ARE ALL CREDIT DEFAULT SWAP DATABASES EQUAL?

Sergio Mayordomo(1) Juan Ignacio Peña (2)* Eduardo S. Schwartz(3)

This version: 13/12/11

Abstract

In this study we compare the six major sources of corporate Credit Default Swap prices: GFI, Fenics, Reuters EOD, CMA, Markit and JP Morgan, using the most liquid single name 5-year CDS of the components of the leading market indexes, iTraxx (European firms) and CDX (US firms) for the period from 2004 to 2010. We find systematic differences between the data sets implying that deviations from the common trend among prices in the different databases are not purely random but are explained by idiosyncratic factors as well as liquidity, global risk and other trading factors. The lower is the amount of transaction prices available the higher is the deviation among databases. Our results suggest that the CMA database quotes lead the price discovery process in comparison with the quotes provided by other databases. Several robustness tests confirm these results.

JEL classification: F33, G12, H63

Keywords: Credit Default Swap prices; Databases; Liquidity

(1)Sergio Mayordomo is at the Department of Research and Statistics, Comisión Nacional del Mercado de Valores (CNMV), c/ Miguel Ángel 11, 28010 smgomez@cnmv.es.

(2) Juan Ignacio Peña is at the Department of Business Administration, Universidad Carlos III de Madrid, c/ Madrid 126, 28903 Getafe (Madrid, Spain) ypenya@eco.uc3m.es. (*) Corresponding author.

(3) Eduardo S. Schwartz is at the Anderson Graduate School of Management, UCLA, 110 Westwood Plaza Los Angeles, CA 90095 (USA) eduardo.schwartz@anderson.ucla.edu.

This paper was partially drafted during the visit of Sergio Mayordomo to the Anderson School at UCLA. We acknowledge financial support from MCI grant ECO2009-12551. We thank Teresa Corzo and other participants in the INFINITI 2011 Conference for useful comments.

1. Introduction

Over the last decade, the Credit Default Swap (CDS) market has grown rapidly.[1] Given the growth and the size of this market, quoted and transaction prices of CDS contracts are widely thought to be a gauge of financial markets' overall situation, as suggested by the GM/Ford credit episode in 2005, the US subprime fiasco in 2007-2009 or the Europe's debt crisis in 2010. Academic and policymakers alike have voiced concerns with respect to the CDS market's role in the above mentioned episodes and its possible influence in other financial markets, credit-oriented or otherwise. However, to properly address current concerns, careful empirical research is needed and therefore dependable CDS price data is a key requirement. The CDS market is an Over the Counter (OTC) market almost entirely populated by institutional investors and therefore, in contrast with an organized exchange like the NYSE, there is no reliable information on prices. The information on prices must be gathered from market participants on the basis of their voluntary participation on periodic surveys, with all the potential shortcomings such a situation may bring about. For instance, Leland (2009) reports that Bloomberg's CDS data is frequently revised weeks after and often disagrees substantially with other data sources such as Datastream. Given that price data deserve special attention, as the validity and power of the empirical results must be based on a dependable data source, in this study we investigate the differences in the main data sources employed by researchers and policymakers in this area. Specifically, we compare the six data sources for CDS prices commonly used in almost all the extant research: GFI, Fenics, Reuters

1 The global notional value of CDSs outstanding at the end of 2004, 2005 and 2006 was $8.42, $17.1 and $34.4 trillion, respectively. The CDS market exploded over the past decade to more than $45 trillion in mid-2007 and more than $62 trillion in the second half of the same year, according to the ISDA. The size of the (notional) CDS market in mid-2007 is roughly twice the size of the U.S. stock market (which is valued at about $22 trillion) and far exceeds the $7.1 trillion mortgage market and $4.4 trillion U.S. treasuries market. However, the notional amount outstanding decreased significantly during 2008 to $54.6 trillion in mid-2008 and $38.6 trillion at the end of 2008. This declining trend followed in 2009 (31.2 in mid-2009 and 30.4 at the end of 2009).

EOD, Credit Market Analysis (CMA) DataVision (CMA hereafter), Markit and JP Morgan.[2] Thus, we study the consistency of these six CDS data sources in the cross section and time series dimensions using the most liquid single name 5-year CDS of the components of the leading market indexes, iTraxx (European firms) and CDX (US firms). First we look at their basic statistical properties. Then we address two specific issues: (i) the factors explaining the divergences from the common trend among different CDS quoted spreads, and (ii) the relative informational advantage of the prices coming from different CDS databases.

Two price time series for the same single name CDS reported by different data sources should, in principle, be very close in the sense that both share a common trend, the underlying true value of the asset. Even if there are deviations from the common trend between the price series reported by the different datasets, one should expect that these deviations are non-systematic and therefore unrelated both to idiosyncratic factors such as firm size and industrial sector, and to systematic market liquidity or trading activity factors. If all the data sources are consistent among them, the use of a given data source should not affect research results and their financial and policy implications. But if there are significant deviations among them, the research implications may be sensitive to the specific data base employed. The inconsistency derived from private price's providers would also imply a damaging lack of market transparency affecting all financial agents such as investors, risk managers, and regulators.

