Journal of Banking and Finance 90 (2018) 146–158

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Detecting abnormal changes in credit default swap spreads using matching-portfolio models[☆]

Fabio Bertoni[a], Stefano Lugo[b,*]

[a] EMLYON Business School. Department of Economics, Finance and Control Avenue Guy de Collongue 22, Ecully 69134, France
[b] Utrecht University, Utrecht School of Economics, Kriekenpitplein 21–22, EC, Utrecht 3584, Netherlands

ARTICLE INFO

Article history:
Received 27 May 2016
Accepted 29 March 2018
Available online 31 March 2018

JEL classification:
G10
G14

Keywords:
Event studies
Credit default swaps
Matching-portfolio models
Size and power of tests

ABSTRACT

We evaluate the size and power of different statistical tests and adjustment methods for matching-portfolio models to detect abnormal changes in credit default swap (CDS) spreads. The sign-test generally dominates the signed-rank test in terms of size, and dominates both the *t*-test and the signed-rank test in terms of power. Traditional adjustment methods often lead to a misspecified sign-test. We propose a new and parsimonious method (the spread-matched method), which leads to a well-specified and more powerful sign-test. The superiority of the spread-matched method is particularly evident for observations characterized by extreme levels of CDS spread. Analyses of CDS samples differing by contract maturity, data source, and time period confirm these results. We perform an event study on rating downgrades to illustrate how the choice of tests and adjustment methods can affect inference.

1. Introduction

In recent years, credit default swap (CDS) spreads have become increasingly popular among academics, practitioners, and regulators for measuring credit risk (for a general survey of studies on CDS contracts, see Augustin et al. 2014), and are now a serious alternative to bond yields for conducting event studies on credit risk. CDS spreads present a series of advantages over bond yields. Unlike bond yields, which require crucial assumptions about the benchmark interest curve, CDS spreads are already a direct measure of credit premiums (Hull et al., 2004). CDS spread data refer to new, fixed-maturity contracts issued every day, whereas bond yield data refer to outstanding bonds whose time to maturity naturally evolves over time. Moreover, compared to bond yields, CDS spreads respond more quickly to changes in credit conditions (e.g., Blanco et al., 2005; Zhu, 2006) and are less affected by liquidity risk (Longstaff et al., 2005).

CDS spreads are therefore particularly well-suited for investigating the impact of specific events on credit risk. Event studies on CDS spreads can be conducted using two main classes of model: factor models (King, 2009; Shivakumar et al., 2011) and matching-portfolio models. The literature lacks an in-depth analysis of this latter class of model, and our paper aims to fill this gap. Specifically, we compare the size (i.e., Type I error) and power (i.e., Type II error) of the statistical tests used to detect Abnormal CDS Spread Changes (*ASCs*) computed with different adjustment methods.

A better understanding of matching-portfolio models is needed for two main reasons. First, matching-portfolio models can be used on larger samples of observations and are less exposed to sample biases than factor models. While matching-portfolio models only require available observations around the event, factor models require available observations over a longer time window, which reduces the number of usable events. As illustrated in Section 2, a lower bound for the loss of usable observations in factor models is between 10% and 20% (depending on the criteria for inclusion), although the actual number could be larger due to missing data in factor proxies and because factor models require event-free estimation windows (Afonso et al., 2012). More importantly, sample selection in factor models is not random; we show that the selection criteria entailed by factor models result in samples that are systematically biased towards companies with lower CDS spreads. For these reasons, matching-portfolio models will most likely continue to be used, at least as a robustness check.

A second reason why a better understanding of matching-portfolio models is needed is that these models are very common in the existing empirical literature. A large number (and, as of today, the vast majority) of the published articles that conduct event

☆ We would like to thank Geert Bekaert (the Editor) and two anonymous referees for their helpful suggestions.

* Corresponding author.
E-mail addresses: bertoni@em-lyon.com (F. Bertoni), s.lugo@uu.nl (S. Lugo).

https://doi.org/10.1016/j.jbankfin.2018.03.009
0378-4266/

Table 1
Summary of event studies using portfolio-matched methods on CDS spreads. This table presents a survey of the existing literature performing event studies using CDS data and a matching-portfolio model. We only include studies published as of December 2016 that focus on the variation of CDS spreads around specific, observable events. No Obs refers to the minimum and maximum number of events included in the analysis. Adjustment methods used to compute Abnormal CDS spread Changes are defined as in Section 3.2.

Authors	Period	Object of study	No Obs	Adjustment methods	Database
Afonso et al. (2012)	*-2010	Rating changes	167	All; Unadj.	CMA/Thomson
Batta et al. (2016)	2001–2010	Earnings announcements	*	All	Markit
Bedendo and Colla (2015)	2008–2011	Sovereign rating changes	44–119	All; Unadj.	Markit
Bertoni and Lugo (2014)	2003–2010	SWF investments	96–391	Rating(4); All	CMA
Callen et al. (2009)	2002–2005	Earnings announcements	383	Unadj.	Lombard risk
Finnerty et al. (2013)	2001–2009	Rating changes	2–934	Rating(6)	Markit
Galil and Soffer (2011)	2002–2006	Rating changes	14–978	Rating(6)	Markit
Horvath and Huizinga (2015)	2010	Introduction EFS Facility	19–44	Unadjusted	Thomson
Hull et al. (2004)	1998–2002	Rating changes	7–114	Rating(3); Unadj.	GFI
Ismailescu and Kazemi (2010)	2001–2009	Sovereign rating changes	57–87	All; Unadj.	Markit
Jorion and Zhang (2007)	2001–2004	Intra-sector contagion	22–272	Rating(5); Unadj.	Markit
Jorion and Zhang (2009)	2001–2005	Creditor contagion	30–128	Rating(2)	Markit
Lehnert and Neske (2006)	2000–2003	Rating changes	8–70	Rating(4)	JP Morgan
Norden and Weber (2004)	1998–2002	Rating changes	24–63	Rating(4)	Undisclosed
Pop and Pop (2009)	2003	Banks bailouts externalities	16–21	Unadj.	CMA; Markit
Wengner et al. (2015)	2004–2011	Rating changes	38–1071	Rating(6); All	Bloomberg
Zhang and Zhang (2013)	2001–2005	Earnings announcements	108–4005	Unadjusted	Markit

studies on CDS spreads employ matching-portfolio models. Table 1 presents a survey of the published event studies on credit risk that use CDS spreads and a matching-portfolio model. It is important to understand whether these studies are based on tests and methods that could potentially result in misspecification (understating the risk of rejecting a true null hypothesis) or that have limited power (having limited ability to reject a false null hypothesis).

The distributional properties of *ASC*s suggest that the tests and adjustment methods currently used in the literature may not be adequate. Regardless of the adjustment method used to compute them, *ASC*s are extremely leptokurtic. We would therefore expect the *t*-test, which is the most commonly used statistical test in the literature to date, to have very limited power and to be generally dominated by non-parametric tests.[1] Traditional adjustment methods also result in a highly skewed distribution of *ASC*s and a high risk of misspecification. Even when adjusting with a rating-matched method (the method most commonly used in the literature), the pre-event level of CDS spreads can differ substantially between the focal firm and the matched portfolio, because ratings are sticky (Altman and Rijken, 2004) and adjust slowly to changes in credit risk (Norden and Weber, 2004). As credit spreads are mean-reverting (Zhu, 2006), traditional adjustment methods result in a violation of the common trend assumption: *ASC*s are systematically biased upward (downward) for events that occur in companies with relatively low (high) CDS spreads.

We therefore propose a new adjustment method, the spread-matched method, which is based on a matched portfolio including companies whose pre-event CDS spreads are similar to that of the focal company. The spread-matched method results in a more symmetric distribution of *ASC*s compared to traditional methods; the cross-sectional median *ASC* is closer to zero on any trading day when computed with the spread-matched method; by minimizing the difference in pre-event levels of CDS spread between the focal firm and the matched portfolio, the spread-matched method also eliminates the systematic bias in *ASCs* for firms characterized by extreme levels of spreads. For these reasons, we predict a significantly lower risk of (especially) Type I and Type II errors when *ASC*s are computed with the spread-matched method compared to traditional adjustment methods.

[1] A similar result is found by Bessembinder et al. (2009) in the context of bond returns.

To address the size and power of different statistical tests on *ASC*s computed with different adjustment methods, we use the standard approach introduced by Brown and Warner (1980, 1985) in the context of financial data. Our two main results are consistent with our expectations. First, we find that – regardless of the adjustment method used – the *t*-test is very weak compared to non-parametric tests. Among the latter, the sign-test is always more powerful, and often better specified, than the Wilcoxon signed-rank test. The sign-test should thus be the preferred statistical test when addressing *ASC*s.

Second, we find that the sign-test is always well-specified when, and only when, used in combination with the spread-matched adjustment method. Traditional adjustment methods lead to severe overrejection of a true null hypothesis, especially when observations are sampled from the tails of the CDS spread distribution. In most sampling situations the spread-matched method also results in a more powerful sign-test compared to other adjustment methods (conditional on the test being well-specified). Several robustness checks, such as using a different CDS maturity or CDS data provider, confirm our main results.

To illustrate how the use of different statistical tests and adjustment methods can affect inference, we also conduct a real event study on credit rating downgrades. Using a sign-test in combination with the spread-matched method, the preferred choice according to our results, we conclude that downgrade announcements are indeed associated with abnormal increases in CDS spreads.

