	39.4%	21.8%
Non-investment grade	738,630	0.0087	0.0000	1.5884	15.5	6,627	37.8%	24.9%

Europe

	Number of observations	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads	Zero spreads
All	1,083,294	0.0011	0.0000	0.4336	184.1	167,924	39.3%	19.4%
Investment grade	922,296	0.0008	0.0000	0.1659	347.2	226,328	39.3%	19.7%
Non-investment grade	160,998	0.0034	0.0000	1.0600	78.3	31,309	39.5%	18.0%

Notes: This table provides summary statistics for daily spread changes for the North American and European samples, as well as for subsamples by issuer rating. Daily CDS spread data are from the daily CDS database of Markit Ltd. and cover the period from February 11, 2003, to July 31, 2011. Panel A reports the results for the relative spread change measure. Panel B reports the results for the absolute spread change measure.

Table 5: Descriptive Statistics (Abnormal Spread Changes)

Panel A: ***Statistics of Relative CDS Spread Changes* (%), 0.01 = 1%**

North America

	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads
Mean-adjusted spreads	-0.0001	-0.0003	0.0379	1.1	111	48.6%
Market model	0.0002	-0.0001	0.0357	1.1	139	49.3%
CDS factor model	0.0001	-0.0002	0.0629	1.0	59	49.2%
Portfolios (rating)	-0.0002	-0.0003	0.0491	1.4	227	49.4%

Europe

	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads
Mean-adjusted spreads	-0.0001	-0.0003	0.0378	0.6	85	48.5%
Market model	-0.0003	-0.0003	0.0323	1.2	162	48.7%
CDS factor model	-0.0001	-0.0002	0.0547	1.1	62	48.9%
Portfolios (rating)	-0.0001	-0.0002	0.0493	1.3	267	49.3%

Panel B: ***Descriptive Statistics of Absolute CDS Spread Changes* (basis points (bps)), 0.01 = 1 bps**

North America

	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads
Mean-adjusted spreads	0.0044	0.0001	1.15	2.2	10,528	50.6%
Market model	0.0047	0.0001	1.16	1.9	10,657	50.8%
CDS factor model	0.0012	0.0002	3.59	1.8	5,373	50.5%
Portfolios (rating)	0.0056	0.0000	1.21	1.6	9,486	48.8%

Europe

	Mean	Median	Standard deviation	Skewness	Excess kurtosis	Positive spreads
Mean-adjusted spreads	0.0017	0.0000	0.50	343.2	163,039	50.2%
Market model	0.0009	-0.0001	0.49	351.6	168,677	49.5%
CDS factor model	0.0007	0.0000	1.12	334.9	81,364	50.2%
Portfolios (rating)	0.0024	0.0000	0.52	303.4	139,298	47.7%

Notes: This table provides summary statistics for daily relative abnormal spread change for each model of abnormal spread changes. The results are based on 5,000 replications of 200 randomly drawn CDS contract–event day combinations. We do not add any abnormal performance. Panel A reports the results for the relative spread change measure. Panel B reports the results for the absolute spread change measure.

Table 6: Size of Tests (Relative Spread Changes)

Panel A: North America

Lower tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	1.98%*	2.08%	1.60%*	2.16%	2.46%
Market model	2.06%*	1.56%*	1.48%*	2.02%*	3.06%*
CDS factor model	2.76%	1.96%*	1.68%*	2.80%	2.86%
Portfolios (rating)	2.42%	2.66%	3.26%*	2.50%	6.98%*

Upper tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	3.10%*	2.32%	2.76%	2.70%	1.86%*
Market model	3.64%*	2.28%	2.90%	2.82%	1.70%*
CDS factor model	5.02%*	2.38%	2.82%	2.78%	2.38%
Portfolios (rating)	3.12%*	1.66%*	1.60%*	2.22%	0.68%*

Panel B: Europe

Lower tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	2.24%	2.84%	1.38%*	2.10%	3.14%*
Market model	3.46%*	3.80%*	2.88%	3.06%*	4.94%*
CDS factor model	2.82%	2.32%	1.44%*	2.28%	3.26%*
Portfolios (rating)	2.24%	2.84%	1.38%*	2.10%	4.14%*