We find that there are systematic departures from the common trend across databases. The analysis suggests that, although the different CDS quotes moved broadly in the same direction, there are very noticeable divergences for some entities in some days. Also, the discrepancies among databases appear to be more marked in specific time

[2] The first five databases are analyzed jointly for American and European firms while the JP Morgan data is employed additionally as a robustness test for a subset of European firms.

periods, probably reflecting market turbulences but it is important to remark that no single database provides quotes that are consistently above or below the quotes from other databases. We also find evidence suggesting that on average the days without trade information have higher quote dispersion than the days with trade price information.

Most importantly, deviations (in absolute value) from the common trend among the different CDS quoted spreads are not purely random, but are related to idiosyncratic factors such as firm size and also to liquidity, global risk and trading factors. We also find that the different data sources do not reflect credit risk information equally efficiently. Our results suggest that the CMA quoted CDS spreads led the credit risk price discovery process with respect to the quotes provided by other databases. All these results are robust to potential endogeneity or multicollinearity problems and to different econometric methodologies.

Our results have a number of important implications for empirical research using CDS prices. First, for US names with low trade frequency, our results cast doubts on the reliability of the existing price information because there are very few recorded trade prices in GFI. Thus, conclusions obtained in papers that have used these data are open, to some extent, to criticism on these grounds.[3] Second, the smaller the firm, the larger are the deviations among databases. Third the lower the market's liquidity and, the higher the VIX volatility index, the larger are the deviations from the common trend in prices across the different databases.

This paper is structured as follows. Section 2 presents a literature review. Section 3 describes the data employed in the analysis. Section 4 motivates the research hypotheses

[3] The aim of this paper is not to study whether the results obtained by previous literature would have changed depending on the data source employed. Our aim is to test whether the deviations among data sources are random or dependent on systematic factors as well as the reliability of the different data sources.

and introduces the methodology. Section 5 shows the empirical results while Section 6 confirms the robustness of the results and presents some extensions. Section 7 concludes.

2. Literature Review

The importance of comparing alternative financial databases is stressed in the classical papers by Rosenberg and Houglet (1974) and Bennin (1980) on the differences between CRSP and COMPUSTAT stock price data. However, in more recent times there are very few papers comparing databases. Schoar (2002) and Villalonga (2004) compare COMPUSTAT with the Longitudinal Research Database and the Business Information Tracking Series from the U.S Bureau of the Census, respectively, and show that different data sources have large impact on the answers to research questions. Despite the widespread use of CDS databases and the high relevance of their accuracy, to the best of our knowledge there exists no study that examines or compares data as well as databases. Our paper is a first attempt to fill this gap in the literature.

The first papers that compare, at least to some extent, different CDS data sources are Nashikkar and Subrahmanyam (2007) and Nashikkar, Subrahmanyam, and Mahanti (2009). However, this comparison is not the main focus of their paper and these authors do not present a detailed analysis. They simply conduct a test to ensure consistency between the CMA and GFI CDS spreads series over a short period when there was an overlap between the two series. They develop this test just to match GFI and CMA series and create a longer dataset given that they have the two data sources for different periods. Moreover, they do not report any results of the tests and simply state that they find consistency.

Mayordomo, Peña and Romo (2011) employ four different data sources (GFI, CMA, Reuters EOD and JP Morgan) to study the existence of arbitrage opportunities in credit derivatives markets focusing their attention to the single names CDSs and asset swaps. Although they find similar results employing any of the four previous data sources at the aggregate level, some differences appear at the individual reference entity level. They report their base results using GFI data but when they use the other data sources they do not find exactly the same number of arbitrage opportunities. For some individual firms they find arbitrage opportunities using GFI but they do not find them using some of the other data sources. In some other cases they find the opposite.

The Mayordomo et al.'s (2011) study above suggests that the differences in CDS prices from different databases can have a material influence on research results and therefore a careful analysis of the publicly accessible databases is called for. In fact, the problem could be potentially even more serious when researchers work with "unique" databases coming from a single dealer's quotes (contributor) and without crosschecking. It is important to emphasize that we use a broad array of CDS data sources where, for most of them, prices are put together based on information provided by several market traders and dealers. Using aggregate prices we focus on the market factors or characteristics that could affect the consistency among quoted prices. Thus, instead of using individual dealer's prices we use aggregated (composite/consensus) prices which allow us to have a more comprehensive perspective on the market.