Our paper contributes to the literature addressing the adequacy of the statistical tests used to identify event-induced abnormal variations in financial data, such as short-term (Brown and Warner, 1980; 1985) and long-term (Barber and Lyon, 1997; Lyon et al., 1999) stock returns, firms' operative performance (Barber and Lyon, 1996), and bond returns (Bessembinder et al., 2009; Ederington et al., 2015). In particular, Bessembinder et al. (2009) show that a matching-portfolio model based on ratings (and maturity) generally results in well-specified and powerful tests on bond abnormal returns, especially when using daily prices. Our findings indicate that this approach is not well-suited to *ASC*s, despite the fact that several event studies in the literature have implicitly assumed as such. Andres et al. (2016) also address event study methodologies using CDS data, but compare different model approaches considering only one variant per model. We focus specifically on different adjustment methods for matching-portfolio models and propose a new method–the spread-matched method–which we find to be a significant improvement compared to the adjustment methods used in the literature to date.

The rest of this paper proceeds as follows. Section 2 describes the dataset used. In Section 3, we set out the different methods for computing abnormal spread changes and the approach used to address the size and power of statistical tests. Our main results are presented in Section 4. In Section 5, we illustrate the use of the different methods and tests in an event study on *ASC*s following credit rating downgrades. We report additional results about size and power for alternative specifications and samples in Section 6. Section 7 concludes.

2. Data

The main analyses presented in this paper are performed on CDS data from Markit, retrieved from Thomson Eikon. As Table 1 illustrates, Markit is the most commonly used data source in CDS-based event studies. Markit data represent composite end-of-day spreads based on the contributions of more than 30 major market participants from the sell-side. Other data providers, such as Credit Market Analysis (CMA),[2] collect their quotes from qualified members of the buy-side community. These differences in data source can have a material impact on spread measurement. Qiu and Yu (2012), for example, show that the number of sell-side quote providers for a single-name CDS in Markit is positively related to the amount of informed trading; while Mayordomo et al. (2013) provide evidence that systematic differences exist between spreads obtained from different data providers. As a robustness check, we therefore also use CMA data for our analyses. These additional results are discussed in Section 6.2.

The Markit dataset spans the period from July 2009 to February 2017. We gather data on 5-year and 10-year CDS spreads on senior unsecured debt. In our main analysis we focus on 5-year maturity contracts, which is the maturity most often considered in the literature. Results obtained for 10-year CDS spreads are presented in Section 6.1. We exclude entities classified as governments or public administrations. For firms with multiple CDS contracts for the same maturity (e.g., in different currencies), we retain the contract characterized by the greatest number of non-missing values. After applying these initial filters, the dataset includes 1,536,598 daily observations with non-missing, 5-year CDS spreads on 1417 firms from 52 different countries. In order to compute *ASC*s we need two consecutive data-points. It is possible to compute daily changes in spreads for 1,445,328 firm-day observations on 1390 firms.

For each firm we consider the domestic rating assigned by S&P, which is the most commonly available rating. We are able to identify a rating history for 1119 of the 1390 firms in our sample. Results obtained by re-integrating unrated firms are discussed in Section 6.3. Finally, we only retain observations where all of the methods considered for computing *ASC*, described in Section 3.2, are applicable. Our final dataset includes 1,233,057 firm-day observations.

It is interesting to compare these figures to the number of events that would potentially be usable if a factor model were instead employed to calculate *ASC*. Factor models require CDS data to be available for a longer time window than matching-portfolio models, which results in fewer usable observations even if we assume that all variables used to calculate factor weights are always available. In order to estimate the severity of this sample selection we consider, as in Andres et al. (2016), a factor model with an estimation window spanning between 150 and 20 days before the event. We only retain those events for which the number of data-points in the observation window is above the 50% or 80% threshold. At the 50% threshold, approximately 9.5% of the observations that are usable in an event study conducted using a matching-portfolio model are excluded from the analysis if a factor model is used. At the 80% threshold, the proportion of observations that are excluded increases to about 20%. The use of a factor model may therefore imply a sizable loss of events from the sample. It is also important to note that these estimates are lower bounds, and that the actual loss in usable observations could be even larger because of missing data in factor proxies and because of the elimination of events for which the estimation window is not event-free (Afonso et al., 2012).

More importantly, the sample selection imposed by factor models is not random: observations that are excluded from factor models are characterized by significantly higher levels of CDS spread. Using the 50% threshold, the average CDS spread of observations included in the sample would be 199.77 basis points (bps), which is substantially less than the average of 246.14 bps for the observations excluded from the sample. The difference is significant at customary confidence levels. It is hard, and beyond the scope of this paper, to determine the extent to which this sample selection bias may affect the results of studies based on factor models. However, the fact that such a bias exists suggests that studies applying factor models to CDS spreads should verify the robustness of their results to sample selection, and the use of a matching-portfolio model is an obvious way to do that.

3. Methodology

3.1. Assessing the size and power of statistical tests

As it is customary in research addressing the size and power of statistical tests for event studies (e.g., Barber and Lyon, 1996; 1997; Bessembinder et al., 2009; Edderington et al., 2015), we follow the method popularized by Brown and Warner (1980); Brown and Warner (1985)) for financial data. Following the approach of Bessembinder et al. (2009), we draw 5000 samples of 200 randomly selected firm-date observations in each sampling situation. The number of observations for each random sample is in line with the number of events included in the studies surveyed in Table 1.

Because observations are randomly selected, we know that the null hypothesis is true. A statistical test is correctly specified if it does not overreject a true null hypothesis. This means that the proportion of random samples affected by a Type I error should not be significantly larger than the theoretical value associated with the desired level of statistical significance. For each test, we compute the upper tail and lower tail rejection rates corresponding to the 5% confidence level (2.5% for each tail),[3] and determine whether there is a statistically significant overrejection of the null hypothesis for each of the two rejection regions (in our analysis, the maximum acceptable rejection rate for a one-tailed confidence level of 2.5% is 2.93%).[4]

To assess the power of the tests, we impose an abnormal shock to the CDS spread of each of the 200 randomly selected observa-

[2] CMA has been acquired by S&P Capital IQ. The CMA dataset is currently referred to as the S&P Capital IQ CDS dataset.

[3] In unreported analyses, we use 1% or 10% confidence levels to compute rejection rates. Results are fully consistent with those presented here.

[4] If the tests performed on each of the 5000 randomly drawn samples were independent and correctly specified, overrejection would follow a Bernoulli process, with a mean (μ) equal to the expected rejection rate (i.e., 2.5% for each tail at the 5% confidence level), and standard deviation equal to: $sd = \left(\frac{\mu(1-\mu)}{5000}\right)^{\frac{1}{2}}$. At the 5% confidence level, the null hypothesis that the test is well-specified for a specific rejection region and confidence level can be rejected if the rejection rate is greater than $\mu + 1.96 \times sd = 2.5\% + 1.96 \times \left(\frac{2.5\% \times (1-2.5\%)}{5000}\right)^{\frac{1}{2}} = 2.93\%$.

tions and test the null hypothesis of no abnormal spread changes using a two-sided test at the 5% confidence level. Because we impose an abnormal shock, we now know that the null hypothesis is false. The power of the statistical test is measured by the proportion of random samples where this false null hypothesis is rejected (i.e., by the incidence of Type II errors.). The power of a test depends on the magnitude and possibly the direction of the imposed shock. It is therefore important to calibrate the imposed shock to a level consistent with that observed in empirical studies on *ASC*. We study power separately for positive and negative shocks and use a conservative value of ± 0.50 bps for our primary analysis. By way of comparison, Jorion and Zhang (2009) estimate the average 1-day *ASC* experienced by creditor companies when a counterparty defaults to be 0.81 bps. We perform our analysis on the full sample as well as on samples that only include investment grade (IG) or speculative grade (SG) companies. In line with Bessembinder et al. (2009), we use a larger shock (± 1.00 bp) for SG companies.

3.2. Computing abnormal spread changes

The CDS spread change corresponding to firm i on day t ($SC_{i,t}$) can be computed as follows:

$$SC_{i,t} = CDS_{i,t} - CDS_{i,t-1} \tag{1}$$

where t and $t-1$ are two consecutive trading days and $CDS_{i,t}$ is the CDS spread of company i on day t. The $SC_{i,t}$ computed in Equation (1) reflects both idiosyncratic changes in the credit risk of firm i and changes in the general market credit risk. The goal of researchers is typically to isolate the idiosyncratic component of $SC_{i,t}$ and determine whether this idiosyncratic component is statistically different from zero for firms for which an event occurred.

Let $G_{i,t}$ be an indicator equal to 1 if an event occurred for firm i at time t, and 0 otherwise. Ideally, we would like to measure the difference between the $SC_{i,t}$ for a company where an event has occurred in t, and the $SC_{i,t}$ the same company would have experienced had no event occurred in t. Formally, we want to measure the following: $E\left[SC_{i,t} \middle| G_{i,t} = 1\right] - E\left[SC_{i,t} \middle| G_{i,t} = 0\right]$. If an event occurs, $E\left[SC_{i,t} \middle| G_{i,t} = 0\right]$ is not directly observable, and it therefore has to be approximated. With matching-portfolio models, this is done by calculating the spread change observed in a portfolio of companies that are similar to the focal company. The logic followed by matching-portfolio models is the same as that in synthetic control methods (Abadie and Gardeazabal, 2003; Abadie et al., 2010) and difference-in-difference analysis (Blundell and Dias, 2002; Imbens and Wooldridge, 2009; Lechner, 2011).