Upper tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	2.18%	1.92%*	3.02%*	2.42%	1.60%*
Market model	2.42%	0.98%*	1.64%*	1.80%*	1.14%*
CDS factor model	2.64%	2.16%	2.50%	2.14%	1.88%*
Portfolios (rating)	2.18%	1.92%*	3.02%*	2.42%	1.60%*

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics ($\alpha = 2.5\%$, one-tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract-event day combinations. For the normal approximation of the 5,000 binomial trials the average rejection rate should be between 2.07% and 2.93% (at the 95% confidence interval).

Table 7: Power of Tests (Relative Spread Changes)

Panel A: North America

Relative CDS Spread Changes: +0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	48.44%	56.84%	75.50%	99.74%	98.88%
Market model	56.82%	63.88%	80.89%	99.70%	98.68%
CDS factor model	59.70%	64.26%	82.53%	99.26%	96.60%
Portfolios (rating)	25.73%	40.88%	55.91%	98.33%	93.66%

Relative CDS Spread Changes: -0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	43.86%	51.82%	61.08%	99.36%	99.30%
Market model	50.14%	56.40%	68.60%	99.46%	99.60%
CDS factor model	51.01%	58.58%	70.35%	99.53%	99.68%
Portfolios (rating)	26.96%	42.22%	57.76%	98.72%	99.86%

Panel B: Europe

Relative CDS Spread Changes: +0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	43.92%	50.28%	74.66%	99.58%	98.26%
Market model	59.52%	63.26%	79.98%	99.54%	97.96%
CDS factor model	60.03%	64.89%	81.06%	99.62%	98.82%
Portfolios (rating)	27.32%	40.26%	55.36%	98.18%	93.38%

Relative CDS Spread Changes: -0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	47.20%	52.42%	60.10%	99.56%	99.48%
Market model	69.14%	70.88%	77.84%	99.76%	99.80%
CDS factor model	70.34%	71.84%	78.65%	99.85%	99.78%
Portfolios (rating)	29.90%	44.54%	59.44%	98.34%	99.78%

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics ($\alpha = 5\%$, two tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract-event day combinations. We add abnormal spread changes (relative) at day zero of 0.5% and -0.5%.

Table 8: Event-Induced Variance and Size of Tests

Panel A: ***Event-Induced Variance Proportional to Individual Security Estimation Window Variance***

Lower tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	2.76%	1.96%*	1.68%*	2.80%	2.86%
k=0.5	3.88%*	2.96%*	2.86%	4.28%*	2.16%
k=1.0	9.88%*	3.04%*	2.92%	5.94%*	2.19%
k=1.5	19.73%*	3.64%*	3.41%*	7.30%*	2.89%

Upper tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	3.02%*	2.38%	2.82%	2.78%	2.38%
k=0.5	3.75%*	2.30%	2.04%*	3.84%*	2.74%
k=1.0	6.80%*	2.98%*	2.41%	7.30%*	3.35%*
k=1.5	11.00%*	3.64%*	3.95%*	8.53%*	3.61%*

Panel B: ***Event-Induced Variance Proportional to Average Estimation Window Variance***

Lower tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	2.76%	1.96%*	1.68%*	2.80%	2.86%
k=0.5	3.59%*	1.91%*	1.86%*	4.17%*	2.14%
k=1.0	9.13%*	2.90%	2.23%	6.29%*	2.97%*
k=1.5	19.56%*	4.11%*	2.49%	7.61%*	3.04%*

Upper tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	3.02%*	2.38%	2.82%	2.78%	2.38%
k=0.5	4.41%*	2.38%	2.84%	6.03%*	2.67%
k=1.0	7.01%*	2.29%	2.86%	8.06%*	3.33%*
k=1.5	9.12%*	2.85%	2.89%	9.67%*	3.56%*

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics (α = 2.5%, one-tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract–event day combinations. For the normal approximation of the 5,000 binomial trials the average rejection rate should be between 2.07% and 2.93% (at the 95% confidence interval).