The only previous paper that employs different CDS prices (trades and quotes) is Arora, Gandhi and Longstaff (2010). They examine how counterparty credit risk affects the pricing of CDS contracts using a proprietary data set. Specifically, their data set spans from March 2008 to January 2009 and includes contemporaneous CDS transaction prices and quotations provided by 14 large CDS dealers for selling protection on the same set of underlying reference firms. The authors find differences across dealers in how counterparty credit risk is priced. That is, counterparty credit risk is not priced symmetrically across dealers and they consider that these asymmetries could be due to differences in the microstructure and legal framework of the CDS market. They argue that dealers may behave strategically in terms of their offers to sell credit protection.

We use aggregate data, which are formed after grouping the information of the market traders and dealers instead of individual dealer prices, to study the potential divergence among the composite CDS spreads. By concentrating on the aggregate prices we focus on the market factors or characteristics that could affect the consistency among quoted prices but we do not try to explain the effect of potential differences among the individual dealers. As Arora, et al. (2010) sustain, the decentralized nature of the CDS markets makes the transaction prices somewhat difficult to observe. This is why most empirical research analyses based on the CDS markets use price quotes instead of transaction prices.

Longstaff, Mithal, and Neis (2005) argue that the composite prices include quotations from a variety of credit derivatives dealers and therefore, these quotations should be representative of the entire credit derivatives market. We complement our analysis using also GFI transaction prices and Fenics prices (elaborated by GFI) which are based on a combination of transaction and judgmental prices, the latter computed using the Hull and White methodology and therefore not dependent on contributors.[4]

3. Data

The six publicly available data sources that we employ in this paper are GFI, Fenics, Reuters EOD, CMA, Markit and JP Morgan. As was mentioned above, the first five databases are analyzed jointly and JP Morgan data is employed additionally as a

[4] As explained in Section 3, we also use an additional single sources data (JP Morgan). JP Morgan data refer to individual dealer's prices and we use them in the robustness analysis.

robustness test for the European firms given that we do not have JP Morgan data for the American firms.

- GFI, which provides traded CDS spreads, is a major inter-dealer broker (IDB) specializing in the trading of credit derivatives. GFI data contain single name CDS transaction prices for 1, 2, 3, 4, and 5 years maturities. They are not consensus or indicative prices.[5] Thus, these prices are an accurate indication of where the CDS markets traded and closed for a given day. GFI data have been used by Hull, Predescu, and White (2004), Predescu (2006), Saita (2006), Nashikkar and Subrahmanyam (2007), Fulop and Lescourret (2007), Nashikkar, Subrahmanyam, and Mahanti (2009) among others.

- Fenics (elaborated by GFI) data are a mixture of traded, quoted and estimated CDS spreads. Fenics' data are credit curves for the whole term structure of maturities, generated hourly (all trading days) for more than 1900 reference entities. Data points in a given name's credit curve can be actual trades or mid prices calculated from the bid/offer quotes. If there are no market references, the Fenics CDS spread is computed using the Hull and White methodology to ensure that a credit curve always exists for each reference entity.[6] Fenics data have been used in Mayordomo et al. (2011) among others[7].

---

[5] Consensus and indicative data are trusted less nowadays given the increased market's volatility. There exist differences of up to 100% between consensus prices from leading providers compared to actual trades on GFI systems. The reason is that consensus process is inherently slow and the prices originate from back office staff which can be swayed by the positions they already hold in their books, and also perhaps because they do not have a front office's market view.

[6] Although Fenics is computed using the approximations mentioned above, it is a reasonably accurate data source. For instance, the median of the absolute difference in basis points between five years CDS premiums as defined by Fenics and the actual quotes or transaction prices registered in other databases for the period between April 2001 and May 2002, is equal to 1.16, 2.01 and 3.82 bps for AAA/AA, A and BBB ratings for a total of 2,659, 9,585 and 8,170 companies respectively.

[7] GFI is a broker which also reports the Fenics prices. The data reported by GFI are transactions prices or bid/quotes in which capital is actually committed. This data is only available when there is a trade. When there is not, GFI constructs the Fenics curve which is available daily with no gaps. To compute the Fenics curve, GFI uses its own information on transactions or quotes. If for a given day neither prices nor quotes are available, Fenics data is computed by means of Hull and White's methodology.

- Reuters EOD provides CDS composite prices. Reuters takes CDS quotes each day from over 30 contributors around the world and offers end of day data for single names CDSs. Before computing a daily composite spread, it applies a rigorous screening procedure to eliminate outliers or doubtful data. Mayordomo et al. (2011), among others, employ CDSs data from Reuters.

- Credit Market Analysis (CMA) DataVision does not categorizes its CDS prices along the “composite” or “consensus” lines but in order to bring more transparency to CDS information, CMA uses a strict aggregation methodology, instead of “composite” or “consensus” methods, depending on the intraday market activity. The data aggregation is not equally weighted but the different weights are based on the respective age and length of the original sample employed (the last contribution is more influential than the older ones). CMA collects its CDS data from a robust consortium which consists of around 40 members from the buy-side community (hedge funds, asset managers, and major investment banks) who are active participants in the CDS market.[8] CMA reports bid, ask and mid prices. Among the papers that employ CMA data are Nashikkar and Subrahmanyam (2007) and Nashikkar, Subrahmanyam, and Mahanti (2009).