Once the matching portfolio for company i has been identified, the average CDS spread of the companies in the portfolio ($I_{i,t}$) is computed. The change in the CDS spread of these matched companies ($\triangle I_{i,t}$) proxies for $E\left[SC_{i,t} \middle| G_{i,t} = 0\right]$. As such, the difference between the observed and the expected change, computed as in Eq. (2), constitutes the *Abnormal* CDS Spread Change for firm i on day t ($ASC_{i,t}$):

$$ASC_{i,t} = \Delta CDS_{i,t} - \Delta I_{i,t} \tag{2}$$

The null hypothesis tested by event studies on CDS spreads is thus $H0: E\left[ASC_{i,t} \middle| G_{i,t} = 1\right] = 0$.

In this paper we consider four different adjustment methods (namely: the unadjusted method, the all-spreads method, the rating-matched method, and the spread-matched method), which correspond to different ways of identifying the companies that constitute the matching portfolio.[5] These adjustment methods are illustrated in the remainder of this section.

3.2.1. The unadjusted method

The first method is based on the assumption that *SC* in equation (1) is an unbiased measure of *ASC* in equation (2). We refer to this method as the *unadjusted method*. The majority of the studies surveyed in Table 1 only apply this method as a robustness check.

The unadjusted method has advantages and disadvantages. On the one hand, it does not require a matching portfolio, which makes its calculation easier and somewhat less arbitrary. On the other hand, the unadjusted method relies on rather strong assumptions, namely that the distribution of events is uncorrelated with changes in systematic risk.

3.2.2. The all-spreads method

One possible way to control for the systematic component of $SC_{i,t}$ is to compare it to the change in spread of the market as a whole. This approach is used by Bertoni and Lugo (2014) and Ismailescu and Kazemi (2010), for instance. The index is computed as the average of all of the available CDS spreads. In the spirit of Lyon et al. (1999), firms are only included if their CDS spread is available in both t and $t-1$, which avoids the potential bias induced by firms entering or exiting the portfolio. We refer to this particular adjustment method as the *all-spreads method*. The main advantage of the all-spreads method is that (unlike the rating-matched method discussed below) it does not require any additional information beyond the CDS spreads. The main drawback of this method is that it assumes that each company is equally exposed to variations in the general level of credit risk.

3.2.3. The rating-matched method

The most common method for building the portfolio of matched companies is based on credit ratings. In the *rating-matched method*, an equally-weighted index of all of the firms within the same credit rating category as firm i is used to compute $\Delta I_{i,t}$. This method is inspired by event studies on bonds and its rationale is that bonds with the same maturity and characterized by similar risk are expected to yield the same return (e.g., Kim et al., 1977). The empirical evidence confirms the existence of a systemic component that depends on the credit risk of companies and that affects their credit spreads (Berndt and Obreja, 2010). This method is used in the two seminal event studies by Hull et al. (2004) and Norden and Weber (2004) and has been used extensively in event studies on corporate bonds (e.g., Bessembinder et al., 2009).

In the rating-matched method, firms are only considered if their CDS spread and credit rating are available in both t and $t-1$. Moreover, firms whose credit rating varies in $t-1$ or t are excluded from the matching portfolio. This approach requires a decision about how rating categories are defined and, as illustrated in Table 1, previous studies use different definitions of rating categories.[6] We consider two different rating-matched methods. The first method uses five categories (AAA/AA; A; BBB; BB; B and below) defined as in Jorion and Zhang (2007). We refer to this as the rating(5)-matched method. The second approach only divides firms into IG versus SG as in Jorion and Zhang (2009), and we refer to this approach as the rating(2)-matched method.

3.2.4. The spread-matched method

Because CDS spreads are a measure of credit risk, a natural alternative to ratings as a way of identifying firms with similar levels of risk is to look at CDS spreads themselves. We suggest measuring the expected CDS spread change around the event using a portfolio

[5] In the matching portfolios, we only include CDS contracts with the same currency and contract clauses as the focal CDS contract. Results are similar when the reference portfolios also include CDS entities that differ from the focal CDS for currency or contract clauses.

[6] CDS-spread event studies using the rating-matched method also differ in terms of the credit rating agency they rely upon. For example, Hull et al. (2004) use Moody's ratings, while Jorion and Zhang (2007) use S&P ratings.

Table 2
Distribution of abnormal CDS spread changes. This table reports the descriptive statistics of the Abnormal CDS Spread Change (*ASC*) in bps computed with different adjustment methods. Statistics are based on 5000 random samples of 200 observations.

Method	Mean	Median	Sd	Skewness	Kurtosis
Unadjusted	0.021	0.000	109.705	0.260	27,903.724
All-spreads	0.020	0.017	109.310	0.315	27,915.045
Rating(2)-matched	0.042	0.001	108.002	0.804	28,109.559
Rating(5)-matched	0.036	0.000	106.270	1.118	28,050.431
Spread-matched	0.013	0.000	110.080	-0.044	26,946.873

of firms characterized by a similar level of pre-event CDS spread. We refer to this approach as the *spread-matched method*. In order to build the matching portfolio, for every observation we identify the firm that has the closest CDS spread to the focal firm in $t-1$. If multiple firms are identified as closest neighbors, we equally-weight them into the matched portfolio. The matching portfolio is composed of one firm in 89.61% of cases, of two firms in 8.16% of cases, and of three or more firms in the remaining 2.23% of cases.

More complex versions of the spread-matched method are conceivable along the lines of the approach taken in matching methods (Li and Prabhala, 2007). First, the matching portfolio could include the closest n companies in terms of CDS spread. The main disadvantage of this alternative approach is that, for larger values of n, the matched firms increasingly differ from the focal firm in terms of credit risk. Second, the matching portfolios could be built to include all firms whose CDS spread is within a pre-specified radius s (e.g., $s = 10$ bps) of the focal company's spread. The main disadvantage of this alternative matching method is that a matching portfolio might be empty if no firm falls within the radius s. Excluding events where no matching portfolio is available for the focal company would introduce a selection bias, as those companies would obviously be characterized by uncommon levels of CDS spread. Finally, the matching could be performed on the average spread over a pre-event period of d trading days. Requiring CDS spreads to be available for all trading days in a certain pre-event window would reduce the number of usable observations, and possibly introduce a selection bias along the lines of that discussed in Section 2 for factor models. In our analysis, we find that none of these more complex approaches produce better results than the simpler approach described above, and there is no evidence that the additional complexity of calibrating n, s, or d is compensated by better empirical performance. Accordingly, we present our results using the simpler spread-matched method described at the beginning of this section.

3.3. *Distributional properties of* ASCs *calculated using the different adjustment methods*

Following Bessembinder et al. (2009), we compute the moments of the distribution using the 5000 random samples of 200 observations rather than the entire population. Results are reported in Table 2.

Table 2 shows that, regardless of the adjustment method used, the distribution of *ASC* is non-normal (a Kolmogorov test rejects the null hypothesis of normal distribution at the 1% confidence level for all of the methods). In particular, the distribution is extremely fat-tailed, with high levels of kurtosis. We would therefore expect the t-test to be very weak, which is indeed what we verify in our analysis. Conversely, the sign-test generally becomes more powerful as kurtosis increases (Randles and Wolfe, 1979).

The difference in skewness between the spread-matched method and the other adjustment methods is remarkable: the skewness of *ASC*s is roughly two orders of magnitude smaller with the spread-matched method than with the rating-matched

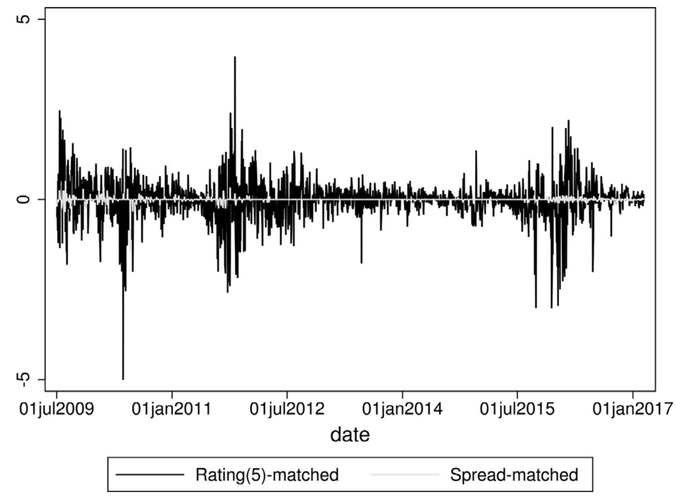

Fig. 1. Median daily abnormal CDS spread changes for different adjustment methods. This figure presents the cross-section median Abnormal CDS Spread Change (*ASC*) in basis points (bps) over time. *ASC*s are computed using the rating(5)-matched method (black line) and the spread-matched method (gray line) as presented in Section 3.2.

methods. A more symmetric distribution generally results in better non-parametric tests,[7] and we would thus expect the spread-matched method to outperform other adjustment methods when non-parametric tests are used.

To provide further evidence that statistical tests are expected to be better specified in combination with the spread-matched method, we graphically illustrate the extent to which the different methods eliminate the systematic component of changes in credit risk when computing *ASC*s. Fig. 1 reports the daily median *ASC*s computed using the rating(5)-matched method and the spread-matched method. An adjustment method should be able to eliminate the systematic component of CDS-spread variation; as such, the median *ASC* should ideally be (close to) zero on any trading day.[8]

It is clear from Fig. 1 that the median *ASC* is generally closer to zero when computed using the spread-matched method. The rating-matched method does not appear to be as effective at removing the systematic component of CDS spread changes as the spread-matched method.

In summary, our analysis of the distributional properties of *ASC*s gives us preliminary evidence that the t-test could be affected by severe Type II errors regardless of the adjustment methods, and that non-parametric tests are expected to perform better in combination with the spread-matched method.