Table 9: Event-Induced Variance and Power of Tests

Panel A: ***Event-Induced Variance Proportional to Individual Security Estimation Window Variance***

Relative CDS Spread Changes: +0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	59.70%	64.26%	82.53%	99.26%	96.60%
k=0.5	49.72%	48.32%	69.75%	92.42%	80.69%
k=1.0	44.56%	27.28%	50.86%	74.54%	53.14%
k=1.5	40.64%	12.72%	33.54%	58.14%	32.87%

Relative CDS Spread Changes: -0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	51.01%	58.58%	70.35%	99.53%	99.68%
k=0.5	54.58%	51.00%	72.98%	95.10%	82.86%
k=1.0	48.80%	29.18%	54.38%	78.08%	53.50%
k=1.5	42.68%	15.18%	34.74%	60.90%	35.76%

Panel B: ***Event-Induced Variance Proportional to Average Estimation Window Variance***

Relative CDS Spread Changes: +0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	59.70%	64.26%	82.53%	99.26%	96.60%
k=0.5	53.07%	46.38%	55.65%	82.65%	58.32%
k=1.0	44.02%	25.09%	26.90%	52.88%	28.30%
k=1.5	41.33%	12.67%	13.70%	35.97%	16.47%

Relative CDS Spread Changes: -0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
k=0.0	51.01%	58.58%	70.35%	99.53%	99.68%
k=0.5	54.58%	51.00%	72.98%	95.10%	82.86%
k=1.0	48.80%	29.18%	54.38%	78.08%	53.50%
k=1.5	42.68%	15.18%	34.74%	60.90%	35.76%

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics (α = 5%, two tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract–event day combinations. We add abnormal spread changes (relative) at day zero of 0.5% and -0.5%.

Table 10: Bootstrapping and Size of Tests

Panel A: North America

Lower tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	2.46%	2.49%	2.42%
Market model	2.62%	2.21%	2.62%
CDS factor model	2.58%	2.36%	2.59%
Portfolios (rating)	2.43%	2.26%	3.16%*

Upper tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	2.46%	2.51%	2.68%
Market model	2.62%	1.96%*	2.62%
CDS factor model	2.58%	2.13%	2.63%
Portfolios (rating)	2.86%	2.16%	1.73%*

Panel B: Europe

Lower tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	2.51%	3.08%*	2.41%
Market model	2.54%	2.12%	2.56%
CDS factor model	2.58%	2.76%	2.44%
Portfolios (rating)	2.40%	2.60%	2.16%

Upper tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	2.50%	2.96%*	2.64%
Market model	2.51%	2.58%	2.61%
CDS factor model	2.54%	2.23%	2.53%
Portfolios (rating)	2.40%	1.96%*	2.60%

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics ($\alpha = 2.5\%$, one-tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract–event day combinations. For the normal approximation of the 5,000 binomial trials the average rejection rate should be between 2.07% and 2.93% (at the 95% confidence interval).

Table 11: Bootstrapping and Power of Tests

Panel A: North America

Relative CDS Spread Changes: +0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	47.48%	56.42%	74.64%
Market model	64.57%	61.76%	75.81%
CDS factor model	66.00%	63.67%	76.51%
Portfolios (rating)	24.77%	41.27%	56.29%

Relative CDS Spread Changes: -0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	43.68%	51.32%	61.36%
Market model	63.19%	60.55%	76.52%
CDS factor model	65.99%	63.86%	75.91%
Portfolios (rating)	27.13%	41.74%	57.70%

Panel B: Europe

Relative CDS Spread Changes: +0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	50.45%	47.39%	72.01%
Market model	63.28%	67.43%	80.95%
CDS factor model	65.02%	68.47%	81.42%
Portfolios (rating)	27.74%	40.77%	53.98%

Relative CDS Spread Changes: -0.50%

	t-Test (time-series)	t-Test (cross-section)	BMP
Mean-adjusted spreads	54.88%	52.99%	62.14%
Market model	64.72%	68.99%	79.66%
CDS factor model	66.92%	69.38%	81.29%
Portfolios (rating)	30.27%	44.24%	59.86%

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics ($\alpha = 5\%$, two tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract-event day combinations. We add abnormal spread changes (relative) at day zero of 0.5% and -0.5%.