- Markit provides composite prices. The Markit Group collects more than a million CDS quotes contributed by more than 30 major market participants on a daily basis. The quotes are subject to filtering that removes outliers and stale observations. Markit then computes a daily composite spread only if it has two or more contributors.

[8] The buy-side community includes major credit-focused houses that receive up to 20,000 e-mail pricing messages a day, covering a wide array of credits; and boutique experts focusing on niche credits. These contributors are spread geographically across Europe and the U.S.. Each of these members contributes their CDS prices to a CMA database which they receive in Bloomberg formatted messages (as well as forms) from their sell-side dealers. Hence, CMA has access to a constant stream and continuously evolving pool of CDS data. The access to OTC communication between buy-side trading desks and their counterparties guarantees that the prices received by CMA from the buy-side community are very likely to be tradable or even executable prices and that they capture market conditions as they evolve throughout the day. Of course it is difficult to know precisely whether all of them are tradable or not.

Once Markit starts pricing a CDS contract, data will be available on a continuous basis, although there may be missing observations in the data. Markit is one of the most widely employed dataset. Papers that employ this dataset include: Acharya and Johnson (2007), Zhang, Zhou and Zhu (2009), Jorion and Zhang (2007), Jorion and Zhang (2009), Zhu (2006), Micu et al. (2004), and Cao, Yu, and Zhong (2010).

- Our last database is J.P. Morgan quotes. It contains mid-market data provided by J. P. Morgan which is one of the leading players and most active traders in the CDS market. The data from J.P. Morgan is employed for a subgroup of European firms as part of the robustness tests (data is not available for US firms). This dataset is employed in Aunon-Nerin, Cossin, Hricko, and Huang (2002), Blanco, Brennan, and Marsh (2005), and Chen, Cheng, and Liu (2008) among others.

To summarize, three of the data sources (Reuters EOD, CMA and Markit) employ data from a variety of contributors (over 30 potential dealers/traders) to report composite prices. GFI reports traded CDS spreads. Fenics is a mixture of traded, quoted and calculated CDS spreads all of them based on the same data source and without depending on contributors. Finally, the last data source is obtained from one of the most active traders in the CDS market (JP Morgan) and reports mid-quoted prices obtained from their own traders. Thus, the information reported by Reuters EOD, CMA and Markit may also contain the information of JP Morgan's quoted CDS spreads[9].

For our analysis we use US firms included in the CDX index, as well as European firms included in the iTraxx index. At any point in time, both the CDX and iTraxx indexes contain 125 names each but the composition of the indexes changes every six months.

[9] CMA and GFI span from January 1, 2004 to March 29, 2010 for all firms. Markit spans from January 1, 2004 to December 8, 2009 for all firms. Fenics spans from January 1, 2004 to June 3, 2009 for most of the firms and from January 1, 2004 to March 29, 2010 in the remaining seven firms. Reuters spans from December 3, 2007 to March 29, 2010. JP Morgan spans from January 1, 2005 to August 13, 2009. For this reason, in the robustness test in which we add JP Morgan data we limit the length of our sample from January 1, 2005 to August 13, 2009 to focus in the cases in which we have observations from JP Morgan.

We do not use all the single names CDSs in these indexes but concentrate on the most liquid single names CDSs. As in Christoffersen, Ericsson, Jacobs, and Jin (2009) we use only the single name CDSs which constitute the iTraxx and CDX indexes over the whole sample period which spans from January 2004 to March 2010. We end up with 47 (43) firms which stay in the iTraxx (CDX) index during the whole sample period and for which we are able to obtain equity price information[10]. We guarantee a minimum consistency between the single name CDS spread obtained from the different data sources by requiring that all of them have the same maturity (5-year), currency denomination (Euros for the European and US Dollars for the American CDSs), seniority (senior CDS spreads), and restructuring clause (Modified-Modified Restructuring for the European and Modified Restructuring for the American CDSs).

Table 1 reports statistics on Number of Trades or Quotes, Number of Trades or Quotes per day, Mean, Standard Deviation, and Median of the CDS spreads.[11] In Panel A we report the names classified by index and sector for both the American and European firms. Panel B provides aggregated CDS summary statistics for all the observations distinguishing between European and American firms over the five data sources. The information is divided into two periods: before and during the financial crisis. The actual sample size of the different data sources differ due to the existence of missing values and different covered periods and for this reason, we report the summary

[10] It could be argued that this selection procedure could introduce some survivorship bias in our sample. It should be noted that the components of the indexes are investment grade CDSs firms which are the most actively traded names in the six months prior to the index roll. If in a given period a single name CDS is excluded from the index it is not necessary due to the fact that the firm enters financial distress but simply because of liquidity reasons. On the other hand if a name is downgraded to non-investment it is, of course, excluded from the index. Notice however that one should expect that the agreement among databases on the CDS price for a given name should be higher for the most liquid names. Thus, this possible survivorship bias will tend to make the prices from different databases more in agreement than they are in fact. Consequently if we find significant disagreement among prices from different sources, the empirical evidence is even more compelling.