3.4. *Common trend and unconfoundedness assumptions using the different adjustment methods*

Because of the similarity between matching-portfolio models and synthetic-control and difference-in-difference approaches, we can borrow some of the well-established theoretical features of these approaches (Blundell and Dias, 2002; Imbens and Wooldridge, 2009; Lechner, 2011; Abadie et al., 2010) to better understand how the choice of adjustment method can affect the size

[7] The Wilcoxon signed-rank test, for instance, assumes that the distribution is symmetric.

[8] Results obtained with the other adjustment methods are not included in Fig. 1 for the sake of readability. The temporal distribution of the cross-section median *ASC*s obtained with these alternative methods is similar to (or worse than) that observed with the rating(5)-matched methods.

and power of statistical tests. Two main assumptions have to be met for statistical tests to be well-specified.

The first key assumption is that the focal firm i and its matching portfolio share a common trend (Lechner, 2011): if no event occurs for the focal company, we should expect its CDS spread to move like (i.e., have a common trend with) the CDS spread of its matching portfolio. Formally: $E\left[ASC_{i,t}\middle|G_{i,t}=0\right]=0$. If this assumption is not met, statistical tests overreject the null hypothesis because a systematic difference between $SC_{i,t}$ and $\triangle I_{i,t}$ can be expected, even when no event occurs.

A second key assumption, known as the unconfoundedness assumption (Imbens and Wooldridge, 2009), states that the event and the abnormal change in spread have to be conditionally independent: no (unobserved) factor should determine both the event and the abnormal change in spread. Formally: $G_{i,t} \perp\!\!\!\perp ASC_{i,t}|G_{i,t}$. Failure to meet this assumption can lead to overrejection and to biases in the results, depending on the type of unobserved relationship between $G_{i,t}$ and $ASC_{i,t}$. This effect is equivalent to an omitted variable bias in linear regression, which can lead to spurious correlation and biases in estimated parameters (Imbens and Wooldridge, 2009; Li and Prabhala, 2007).

Adjustment methods meet the common trend and unconfoundedness assumptions in different sets of circumstances. In general, simpler adjustment methods are valid under more stringent assumptions about the distribution of the events. The simplest method, the unadjusted method, is only valid if the distribution of events is uncorrelated with changes in systematic risk. In most empirical applications, such as those in Table 1, events are hardly randomly distributed. Even with random events, the mere existence of a trend in credit risk, i.e., $E[SC_{i,t}]\neq 0$, can result in the violation of $E\left[ASC_{i,t}\middle|G_{i,t}=0\right]=0$. The all-spreads method may meet the common trend assumption to the extent that a unique systematic trend exists for all companies. However, significant differences exist in this respect among companies characterized by different levels of credit risk (Berndt and Obreja, 2010).

The rating-matched method partially controls for cross-sectional differences in credit risk. However, it can still fail to meet the common trend and unconfoundedness assumptions more often than the spread-matched method. A specific reason for this is that credit spread movements are a mean-reverting process (Zhu, 2006). The literature (e.g., Barber and Lyon, 1996) has highlighted that an effective way of meeting the common trend assumption in cases of mean reversion is to ensure that the focal firm and the matching portfolio exhibit similar pre-event levels of the variable of interest (in our case, the CDS spread). However, since ratings tend to be sticky (Altman and Rijken, 2004) and slow to adjust to changes in credit conditions (Norden and Weber, 2004), the credit risk of the focal firm can differ substantially from the average of firms in the same rating category.

Because of mean reversion, firms characterized by particularly low (high) levels of spreads compared to the average in their rating category are likely to exhibit an SC systematically higher (lower) than $\triangle I$, which in turns results in ASCs that are systematically higher (lower) than zero and, ultimately, in misspecified statistical tests. The size of this bias is proportional to the difference in pre-event levels of the variable of interest between the focal firm and the matching portfolio (Imbens and Wooldridge, 2009). By minimizing this difference, the spread-matched adjustment methods can therefore provide significantly improved results compared to the rating-matched methods.

To illustrate this feature, we calculate the ASC of CDS contracts across quantiles of CDS spread within the relevant rating category. For each rating category, we divide CDS contracts into 20 quantiles of CDS spread (e.g., the first quantile includes CDS spreads that are in the bottom 5% for that rating category, and the last quantile includes CDS spreads that are in the top 5% for that rating category). By construction, each rating category is represented in each quantile in the same proportion as in the general population. Fig. 2 illustrates, for each quantile, the median ASC by quantiles calculated using the rating(5)-matching and the spread-matched method.

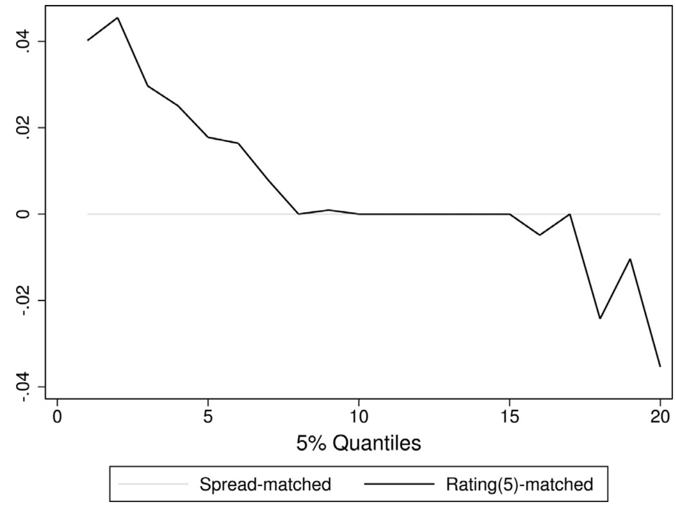

Fig. 2. Median abnormal CDS spread changes across quantiles of within-rating CDS spread. This figure presents the median Abnormal CDS Spread Change (*ASC*) in basis points (bps) across quantiles of within-rating CDS spread ranging from the Bottom 5% (1) to the Top 5% (20). Quantiles of the distribution are defined within rating classes (namely: AAA/AA; A; BBB; BB; B and below), so that each rating class is represented in each of the 20 groups in the same proportion as in the general population. *ASC*s are computed using the rating(5)-matched (black line) and spread-matched (gray line) methods as presented in Section 3.2.

Fig. 2 shows that the median ASC calculated using the rating(5)-matching is positive for CDS contracts that have a low CDS spread and negative for contracts that have a high CDS spread. If the spread-matched method is used instead, the median ASC is virtually null for each quantile. As the common trend assumption appears to be violated, we expect statistical tests to be misspecified in combination with the rating-matched method, but not with the spread-matched method. This is true in particular when events are more likely to occur for firms that already have high (or low) levels of CDS spread, i.e., when the unconfoundedness assumption is also violated. As discussed in Section 5, credit rating downgrades are a good example of such events.

4. Results

In this Section we report the main results of our analysis. We begin, in Section 4.1, by presenting our analysis of the whole sample of rated firms and of the subsamples of IG and SG firms. In Section 4.2, we show the results across subsamples of companies that are characterized by extreme levels of pre-event CDS spreads.

4.1. Analysis of the whole sample

4.1.1. Size of tests and adjustment methods

In Table 3, we report rejection rates with the t-test, sign-test, and signed-rank test for randomly drawn samples of observations from the dataset of usable rated firms (Panel A). In line with Bessembinder et al. (2009), we also show results obtained by sampling from a dataset including IG (Panel B) or SG (Panel C) firms only. As discussed in Section 3.1, rejection rates for each tail are computed at the 5% confidence level (i.e., 2.5% each tail). An adjustment method and statistical test combination is considered to be correctly specified when it does not result in a significant (at the 5% confidence level) overrejection for any of the two regions.

Table 3
Size of statistical tests and adjustment methods. This table presents rejection rates (in %) of the null hypothesis of no Abnormal CDS Spread Change for each tail at a theoretical significance level of 5%. Panel A reports results for samples that include all rated firms. Results for samples including IG or SG firms only are reported in Panels B and C respectively. The population firms are all rated corporate issuers. *N* refers to the number of firm-day observations in the sampling population. Adjustment methods are as defined in Section 3.2.

	t-test		Sign-test		Signed-rank test	
	2.50%	97.50%	2.50%	97.50%	2.50%	97.50%
Panel A: All rated (N=1233057)						
Unadjusted	2.26	1.00	8.46[a]	0.38	8.64[a]	0.62
All-spreads	1.82	1.32	1.38	3.16[a]	2.26	2.82
Rating(2)-matched	1.60	1.22	1.88	2.58	2.80	2.40
Rating(5)-matched	1.36	1.08	1.84	2.28	2.80	2.28
Spread-matched	1.56	1.64	2.00	1.68	2.30	2.38
Panel B: IG only (N=928732)						
Unadjusted	3.18[a]	1.48	9.64[a]	0.48	8.72[a]	0.62
All-spreads	2.32	1.90	1.08	3.64[a]	1.68	3.80[a]
Rating(2)-matched	2.64	1.58	1.74	2.30	3.16[a]	1.80
Rating(5)-matched	2.62	1.98	1.94	2.44	2.98[a]	2.02
Spread-matched	2.44	1.52	2.40	2.02	3.06[a]	2.24
Panel C: SG only (N=304325)						
Unadjusted	1.96	0.86	7.10[a]	0.36	7.64[a]	0.54
All-spreads	1.96	0.80	3.34[a]	1.26	4.46[a]	1.06
Rating(2)-matched	1.96	1.14	1.68	2.64	2.46	2.30
Rating(5)-matched	2.04	1.04	1.54	3.04[a]	2.84	2.68
Spread-matched	1.26	1.16	2.04	2.42	2.36	2.82

[a] Significant overrejection of the null hypothesis at the 5% confidence level.