A Absolute Spreads

Table 12: Size of Tests (Absolute Spread Changes)

Lower tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	4.50%*	0.76%*	0.42%*	1.06%*	0.52%*
Market model	4.68%*	0.78%*	0.72%*	1.98%*	1.44%*
CDS factor model	2.94%*	4.46%*	1.06%*	2.20%	2.50%
Portfolios (rating)	4.64%*	1.36%*	3.68%*	3.94%*	7.90%*

Upper tail (2.5%)

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	7.28%*	1.68%*	6.78%*	4.04%*	7.58%*
Market model	7.46%*	1.72%*	4.08%*	3.30%*	3.46%*
CDS factor model	4.04%*	1.74%*	4.10%*	3.80%*	2.66%
Portfolios (rating)	6.18%*	0.84%*	1.45%*	1.50%*	0.60%*

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics (α = 2.5%, one-tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract-event day combinations. For the normal approximation of the 5,000 binomial trials the average rejection rate should be between 2.07% and 2.93% (at the 95% confidence interval).

Table 13: Power of Tests (Absolute Spread Changes)

Absolute CDS Spread Changes: +0.5 bps

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	8.28%	5.50%	87.02%	100.00%	100.00%
Market model	8.78%	5.72%	87.00%	100.00%	100.00%
CDS factor model	4.36%	6.34%	88.60%	100.00%	100.00%
Portfolios (rating)	7.04%	4.08%	86.80%	100.00%	99.70%

Absolute CDS Spread Changes: -0.5 bps

	t-Test (time-series)	t-Test (cross-section)	BMP	Rank	Generalized sign
Mean-adjusted spreads	5.38%	2.98%	66.10%	100.00%	99.86%
Market model	5.66%	3.00%	73.40%	100.00%	99.96%
CDS factor model	5.86%	3.12%	74.58%	100.00%	99.98%
Portfolios (rating)	5.46%	6.54%	82.98%	100.00%	100.00%

Notes: This table reports the day zero rejection rates of the null hypothesis of no abnormal performance for different test statistics ($\alpha = 5\%$, two tailed). The results are based on 5,000 replications of 200 randomly drawn CDS contract–event day combinations. We add abnormal performance at day zero of +0.5 and -0.5 bps.

B Test Statistics

Besides the selection of an appropriate spread change measure and a model of normal spread changes, a well-specified test statistic must be defined. The following discusses the adaptation of several parametric as well as non-parametric test statistics used in classical event studies to the case of CDS spread changes. All parametric test statistics are built upon the same null hypothesis. Under the validity of the null hypothesis, the average abnormal spread change on the event day should be equal to zero:

$$H_0 : \Delta AAS_0 = \frac{1}{N}\sum_{i=1}^{N} \Delta AS_{i,0} = 0 \tag{11}$$

If the event of interest induces an abnormal spread change that is significantly different from zero, the null hypothesis should be rejected. Since the direction of the event-induced spread change is unknown (ex ante) in most applications, it is common to apply two-tailed tests. A simple t-test can be derived from the assumption that abnormal spread changes are independent and identically normally distributed. The test statistic equals the quotient of the average abnormal spread change and its estimated standard deviation:

$$t = \frac{\Delta AAS_0}{S(\Delta AAS_0)} \tag{12}$$

The statistic follows a t-distribution with N - D degrees of freedom, where N denotes the number of events and D the number of parameters of the normal spread change model. The distribution of the test statistic is asymptotically normal in sufficiently large samples. The estimated standard error is based on the estimation window observations.[12]