[11] Detailed descriptive statistics (number of trades or quotes, the mean, standard deviation, median, skewness, kurtosis, and autocorrelation) for all the single name CDS spreads obtained from all the data sources used in this study are available upon request.

statistics for the cases in which we have common observations (trades and quotes) in all the data sources in Panel C. Panel D reports the summary statistics for the cases in which there is no trade but there are quotes in all the data sources.

For European firms before the crisis there is price information on trades in 35% of the days and there is, on average, a considerable degree of agreement among the prices provided by the different databases. However there are a few cases where some noteworthy differences can be found.

During the crisis however there is information on transaction prices only in 15% of the days. At the individual level we find that the skewness is usually positive but there are some noticeable differences in persistence: GFI prices are clearly less persistent (0.83) than the quotes from the other databases. In fact, all the names which trade less than the 10% of the days have first-order autocorrelation coefficients around 0.6.[12]

Regarding American firms, we observe that before the crisis there is price information on trades in 15% of the days and there is, on average, a fair amount of agreement among the prices provided by the different databases. In all cases there is a high degree of persistence with first order autocorrelation coefficient near one. However, as is the case with European firms, there are a few names with a relatively similar number of observations where some salient differences can be found.[13]

[12] At the individual level we find that the discrepancies, even in the cases in which a comparable number of observations across the different data sources are available, are much more frequent and more remarkable as the cases of PPR, Volvo, Enel and some others suggest. Specifically, in the case of Volvo there is a difference of 33 b. p. between the highest average price (Reuters, 253b.p.) and the lowest (CMA, 220 b.p.). Other discrepancies are found even between the datasets (CMA and Markit) with more similar quotes, as in the cases of PPR (16 b.p.) and Enel (9 b.p.).

[13] Before the crisis, Cigna's CDS quoted spreads obtained from Fenics are more than 30% higher than CMA and Markit quotes. On the other hand, General Electric's CDS spreads obtained from Fenics are 62% lower than the ones obtained from CMA and Markit.

Transaction prices are only available in 2% of the days during the crisis, and the discrepancies are both more frequent and more remarkable.[14]

In summary, the preliminary analysis suggests that the crisis has had a strong effect on the degree of disagreement of the different databases in several individual reference entities, and especially so for US names.

It is worth noting that, although the total averages are in most cases fairly close, there could be some noteworthy discrepancies both at the entity (as the preliminary analysis above suggest) and also at the cross-sectional level that cannot be captured by these statistics.

To clarify this point, we first compute the absolute value of the average difference across pairs of data sources[15] (CMA - Markit, CMA - Fenics, CMA - Reuters, Markit - Fenics, Markit - Reuters, Fenics – Reuters) and then divide it by the average CDS spread across the four previous data sources (CMA, Markit, Fenics, and Reuters) for each firm every day. Then, we calculate the average of the previous series every day across the total number of firms. This is the Data Sources' Average Absolute Discrepancies (AAD) time series and is shown in Figure 1 for days with trades (Trade) and for days without trades (No Trade). The average value of Trade is 0.031 and its volatility is 0.021. The average value of No Trade is 0.053 and its volatility is 0.017. The two sample unpooled t-test with unequal variances has a t-statistic of 33.68 under the null of equal means, suggesting that on average the days without trade information have higher quote dispersion. The AAD series show a very dynamic behaviour, with

[14] For instance in the case of American International Group there is a difference of 111 b.p. between the Reuters quote (825 b.p.) and the CMA quote (714 b.p.) and 80 b.p. between the Reuters quote (815 b.p.) and the Markit quote (745 b.p.). Other notable disagreements between the highest and the lowest prices in names with a relatively similar number of observations across the different data sources are found for Comcast (the average difference between Markit and Reuters CDS spreads is 61 b.p.), General Electric (the average difference between CMA and Reuters is 34 b.p.), and XL Capital (the average difference between CMA and Reuters is 57 b.p. and between CMA and Markit 34 b.p.), among others.

[15] GFI data is not used due to the scarcity of transaction prices during the crisis.

some noticeably turbulent episodes, for instance in 2005 given the impact of the crisis experienced by General Motors (GM) and Ford in May 2005 on the credit default swap (CDS) market. Both firms' CDS premia increased sharply just before the downgrading of their credit ratings in May 2005. All other CDS premia also rose markedly during this period for US and European firms. The more salient episodes in the AAD series are in September 2008 in the days surrounding the Lehman Brothers collapse when the AAD took its highest value to date (10%). In summary the data suggest that discrepancies from the common trend among databases are persistent and related with market-wide significant episodes. We address the modelling of these discrepancies in Section 4.