Two main conclusions can be drawn from the results reported in Table 3. First, the *t*-test exhibits rejection rates well below the theoretical expected value. This suggests that, while size does not appear to be a serious concern for the *t*-test, power might well be. As discussed later, analyses focusing on the power of statistical tests support this conclusion. Of the two non-parametric tests, the Wilcoxon signed-rank test is misspecified more often than the sign-test. The analyses shown in Table 3 therefore suggest that the sign-test may be generally preferable to the other two statistical tests considered.

Second, the sign-test is always well-specified when, and only when, it is used in combination with the spread-matched and rating(2)-matched methods. The rating(5)-matched method results in a misspecified sign-test when SG firms are considered. The all-spreads method and the unadjusted method result in a misspecified sign-test in all of the three sampling situations. Finally, the unadjusted method results in the most severely misspecified tests, with one-tail rejection rates as high as 9.64%.

4.1.2. Power of tests and adjustment methods

We now examine the Type II error for the different statistical tests and methods. Table 4 reports rejection rates when a positive or negative shock is imposed on *ASCs* computed using different adjustment methods. Rejections are based on two-sided tests at the 5% confidence level. Again, we consider sampling from all rated firms (Panel A) as well as from IG (Panel B) or SG (Panel C) firms only.

As expected, the *t*-test is substantially less powerful than non-parametric tests. As shown in Panel A, the *t*-test rejects the null hypothesis in only between 6.14% and 11.90% of cases. The sign-test is clearly the most powerful: in each of the 30 different combinations of adjustment method, sampling situation, and direction of shock, the sign-test dominates both the *t*-test and the signed-rank test, rejecting the null hypothesis in between 90.44% and 94.98% of the cases in which it is well-specified. If we also take into account the fact that the sign-test is generally better specified than the signed-rank test, we can conclude that the sign-test dominates the other two statistical tests.

The two adjustment methods resulting in a well-specified sign-test in all three sampling situations, i.e., the spread-matched and the rating(2)-matched methods, exhibit similar rejection rates in Panels A and B. When SG firms are considered, the spread-matched method clearly dominates the rating(2)-matched method, with rejection rates that are higher by approximately 15 percentage points for both positive and negative shocks.[9]

All in all, the results in Tables 3 and 4 indicate that the sign-test is the best approach for identifying abnormal CDS spread changes. As for the adjustment methods, the spread-matched adjustment method generally appears to be the best suited when the risk of both Type I and Type II errors is considered.

4.2. Extreme CDS spread levels

As discussed in Section 3.4, the verified superiority of the spread-matched method may depend on its ability to minimize the difference in pre-event CDS spread between the event company and the matched portfolio. If that is the case, the excessive incidence of Type I errors with traditional adjustment methods should be particularly severe for observations characterized by extremely low or high levels of pre-event CDS spread. In this section, we specifically assess this conjecture by looking at the size and power of statistical tests when events are sampled from the two tails of the distribution. We construct two samples. The Bottom 5% sample includes observations in the first 5% of the distribution of pre-event CDS spreads in each of the five rating classes considered (namely, AAA/AA; A; BBB; BB; B and below). Similarly, the Top 5% sample includes the highest 5% of CDS spreads in each rating class. By construction, each rating class is represented in each of these two samples in the same proportion as in the whole population.

The size of tests is presented in Table 5. Panels A and B report rejection rates for random samples drawn from the Bottom 5% and Top 5% samples, respectively.

[9] Note that, although the sign test is more powerful when *ASCs* are computed with the unadjustment method, we have shown in the previous section that this adjustment method results in severely misspecified tests.

Table 4
Power of statistical tests and adjustment methods. This table presents rejection rates of the null hypothesis of no Abnormal CDS Spread Change (*ASC*) at the 5% confidence level, when a negative or positive shock is imposed on the *ASC* of each randomly selected event. A shock of ±0.5 bps is used for samples including all rated firms (Panel A) and IG firms only (Panel B); a 1-bps shock is used for samples including SG firms only (Panel C). *ASCs* are computed using the different adjustment methods illustrated in Section 3.2. *N* refers to the number of firm-day observations in the sampling population. Numbers in squared brackets refer to combinations of tests and methods that are misspecified according to the analysis reported in Table 3.

	Negative shock			Positive shock		
	t-test	Sign	Signed-rank	*t*-test	Sign	Signed-rank
Panel A: All rated (N=1233057)						
Unadjusted	11.90	[99.94]	[94.38]	7.68	[97.86]	[71.88]
All-spreads	8.56	[58.02]	46.14	6.88	[67.22]	46.88
Rating(2)-matched	7.92	91.94	78.62	6.14	92.76	75.38
Rating(5)-matched	7.56	94.94	81.98	6.96	94.98	79.74
Spread-matched	7.86	90.44	68.20	7.04	90.78	65.70
Panel B: IG only (N=928732)						
Unadjusted	[39.02]	[100.00]	[99.56]	[33.18]	[99.84]	[94.14]
All-spreads	14.70	[65.96]	[55.64]	15.64	[80.36]	[67.60]
Rating(2)-matched	43.30	98.94	[97.60]	38.40	98.66	[95.94]
Rating(5)-matched	43.58	99.22	[97.96]	39.46	99.16	[96.66]
Spread-matched	30.30	98.04	[89.92]	27.74	97.26	[87.94]
Panel C: SG only (N=304325)						
Unadjusted	8.34	[96.58]	[71.54]	4.84	[73.96]	[32.46]
All-spreads	8.54	[75.96]	[59.76]	4.74	[59.74]	[39.02]
Rating(2)-matched	6.32	42.90	31.72	4.28	46.82	31.58
Rating(5)-matched	5.62	[62.84]	42.62	4.16	[67.04]	44.12
Spread-matched	5.26	57.24	35.42	5.06	62.46	39.14

Table 5
Size of statistical tests and adjustment methods for extreme CDS spread levels. This table presents rejection rates (in %) of the null hypothesis of no Abnormal CDS Spread Change for each tail at a theoretical significance level of 5%. Adjustment methods are as defined in Section 3.2. Rejection rates are based on randomly selected samples from a population including only the bottom (Panel A) or the top (Panel B) 5% of observations by level of pre-event CDS spread within each rating class (namely: AAA/AA; A; BBB; BB; B and below). *N* refers to the number of firm-day observations in the sampling population.

	t-test		Sign-test		Signed-rank test	
	2.50%	97.50%	2.50%	97.50%	2.50%	97.50%
Panel A: Bottom 5% (N=61856)						
Unadjusted	0.40	2.16	2.24	2.24	1.78	3.76[a]
All-spreads	1.80	1.98	0.70	5.54[a]	1.46	3.96[a]
Rating(2)-matched	0.80	2.06	0.26	9.32[a]	0.58	7.28[a]
Rating(5)-matched	0.40	1.70	0.28	9.86[a]	0.66	8.84[a]
Spread-matched	0.96	0.86	2.48	1.50	2.66	2.22
Panel B: Top 5% (N=61518)						
Unadjusted	1.48	0.36	8.22[a]	0.36	8.98[a]	0.62
All-spreads	1.32	0.50	3.80[a]	1.26	5.22[a]	1.22
Rating(2)-matched	1.32	0.68	4.56[a]	1.10	6.26[a]	0.98
Rating(5)-matched	1.38	0.54	5.32[a]	0.96	6.70[a]	1.04
Spread-matched	1.34	0.94	2.72	1.68	3.30[a]	1.96

[a] Significant overrejection of the null hypothesis at the 5% confidence level.

Again, we observe that the *t*-test is never misspecified, but its very low rejection rates suggest it has very low power (as confirmed in Table 6 below). The superiority of the spread-matched method in terms of size is evident in both sampling situations: while traditional adjustment methods can result in rejection rates 2 to 4 times higher than the asymptotic limit of 2.5%, the spread-matched method results in a well-specified sign-test. When the spread-matched method is not used, a severe overrejection of the true null hypothesis occurs on the right tail for the Bottom 5% sample and on the left tail for the Top 5% sample. This is consistent with the fact that traditional adjustment methods do not correctly control for mean reversion. Rejection rates when a shock is imposed are presented in Table 6. The spread-matched method also generally dominates the all-spreads and rating-matched methods in terms of the power of the sign test. The only notable exception is when negative shocks are imposed on observations from the Top 5% sample. In that case, rejection rates are higher with the rating(5)-matched method, a result consistent with the severe overrejection on the left tail associated with this method.

In conclusion, when the event of interest systematically occurs for firms characterized by low (high) levels of spread, traditional adjustment methods very often lead to the conclusion that the event has a positive (negative) impact on spreads even if that is not actually the case. On the other hand, the spread-matched method always results in a well-specified sign-test even in these extreme sampling situations. Depending on the sample and the direction of the shock, the spread-matched method also generally reduces the incidence of Type II errors compared to the rating-matched methods.

Table 6
Power of statistical tests and adjustment methods for extreme CDS spread levels. This table presents rejection rates of the null hypothesis of no Abnormal CDS Spread Change (*ASC*) at the 5% confidence level, when a negative or positive shock of 0.5 bps is imposed on the *ASC* of each randomly selected event. Rejection rates are based on randomly selected samples from a population including only the bottom (Panel A) or the top (Panel B) 5% of observations by level of pre-event CDS spread within each rating class (namely: AAA/AA; A; BBB; BB; B and below). *ASCs* are computed using the different adjustment methods illustrated in Section 3.2. *N* refers to the number of firm-day observations in the sampling population. Numbers in squared brackets refer to combinations of tests and methods that are misspecified according to the analysis reported in Table 5.