This simple t-test may lead to biased inferences whenever an event induces a variance increase in the spread changes. An estimation of the standard deviation based on the estimation window observations is most likely downward biased in that case. Thus, a valid null hypothesis is rejected too often. Brown and Warner (1985) propose to estimate the standard deviation from the cross section of the event window. For the case of CDSs, the cross-sectional test is accordingly defined as

$$t_{CS} = \frac{\Delta AAS_0}{S_{CS}(\Delta AAS_0)} \tag{13}$$

$$S_{CS}(\Delta AAS_0) = \sqrt{\frac{1}{N-1}\left(\sum_{i=1}^{N} \Delta AS_{i,0} - \Delta AAS_0\right)^2} \tag{14}$$

[12]Since the standard deviation is estimated from the estimation window observations, the estimator should be adjusted for forecast errors. The adjustment depends on the model of normal spread changes. We calculate all test statistics with the appropriate forecast error adjustments within our simulation.

Another parametric test is the standardized residuals test introduced by Patell (1976) . The major difference from the previous test is the standardization of abnormal spread changes before calculating the average value:

$$\Delta AS^*_{i,0} = \frac{\Delta AS_{i,0}}{S(\Delta AS_i)} \tag{15}$$

$$t_P = \frac{1}{N} \sum_{i=1}^{N} \Delta AS^*_{i,0} \sqrt{\frac{N(D-4)}{D-2}} \tag{16}$$

where $\Delta AS^*_{i,0}$ denotes the standardized abnormal spread changes, N is the number of events, and D is the number of parameters of the normal spread change model. The standardization leads to an equal weight for each event. Since the Patell (1976) test is not well specified for an event-induced variance increase, we apply the test of BMP (1991). This test is a hybrid of the Patell (1976) and cross-sectional tests that is robust to event-induced variance increases. The test statistics can be constructed by applying the cross-sectional test to standardized abnormal spread changes:

$$t_{BMP} = \frac{\Delta AAS^*_0}{S_{CS}(\Delta AAS^*_0)} \tag{17}$$

$$S_{CS}(\Delta AAS^*_0) = \sqrt{\frac{1}{N-1} \left(\sum_{i=1}^{N} \Delta AS^*_{i,0} - \Delta AAS^*_0 \right)^2} \tag{18}$$

where ΔAAS^*_0 denotes the average of the standardized abnormal spread changes on the event day.

All aforementioned tests rely on the assumption that abnormal spread changes are normally distributed. However, this assumption does not seem feasible when a distribution exhibits substantial skewness and/or excess kurtosis. In that case non-parametric test statistics can be used instead. The Corrado (1989) rank test and the generalized sign test proposed by Cowan (1992) are commonly used in event studies.

The rank test is based on the transformation of abnormal spread changes into ranks for each time series:

$$R_{i,\tau} = rank\left[AS_{i,\tau}\right] \tag{19}$$

Tied ranks should be treated by the method of midranks according to Corrado (1989). We further correct for missing observations, as proposed by Corrado and Zivney (1992), based on a uniform transformation of ranks:

$$U_{i,\tau} = \frac{R_{i,\tau}}{1+M_i} \tag{20}$$

where M_i denotes the number of non-missing observations for time series i. Under the validity of the null hypothesis, the average of the transformed rank should not deviate significantly from

0.5. Based upon this assumption, the test statistics is defined as

$$t_{Rank} = \frac{\frac{1}{N}\sum_{i=1}^{N} U_{i,\tau} - 0.5}{S(U)} \tag{21}$$

$$S(U) = \frac{1}{T}\sum_{\tau}\frac{1}{N^2}\sum_{i=1}^{N}[U_{i,\tau} - 0.5]^2 \tag{22}$$

The generalized sign test is derived from the proportion of positive and negative abnormal spread changes on the event date. Under the validity of the null hypothesis, this proportion should not differ from the proportion of positive and negative abnormal spread changes of the estimation windows:

$$t_{Sign} = \frac{p_0^+ - p_{est}^+}{\sqrt{\frac{1}{N}p_{est}^+(1 - p_{est}^+)}} \tag{23}$$

where p_0^+ and p_{est}^+ are the percentages of positive abnormal spread changes on the event date and in the estimation window, respectively.