4. Research Hypotheses and Methodology

The main analysis of the data is based on two testable hypotheses. These hypotheses and the methodology employed to perform the empirical tests are detailed in this section.

Hypothesis 1: The volatility of the deviations from the common trend of the quoted prices provided by the different CDS data sources is not related to systematic factors.

In other words, large deviations (in absolute value) from the common trend appear randomly among databases and are unrelated with risk and liquidity factors (global or idiosyncratic). The test of Hypothesis 1 is based on a regression in which the dependent variable is the logarithm of the standard deviation of the 5-year quoted CDS spreads reported by the different data sources which is denoted by $\log(sd(CDS))_{i,t}$ This variable is computed with the j available CDS quoted spreads *(j = 1,..,4* where *1*= CMA, *2* = Markit, *3* = Reuters and *4* = Fenics) for a given underlying firm *i (i = 1,..,90)* on every date t as follows: $\log(sd(CDS\ ))_{i,t} = \log((1/n\sum_{j=1}^{n}[CDS_{j,i,t} - (1/n\sum_{j=1}^{n}CDS_{j,i,t})]^2)^{0.5})$, where

n is the number of data sources from which we observe CDS spreads, with the maximum n equal four whenever CMA, Markit, Reuters and Fenics report the CDS spreads for firm i at time t.

By defining the dependent variable in this way[16] we get rid of the common trend (the average) and concentrate on the deviations from the common trend. The regression equation is as follows:

$$\log(sd(CDS))_{i,t} = \alpha + \beta' X_{k,i,t} + u_{i,t} \qquad (1)$$

where the vector $X_{k,i,t}$ includes k explanatory variables: the logarithm of the firm market capitalization, a trade dummy, the number of days without a trade, the interaction of the number of days without a trade one day ago and the trade dummy , the CDS bid-ask spread, the VIX Index and a number of databases dummy.[17] The vector β' includes the regression coefficients corresponding to these k variables while the parameter α is the intercept of the regression. The residual term is denoted by $u_{i,t}$. The trade dummy is equal to one if there is a trade in the GFI platform at the current date in the 5-year maturity contract, and zero otherwise. [18] The number of days without a trade variable measures the number of days without a trade up to the current date. The interaction variable is constructed as the interaction between the number of days without a transaction up to one day before the current date and the trade dummy. The last variable intends to give an indication about how many data points were used to compute the dependent variable and is a dummy which equals one when all, or all minus one, of

[16] We take logs to induce the data to meet the assumptions of the regression method that is to be applied; because the distribution of the standard deviation variable is strongly right skewed (the skewness of the original series is 25.10 while the skewness of the log series is 0.21).

[17] Hausman's test rejects the random effects specification in favor of a fixed-effects specification, with a p-value of 0.05.

[18] We are considering trades for the 5-year maturity contract only given that the number of trades in the other maturity contracts is very low. The total number of trades according to GFI information during the sample period and for the firms that we consider is 26,126 while the number of trades which occurred in the other maturities (1 and 3 years contracts) is 1,100 confirming that the most liquid contract is the 5-year CDS contract.

the data sources report a price.[19] If the null hypothesis is true no significant coefficients should be found in equation (1) because differences in price dispersion between databases should be purely random.

Hypothesis 2: The different data sources reflect credit risk information equally efficiently or, equivalently, all databases contribute equally to the price discovery process. Given that transaction prices are very scarce for some firms, only quoted prices are employed and therefore the comparison is among CMA, Markit, Fenics and Reuters.

To test Hypothesis 2 we employ the Gonzalo and Granger's (1995) model which is based on the following Vector Error Correction Model (VECM) specification and it is used to study the effectiveness of the different data sources in terms of price discovery:

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{p} \Gamma_i \Delta X_{t-i} + u_t \qquad (2)$$

where equation (2) is formed by a vector autoregressive (VAR) system formed by two equations defined from the vector X_t which includes a pair of CDS quotes or prices of the same underlying firm from two different databases and an error correction term which is defined by the product $\beta' X_{t-1}$ where $\beta' = (1 - \beta_2 - \beta_3)$ are estimated in an auxiliary cointegration regression. The series for the pair of CDS prices included in X_{t-1} must be cointegrated to develop this analysis and the cointegrating relation is defined by $\beta' X_{t-1} = (CDS_{SOURCE\ A,t-1} - \beta_2 - \beta_3 CDS_{SOURCE\ B,t-1})$ which can be interpreted as the long-run equilibrium. The parameter vector $\alpha' = (\alpha_1, \alpha_2)$ contains the error correction coefficients measuring each price's expected speed in eliminating the price difference and it is the base of the price discovery metrics. The parameter vector Γ_i for $i = 1,..p$, with p indicating the total number of lags, contains the coefficients of the VAR system