	Negative shock			Positive shock		
	t-test	Sign	Signed-rank	*t*-test	Sign	Signed-rank
Panel A: Bottom 5% (N=61856)						
Unadjusted	27.82	100.00	[100.00]	56.54	100.00	[100.00]
All-spreads	9.58	[74.88]	[60.06]	12.20	[90.38]	[72.94]
Rating(2)-matched	6.98	[99.46]	[97.56]	12.48	[100.00]	[99.60]
Rating(5)-matched	6.06	[99.64]	[97.56]	12.68	[100.00]	[99.82]
Spread-matched	43.66	100.00	100.00	31.92	100.00	99.98
Panel B: Top 5% (N=61518)						
Unadjusted	2.24	[84.82]	[29.22]	1.64	[38.98]	[5.72]
All-spreads	2.14	[26.22]	[17.52]	1.78	[11.92]	[6.18]
Rating(2)-matched	2.36	[33.08]	[20.32]	1.84	[11.74]	[6.36]
Rating(5)-matched	2.36	[49.02]	[23.94]	1.76	[17.22]	[5.88]
Spread-matched	2.44	32.58	[10.98]	2.22	24.62	[7.16]

5. Illustrative example: rating downgrades

To illustrate how the choice of statistical tests and adjustment methods can affect inference, we perform an event study that investigates the *ASC*s associated with S&P credit rating downgrades. An event study on credit rating downgrades serves our purpose for two reasons. First, the extent to which rating actions determine significant changes in CDS spreads is one of the topics most frequently investigated by means of CDS-based event studies, as shown in Table 1. Both Hull et al. (2004) and Norden and Weber (2004) address this question mainly (if not exclusively) by focusing on a *t*-test based on *ASC*s computed using a rating-matched method.[10] Our results show that the *t*-test is very weak, leading to a high risk of Type II errors. Wengner et al. (2015) take a different approach, basing their inference on the results of a Wilcoxon signed-rank test, which we have shown to be a test with a high risk of misspecification, especially in combination with traditional adjustment methods. These studies come to different conclusions about the effect of downgrades: Hull et al. (2004), for example, find no significant change in CDS spreads around the announcement. Whether downgrades are associated with abnormal (positive) *ASC*s or not thus remains an open question with important practical and regulatory implications (Altman and Rijken, 2004; Kliger and Sarig, 2000; Kisgen and Strahan, 2010). Our findings suggest that a sign-test will generally be preferable to both a *t*-test and a Wilcoxon signed-rank test in answering this question.

The second reason for focusing on credit downgrades is that observations characterized by extreme levels of pre-event CDS spread are likely to constitute a substantial portion of the sample. Both Hull et al. (2004) and Norden and Weber (2004) find a significant increase in CDS spreads during the period preceding a downgrade, reflecting the increase in perceived credit risk. Accordingly, firm-day observations associated with downgrades are likely to be characterized by relatively high CDS spreads for their rating category. As Fig. 3 illustrates, we find this to be the case in our sample. More than 25% of the observations associated with a downgrade come from the Top 5% sample (and more than 40% are from the Top 10% of the population). Observations characterized by extremely high levels of CDS spread are thus severely over-represented in the event sample. As shown in Section 4.2, in this case the incidence of Type I errors can be particularly severe when adjustment methods other than the spread-matched method are used. The use of the spread-matched method could also reduce the risk of Type II errors, as the expected shock from downgrades is positive. All in all, the use of a sign-test in combination with the spread-matched method can significantly reduce the risk of both Type I and Type II errors compared to the tests and adjustment methods used in previous event studies on credit downgrades.

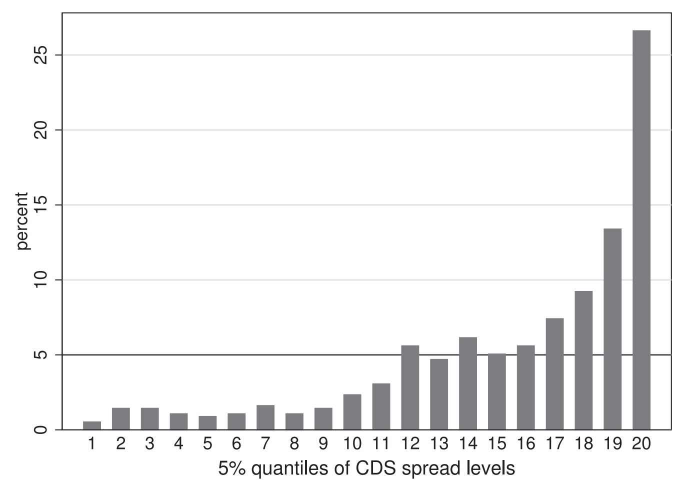

Fig. 3. Distribution of S&P downgrades across quantiles of within-rating CDS spread. This figure represents the distribution of S&P downgrades across quantiles of pre-event CDS spread ranging from the Bottom 5% (1) to the Top 5% (20). Quantiles of the distribution are defined within rating classes (namely: AAA/AA; A; BBB; BB; B and below), so that each rating class is represented in each of the 20 groups in the same proportion as in the general population.

In our event study, we follow Hull et al. (2004) and eliminate all of the downgrades that were preceded by other confounding rating events in the previous 90 calendar days, including downgrades or upgrades by other rating agencies (i.e., Moody's and Fitch). We identify 590 usable downgrades. For each adjustment method, Table 7 reports: the mean and median *ASC*; the share of

[10] Hull et al. (2004) focus on the rating class of the company before the rating action, while Norden and Weber (2004) change the index on the day of the event to reflect the company's new rating class. We follow the approach of Hull et al. (2004). Norden and Weber (2004) also perform a sign-test and a signed-rank test; however, they base their inference mainly on the results from the *t*-test.

Table 7
Illustrative example: S&P downgrades. This table presents the results of an event study on the Abnormal CDS Spread Change (*ASC*) associated with credit rating downgrades by S&P. Downgrades preceded by other rating events–including from Moody's or Fitch–during the previous 90 calendar days are excluded. After this filter is applied, there are 590 downgrade events for which *ASC*s can be computed using all of the adjustment methods considered. The corresponding statistic is presented for each test and adjustment method. *%Pos* indicates the share of positive *ASCs* over the number of positive and negative *ASCs* (ties excluded).

N = 590	Mean	Median	% Pos	*t*-test t	Sign z	Signed-rank z
Unadjusted	1.123	0.000	54.99%	0.522	2.300[a]	2.336[a]
All-spreads	1.305	0.596	55.59%	0.615	2.717[b]	2.831[b]
Rating(2)-matched	1.316	0.556	56.88%	0.616	3.338[b]	3.030[b]
Rating(5)-matched	1.295	0.513	57.14%	0.604	3.443[b]	3.423[b]
Spread-matched	0.740	0.489	55.89%	0.269	2.809[b]	3.413[b]

[*] : p-value < 10%
[a] : p-value < 5%.
[b] : p-value < 1%.

events characterized by $ASC > 0$ (*%Pos*); and the value of the statistic associated with each of the three statistical tests considered.

Three important observations can be made regarding Table 7. First, when the *t*-test is used, none of the adjustment methods result in the rejection of the null hypothesis. Given the limited power of this test, however, we should be careful about inferring that downgrade announcements have no impact on CDS spreads. Second, non-parametric tests in conjunction with the unadjusted, all-spreads, and rating-matched methods firmly reject the null hypothesis. Again, however, we should be careful about taking these results at face value: we show in Section 4 that these tests are often misspecified when *ASC*s are computed using traditional adjustment methods.

Third, our analysis illustrates that the sign-test conducted in association with the spread-matched method has the best characteristics in terms of size and power. When the spread-matched method is used, a sign test rejects the null hypothesis at the 1% confidence level. We can therefore infer that downgrade announcements are indeed associated with a statistically significant increase in CDS spreads.

6. Additional results and robustness

We replicate the analysis presented in Section 4 using alternative CDS populations, which differ from our main analysis by contract maturity, data source, period covered, and the inclusion of unrated firms. We also consider an alternative methodology for computing *ASCs* taking into account the sector of the target firm. For the sake of conciseness we only report rejection rates for the sign-test, which–as discussed in Section 4–is the best statistical test in terms of both size and power.[11] For the Bottom (Top) 5% sample we report rejection rates for the right (left) tail when no shock is imposed and rejection rates obtained by imposing a negative (positive) shock. The main results of these robustness checks are reported in Table 8 and are briefly discussed below. All of these additional analyses produce results that are fully consistent with the evidence presented in Section 4.

6.1. Different maturities

All of the CDS-data event studies conducted to date are based on 5-year maturity contracts. This maturity has quickly become the main standard for CDS contracts, meaning better data quality and availability. Yet, for some research projects it may be important to address how certain events affect the perceived credit risk of target firms over different horizons. As a robustness check, we thus repeat our analyses using Markit spread data for 10-year maturity CDS contracts. The number of usable firm-day observations is 430,570, a 65% reduction compared to 5-year CDSs over the same period.

The superiority of the spread-matched method is even more remarkable when 10-year CDS are considered. All of the other methods result in highly misspecified sign-tests when sampling from the two tails. The risk of Type II errors is also significantly reduced using the spread-matched method. Compared to rating-matched methods, the difference in rejection rates when the null is false can be as high as 50 percentage points.

6.2. Different data sources and periods

CDS spreads reported by different data sources can differ substantially and systematically (Mayordomo et al., 2013). It is therefore important to address the extent to which the conclusions reached based on Markit data can be extended to different databases. For this robustness check we use CMA data (retrieved from Datastream), which is the second most commonly used source for CDS-based event studies (see Table 1). Mayordomo et al. (2013) find that CMA data lead the price discovery process, making them a particularly appealing alternative to Markit data for event studies. CMA data are available via Datastream for the period from January 2004 to September 2010. After excluding quotes not based on actual trades and observations where we cannot compute *ASCs* with all of the adjustment methods considered, we are left with 865,798 firm-day observations.

This dataset covers the global financial crisis, a period characterized by extreme volatility and extreme values of CDS spreads, both in levels as well as in terms of daily changes. To better appreciate differences in the size and power of statistical tests before and during the crisis, we split the dataset into Pre-crisis (408,411 observations) and Crisis (457,387 observations) samples. In line with Mayordomo et al. (2013), we set the beginning of the crisis on August 1, 2007.[12] In general, we expect a significant deterioration in the size and power of statistical tests during the crisis, as volatility, kurtosis, and (absolute) skewness of daily spread changes increase. A relatively higher number of firms can be characterized as presenting uncommon levels of CDS spread during the crisis.

[11] All of these robustness checks confirm that the sign-test is better specified than the signed-rank test and more powerful than both the *t*-test and the signed-rank test in each of the alternative CDS populations we analyze in this section. Rejection rates for other tests, tails, and direction of the imposed shock not reported in Table 8 are consistent with the results presented in Section 4 and are available upon request.

[12] Mayordomo et al. (2013) set the beginning of the crisis using a Bai and Perron (2003) test to identify a structural break in the time series of corporate CDS spreads.

Table 8
Additional results on the size and power of statistical tests and adjustment methods. This table summarizes the main results for a number of additional analyses performed using different sampling situations, data sources, and definitions of *ASCs*. Unless otherwise stated, *ASCs* are computed using the different adjustment methods illustrated in Section 3.2. For each new sampling situation or method, we repeat the Panel A analyses of Tables 3 and 4, as well as those of Panels A and B from Tables 5 and 6. Reported rejection rates at the 5% confidence level are based on the sign-test. Bottom 5% and Top 5% samples are defined as in Table 5. For the Bottom (Top) 5% sample, we only report rejection rates for: i) the right (left) tail when no event occurs, and; ii) a two-tail test when a negative (positive) 0.5 bps shock is imposed. Panel A is based on a database of 10-year maturity corporate CDSs (Markit data). Panels B and C are based on 5-year corporate CDS data from Credit Market Analysis (CMA), spanning the period from January 1st, 2004 to July 31st, 2007 (Panel B) and from August 1st, 2007 to September, 30th, 2010 (Panel C). In Panel D, the Markit 5-year CDS database is used, including observations where the S&P domestic rating is missing. Panel E presents rejection rates for a "spread & sector" alternative method computed selecting the spread-matched firm from those in the same industry as the examined firm. Industries are defined based on the Thomson Reuters Business Classification (TRBC) at the Economic Sector level. *N* refers to the number of firm-day observations in the sampling population. Numbers in square brackets refer to combinations of tests and methods that are misspecified.

	All observations				Bottom 5%		Top 5%	
	No shock		Shock		No shock	Shock	No shock	Shock
	2.5%	97.5%	Neg.	Pos.	97.5%	Neg.	2.5%	Pos.
Panel A: 10-year CDS (N= 430570)								
Unadjusted	4.94[a]	0.72	[100.00]	[97.36]	3.76[a]	[100.00]	3.80[a]	[97.92]
All-spreads	2.88	1.60	68.68	48.96	6.60[a]	[58.14]	4.94[a]	[14.66]
Rating(2)-matched	2.28	1.82	92.90	70.20	14.94[a]	[97.72]	3.96[a]	[27.84]
Rating(5)-matched	1.78	2.18	95.34	75.32	12.94[a]	[99.26]	4.90[a]	[45.80]
Spread-matched	1.98	2.50	99.70	70.22	2.00	100.00	1.76	77.80
Panel B: CMA data, Pre-crisis (N= 408411)								
Unadjusted	6.22[a]	0.50	[100.00]	[100.00]	8.04[a]	[100.00]	6.20[a]	[44.18]
All-spreads	1.06	4.00[a]	[99.78]	[99.70]	18.22[a]	[100.00]	9.10[a]	[37.08]
Rating(2)-matched	1.48	2.40	100.00	99.98	16.80[a]	[100.00]	5.96[a]	[40.12]
Rating(5)-matched	1.92	2.42	99.98	99.98	16.00[a]	[100.00]	5.52[a]	[40.00]
Spread-matched	2.42	2.20	99.82	99.84	2.46	100.00	1.62	30.10
Panel C: CMA data, Crisis (N= 457387)								
Unadjusted	2.58	1.86	62.78	59.36	19.94[a]	[97.82]	5.92[a]	[3.94]
All-spreads	3.22[a]	1.38	[22.60]	[12.84]	0.22	65.84	7.64[a]	[5.06]
Rating(2)-matched	3.90[a]	0.82	[44.50]	[24.64]	0.46	90.10	5.66[a]	[3.94]
Rating(5)-matched	3.84[a]	1.36	[42.12]	[31.18]	0.46	89.48	4.60[a]	[3.94]
Spread-matched	2.40	1.78	45.08	43.98	1.60	98.64	2.10	4.66
Panel D: including unrated firms (N= 1443278)								
Unadjusted	8.98[a]	0.30	[99.96]	[98.98]				
All-spreads	1.26	3.32[a]	[58.84]	[68.16]				
Spread-matched	2.34	2.10	91.86	91.42				
Panel E: Spread and sector matching (N= 1190727)								
Spread-matched	2.22	1.88	90.92	90.90	1.94	100.00	2.36	26.32
Spread & Sector	2.08	2.22	93.88	93.94	3.12[a]	100.00	2.74	28.74

[a] significant overrejection of the null hypothesis at the 5% confidence level.

The misspecification of statistical tests in combination with traditional adjustment methods can thus become particularly severe in such a context, even when sampling from the entire population.

The results in Panel B of Table 8 refer to the Pre-Crisis period and are largely aligned with those obtained using Markit data. Only the spread-matched method results in a well-specified sign-test when sampling from the two tails. Over the general sample, both the spread-matched and rating(5)-matched methods appear to be acceptable. During the financial crisis (Panel C), the spread-matched method dominates the rating-matched method in both size and power: the rating-matched methods result in a severely misspecified sign-test not only when sampling from the tails, but also when sampling from the general population. Moreover, the rating-matched method has a larger Type II error than the spread-matched method regardless of the direction of the imposed abnormal change (the difference is more evident with positive shocks). In summary, the spread-matched method should be the preferred adjustment method regardless of the CDS dataset used, especially when the analyses include periods characterized by extreme trends and spread change volatility.

6.3. Unrated firms

A further advantage of the spread-matched method over rating-matched methods is that it can be implemented even when no ratings are available. This could increase the number of usable events, and reduce any selection bias linked to systematic differences between firms that arise because they are rated or not by a specific rating agency. Because of such systematic differences, the exclusion or inclusion of unrated firms might significantly affect the results of the analysis. Accordingly, we repeat the analysis on the size and power of tests including all observations in the dataset, regardless of the availability of a credit rating. This not only means that observations where no rating is available can be randomly sampled as events, but that they can also be included in a matching portfolio. We compare the performance of the spread-matched method with the performance of the other two methods that do not require ratings to be computed (namely, the unadjusted and the all-spreads methods).[13] Results for the size and power of the sign-test with the spread-matched method, reported in Panel D of Table 8, are virtually identical to those for rated firms only as presented in Panel A of Tables 3 and 4. The other two adjustment methods not based on ratings–unadjusted and all-spreads–again result in a misspecified sign-test.

[13] The Bottom 5% and Top 5% samples cannot be considered in this case, because they require information about credit ratings.

6.4. Matching by spread and sector

Industry sector can play a key role in explaining the cross-section of CDS spread changes (e.g., Aretz and Pope, 2013). We thus examine whether the performances of the spread-matched method can be significantly enhanced by matching within the same sector. We refer to this alternative method as the spread & sector adjustment method. Industries are defined based on the Thomson Reuters Business Classification (TRBC) at the Economic Sector level. Around 3.5% of the observations do not have a suitable match in the same industry and are thus removed from the population. The spread & sector method results in a slightly more powerful sign-test compared to the spread-matched method; for the general sample, rejection rates are around 3 percentage points higher. However, this alternative method also results in a slightly misspecified sign-test when sampling from the Bottom 5% sample. Overall, matching within the same industry does not seem to substantially improve the spread-matched method.

7. Conclusions

CDS spreads are an important and increasingly used tool for assessing how specific events affect companies' credit risk. Previous studies using matching-portfolio models have relied on different adjustment methods and have used different statistical tests to identify abnormal changes in CDS spreads. In this study, we argue that the adjustment method most commonly used in the literature, the rating-matched method, is likely to result in misspecified and/or weak tests for the null hypothesis of no *ASC*s because, due to ratings stickiness and mean reversion in spreads, it results in a violation of the common trend assumption. This issue can be particularly relevant for events that are not independent from the pre-event CDS spread of affected companies, i.e., when the unconfoundness assumption is also violated. This is the case for credit rating downgrades, for instance, which are more likely to occur for companies that have above-average CDS spreads in their rating category.

We thus propose a new adjustment method, the spread-matched method, which uses as a reference index a matching portfolio of firms characterized by similar levels of CDS spread (rather than ratings) before the event occurs. In order to compare our proposed method with those used in previous studies, we assess the incidence of Type I and Type II errors for different statistical tests when using different adjustment methods to compute *ASC*s.