[19] We do not employ values from zero to four given that we only have observations on Reuters EOD after December 2007 which is very close to the beginning of the crisis and may reflect something different to what we want to study in this paper.

measuring the effect of the lagged first difference in the pair of CDS quotes on the first different of such quotes at time *t*.[20] Finally, u_t denotes a white noise vector. The percentages of price discovery of the CDS quote *i* (where *i =1, 2*) can be defined from the following metrics GG_i, *i=1,2* which are based on the elements of the vector α':

$$GG_1 = \frac{\alpha_2}{-\alpha_1 + \alpha_2}; \qquad GG_2 = \frac{-\alpha_1}{-\alpha_1 + \alpha_2} \qquad (3)$$

The vector α' contains the coefficients that determine each market's contribution to price discovery. Thus, given that $GG_1+GG_2=1$ we conclude that market 1 leads the process of price discovery with respect to market 2 whenever market 1 price discovery metric GG_1 is higher than 0.5. If the null hypothesis is true (no dominant market) the percentage of price discovery will be the same for the names from all databases and equal to 0.5. We estimate the price discovery metric for each firm using pairs of CDS spreads and then test whether the average price discovery metric is significantly higher than 0.5 using the mean t-statistic: $Mean \quad t-stat = \frac{(Mean\,(PDMetrics\,) - 0.5)}{Std.Dev\,(PDMetrics\,)/\sqrt{\#\,metrics}}$,

where # metrics denotes the number of firms for which it is estimated the price discovery metric from a given pair of CDS spreads.

5. Empirical Results

5.1 Regression Results: Hypothesis 1

Table 2 shows the regression results obtained from fitting equation (1) to data from the five databases. Column 1 reports the results for the whole sample whereas Column 2 reports the results for European firms and Column 3 for US firms. Negative and significant coefficients for the explanatory variable measuring size (log (market cap)) are found suggesting that the CDS prices for large firms tend to be more in agreement

[20] The optimal number of lags is determined by means of the Schwarz information criteria.

among databases than the prices for small firms. Or in other words, the volatility of the deviations from the common trend is lower for large firms. This effect is also noticeably stronger for US firms. The coefficients for the explanatory dummy variable “trade” are negative and significant suggesting that when there are transaction prices available for a given day, the quotes from different contributors tend to agree more closely. This is in agreement with the results on basic statistical properties summarized in the Section 3 above. Consequently, the positive (but only significant for US firms) effect found for the variable days w/o trade implies that the longer the period without transaction price information, the greater the disagreement among quotes because, the weaker is the referential value of the previous price. The interaction between the trade dummy and the number of days without a trade one day before the current date has a negative sign (but non-significant for European firms) indicating that the effect of the trade is more influential when the number of days without price trade information is larger.[21]

Regarding the liquidity variable, the bid-ask spread, has, as expected, positive and significant coefficients implying that the more illiquid is the market, the more difficult is to infer appropriate prices and the higher are the deviations from the common trend among the different data sources.

Regarding the global risk proxy, the effect of the VIX index is positive and significant. The higher the global risk, the higher the dispersion from the common trend among individual CDS spreads. [22, 23]

[21] One possible explanation is that traders will pay more attention to the new information reported by GFI when there has been no recorded trading activity for some time.

[22] The CDS bid-ask spread and VIX Index variables should not cause any collinearity problem given that their correlation is 0.480 (this is the highest correlation among explanatory variables). However, we further investigate this aspect and others regarding potential endogeneity problems derived from the use of the VIX and CDS liquidity variables in the robustness test section. The VIX and CDS liquidity are the two variables that have the highest correlation with the dependent variable (0.46 and 0.41, respectively).

[23] Our results do not change materially when we proxy the global risk measure by means of the VDAX Index, the difference between LIBOR and Treasury Bill, the CDS indexes (iTraxx and CDX) or the square of the MSCI Index returns instead of the VIX index.

The dummy variable Max Quotes is equal to one when at least three data sources report a price and zero otherwise. The intuition is that the higher the number of quotes employed to calculate the cross-sectional standard deviation, the higher should this standard deviation be. This variable is significant and has a positive sign as expected.

To summarize, the empirical evidence strongly rejects Hypothesis 1. The volatility of the deviations from the common trend of the quoted prices provided by the different CDS data sources is not random but related to systematic factors. In other words, large deviations (in absolute value) from the common trend among databases do not appear randomly but are significantly related with risk and liquidity factors. The economic implication of this result is that, in specific market circumstances, the deviations of the prices from the common trend will tend to grow on average. Some prices will be closer to the trend and some prices will be far away from it but the average distance between them will increase, making the prices less homogeneous and making it more difficult for agents to asses the CDS fair value and for researcher using the data to decide what database gives the market prices' most reliable account. Also, model (1) does a pretty good job in explaining the dispersion among prices for the overall sample as measured by the R^2 (48%), and also for the European (37%) and US (46%) samples.[24]