Regardless of the adjustment method used, the signed-rank test is often misspecified – leading to high Type I errors – and the *t*-test is very weak – leading to high Type II errors. In most sampling situations, the sign-test is better specified than the signed-rank test and substantially more powerful than the *t*-test and the signed-rank test.

Unlike traditional adjustment methods, the spread-matched method always results in a well-specified sign-test. Consistent with our expectations, this difference is particularly evident for events involving firms with extreme CDS spreads, where the other adjustment methods fail to adequately control for the mean reversion of CDS spreads. Analyses based on CDS data for different maturities, periods, and data providers confirm the superiority of the spread-matched method over traditional adjustment methods. The spread-matched method also has three additional advantages over rating-matched methods: it is more parsimonious, it allows the analysis of larger samples, and it does not suffer from the potential sample selection bias that can arise when excluding firms without a credit rating.

In summary, our study suggests that when performing credit risk event studies using matched-portfolio models, the risk of both Type I and Type II errors can be minimized by using a sign-test in combination with the spread-matched adjustment method.

References

Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic control methods for comparative case studies: estimating the effect of california's tobacco control program. J. Am. Stat. Assoc. 105 (490), 493–505.

Abadie, A., Gardeazabal, J., 2003. The economic costs of conflict: a case study of the basque country. Am. Econ. Rev. 93 (1), 113–132.

Afonso, A., Furceri, D., Gomes, P., 2012. Sovereign credit ratings and financial markets linkages: application to european data. J. Int. Money. Finance. 31 (3), 606–638. doi:10.1016/j.jimonfin.2012.01.016.

Altman, E.I., Rijken, H.A., 2004. How rating agencies achieve rating stability. J. Bank. Finance 28 (11), 2679–2714. doi:10.1016/j.jbankfin.2004.06.006.

Andres, C., Betzer, A., Dourmet, M., 2016. Measuring Abnormal Credit Default Swap Spreads. Working Paper.

Aretz, K., Pope, P.F., 2013. Common factors in default risk across countries and industries. Eur. Financial Manag. 19 (1), 108–152.

Augustin, P., Subrahmanyam, M.G., Tang, D.Y., Wang, S.Q., 2014. Credit default swaps: a survey. Found. Trends Finance 9 (1–2), 1–196.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. J. Appl. Econ. 18 (1), 1–22. doi:10.1002/jae.659.

Barber, B.M., Lyon, J.D., 1996. Detecting abnormal operating performance: the empirical power and specification of test statistics. J Financ Econ. 41 (3), 359–399. doi:10.1016/0304-405X(96)84701-5.

Barber, B.M., Lyon, J.D., 1997. Detecting long-run abnormal stock returns: the empirical power and specification of test statistics. J. Financ. Econ. 43 (3), 341–372. doi:10.1016/S0304-405X(96)00890-2.

Batta, G.E., Qiu, J., Yu, F., 2016. Credit derivatives and analyst behavior. Account. Rev. 91 (5), 1315–1343. doi:10.2308/accr-51381. https://doi.org/10.2308/accr-51381

Bedendo, M., Colla, P., 2015. Sovereign and corporate credit risk: evidence from the Eurozone. J. Corporate Finance 33, 34–52.

Berndt, A., Obreja, I., 2010. Decomposing european CDS returns. Rev. Financ. 14 (2), 189–233.

Bertoni, F., Lugo, S., 2014. The effect of sovereign wealth funds on the credit risk of their portfolio companies. J. Corporate Finance 27, 21–35.

Bessembinder, H., Kahle, K.M., Maxwell, W.F., Xu, D., 2009. Measuring abnormal bond performance. Rev. Financial Stud. 22 (10), 4219–4258.

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

Blundell, R., Dias, M.C., 2002. Alternative approaches to evaluation in empirical microeconomics. Portuguese Econ. J. 1 (2), 91–115.

Brown, S.J., Warner, J.B., 1980. Measuring security price performance. J. Financ. Econ. 8 (3), 205–258. doi:10.1016/0304-405X(80)90002-1.

Brown, S.J., Warner, J.B., 1985. Using daily stock returns: the case of event studies. J. Financ Econ. 14 (1), 3–31. doi:10.1016/0304-405X(85)90042-X.

Callen, J.L., Livnat, J., Segal, D., 2009. The impact of earnings on the pricing of credit default swaps. Account. Rev. 84 (5), 1363–1394.

Ederington, L., Guan, W., Yang, L.Z., 2015. Bond market event study methods. J. Bank. Finance 58, 281–293. doi:10.1016/j.jbankfin.2015.03.013.

Finnerty, J.D., Miller, C.D., Chen, R.-R., 2013. The impact of credit rating announcements on credit default swap spreads. J. Bank. Finance 37 (6), 2011–2030. doi:10.1016/j.jbankfin.2013.01.028.

Galil, K., Soffer, G., 2011. Good news, bad news and rating announcements: an empirical investigation. J. Bank. Finance 35 (11), 3101–3119. doi:10.1016/j.jbankfin.2011.04.010.

Horvath, B.L., Huizinga, H., 2015. Does the european financial stability facility bail out sovereigns or banks? an event study. J. Money Credit Bank. 47 (1), 177–206. doi:10.1111/jmcb.12173.

Hull, J., Predescu, M., White, A., 2004. The relationship between credit default swap spreads, bond yields, and credit rating announcements. J. Bank. Finance 28 (11), 2789–2811. doi:10.1016/j.jbankfin.2004.06.010.

Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Lit. 47 (1), 5–86.

Ismailescu, I., Kazemi, H., 2010. The reaction of emerging market credit default swap spreads to sovereign credit rating changes. J. Bank. Finance 34 (12), 2861–2873. doi:10.1016/j.jbankfin.2010.05.014.

Jorion, P., Zhang, G., 2007. Good and bad credit contagion: evidence from credit default swaps. J. Financ. Econ. 84 (3), 860–883. doi:10.1016/j.jfineco.2006.06.001.

Jorion, P., Zhang, G., 2009. Credit contagion from counterparty risk. J. Finance 64 (5), 2053–2087.

Kim, E.H., McConnell, J.J., Greenwood, P.R., 1977. Capital structure rearrangements and me-first rules in an efficient capital market. J. Finance 32 (3), 789–810. doi:10.1111/j.1540-6261.1977.tb01989.x.

King, M.R., 2009. Time to Buy or Just Buying Time? The Market Reaction to Bank Rescue Packages. BIS Working Paper no. 288.

Kisgen, D.J., Strahan, P.E., 2010. Do regulations based on credit ratings affect a firm's cost of capital? Rev. Financial Stud. 4324–4347.

Kliger, D., Sarig, O., 2000. The information value of bond ratings. J. Finance 55 (6), 2879–2902. doi:10.1111/0022-1082.00311.

Lechner, M., 2011. The estimation of causal effects by difference-in-difference methods. Found. Trends Econ. 4 (3), 165–224.

Lehnert, T., Neske, F., 2006. On the relationship between credit rating announcements and credit default swap spreads for european reference entities. J. Credit Risk 2 (2), 83–90.
Li, K., Prabhala, N.R., 2007. Self-selection models in corporate finance. In: Eckbo, B.E. (Ed.), Handbook of Empirical Corporate Finance. North-Holland, pp. 39–86.
Longstaff, F., Mithal, S., Neis, E., 2005. Corporate yield spread: default risk or liquidity? new evidence from the credit-default swap market. J. Finance 60 (5), 2213–2254.
Lyon, J.D., Barber, B.M., Tsai, C.-L., 1999. Improved methods for tests of long-run abnormal stock returns. J. Finance 54 (1), 165–201.
Mayordomo, S., Peña, J.I., Schwartz, E.S., 2013. Are all credit default swap databases equal? Eur. Financial Manag. 20 (4), 677–713. doi:10.1111/j.1468-036X.2013.12023.x.
Norden, L., Weber, M., 2004. Informational efficiency of credit default swap and stock markets: the impact of credit rating announcements. J. Bank. Finance 28 (11), 2813–2843. doi:10.1016/j.jbankfin.2004.06.011.
Pop, A., Pop, D., 2009. Requiem for market discipline and the specter of TBTF in japanese banking. Q. Rev. Econ. Finance 49 (4), 1429–1459. doi:10.1016/j.qref.2009.08.001.
Qiu, J., Yu, F., 2012. Endogenous liquidity in credit derivatives. J. Financ. Econ. 103 (3), 611–631. doi:10.1016/j.jfineco.2011.10.010.
Randles, R.H., Wolfe, D.A., 1979. Introduction to the Theory of Nonparametric Statistics, 1. Wiley New York.
Shivakumar, L., Urcan, O., Vasvari, F., Zhang, L., 2011. The debt market relevance of management earnings forecasts: evidence from before and during the credit crisis. Rev. Account. Stud. 16 (3), 464–486. doi:10.1007/s11142-011-9155-6.
Wengner, A., Burghof, H.-P., Schneider, J., 2015. The impact of credit rating announcements on corporate CDS markets–are intra-industry effects observable? J. Econ. Bu.s 78, 79–91.
Zhang, G., Zhang, S., 2013. Information efficiency of the u.s. credit default swap market: evidence from earnings surprises. J. Financial Stab. 9 (4), 720–730. doi:10.1016/j.jfs.2011.10.005.
Zhu, H., 2006. An empirical comparison of credit spreads between the bond market and the credit default swap market. J. Financ. Serv. Res. 29 (3), 211–235.