As the bulk of the CDS spreads that we employ in our analysis are based on the information revealed by the traders or dealers, it is possible that the degree of divergence among the different data sources may be influenced by the number of contributors which are reporting quoted or traded CDS spreads. In an extreme case in which all the composite prices are constructed using the same group of contributors, the

[24] We also performed separate analysis before and during the crisis. We find that the explanatory variables referred to the trades are not significant before the crisis but they are significant and with the same signs as reported in Table 2 for the crisis period. We also considered the use of a crisis dummy but since the liquidity is much lower during the crisis and the number of trades in US is much lower during that period the use of the crisis dummy may cloud the effect of some of the potential explanatory variables that we use. Moreover, we use the VIX as a potential proxy for times of financial distress

prices should be very similar and the volatility of the deviations from the common trend should be close to zero. The problem is that we do not have access to the identity of the contributors that are reporting prices to the different data sources. However, we have access to the number of contributors that are reporting prices to Markit for the 5-year CDS spread. The different data sources may have different contributors but there should be some common group of contributors which presumably are the most influential traders and for that reason the most active agents in terms of contributed prices. Moreover, there could be other contributors whose participation is less significant in the sense that they report prices less frequently, or they could report prices to a few data sources but not to the others. This could imply that when the number of contributors is small the prices might be provided by the most influential and active traders which, on the other hand, could be common to all the data sources. Therefore, we conjecture that the lower the number of contributors, the higher should be the importance of the common contributors and the lower the divergence from the common trend among the different data sources. To test this conjecture we include the variable "number of contributors" as an additional explanatory variable in equation (1) and run the corresponding regression. The results are shown in Column (2) of Table 3; Column (1) repeats the benchmark results from Table 2 for comparison purposes. The coefficient on the number of contributors is positive and significant which is consistent with our conjecture on the effect of the number of contributors. The coefficients for the remaining variables do not change materially in sign or in magnitude with respect to the ones obtained in the baseline regression (Column (1) of Table 3).

To test if these results are affected by possible collinearity due to the relatively high correlation between both the CDS bid-ask spread and the VIX Index with the number of contributors (-0.212 and -0.361, respectively) we repeat the previous regression but using as explanatory variable the residual of the regression of the number of contributors onto the VIX Index and the CDS bid-ask spread. The residual proxies the number of contributors net of the global risk and the illiquidity effect in the CDS market. These results are shown in Column (3) of Table 3.[25] The results are almost identical to the ones observed in Column (2) and consistent with our conjecture on the effect of the most relevant contributors, and also that collinearity between the three previous variables is not a serious issue in our case.

It should be noted that transactions are not necessarily made through the GFI platform, but they could occur in any other platform. The advantage of GFI data is not that it includes all the CDS contracts traded but that it is a transparent source in which the market participants can observe real transaction prices and not just quotes. Although there is no available data for all the transaction prices since the beginning of our whole sample, we can employ an additional information source for a shorter time period; namely, a "trade information warehouse" that captures the majority of information on CDS trades covering corporate and sovereign borrowers. This warehouse was established by the Depository Trust & Clearing Corporation (DTCC) which keeps a record of outstanding CDSs involving major dealers as counterparties. According to the DTCC calculations around 90-95% of the CDS trades are settled and confirmed through them. The DTCC does not provide all the trade details, which are private information, but it reports weekly data on the gross and net exposures and the number of CDS outstanding contracts on 1,000 corporate and sovereign borrowers. We have this weekly information for the 90 firms that constitute our sample from the 7th of November, 2008 to the last sample date (the 29th of March, 2010).

[25] In order to estimate the coefficients presented in Column (3) of Table 3 we use the bootstrap methodology to correct any potential bias in the standard errors due to the use of a generated regressor.

To test for the importance of trades in the deviations of the CDS prices, we substitute the trading controls employed in equation (1) by a weekly variable which reports the total number of outstanding CDS contracts traded on a given reference firm. This allows us to control for both the cumulative information on a given firm attending to the total number of contracts and the trend in trading activity. The hypothesis we test is whether a higher number of CDS contacts traded on a given reference firm lowers the volatility of the deviations from the common trend across data sources. We find that the total number of CDS contracts traded on a reference firm has a significant and negative effect on the dispersion between data sources while the signs and levels of significance of the other variables remain unchanged with respect to the ones observed in the baseline regression results (Column (1) of Table 2). The implication of this is that the higher the market activity, the lower is the volatility of the deviations from the common trend of the quoted prices provided by the different CDS data sources. This fact is obviously at odds with Hypothesis 1 being true. Additionally, given that we are employing daily information but this variable is constructed on a weekly frequency, we lagged the variable one week and obtain a significant negative coefficient on that variable while the signs and the significance of the other variables remain unchanged. We also use the number of weekly traded contracts, lagged one week, instead of the total number of outstanding traded contracts and obtain similar results.[26]

5.3 Regression Results: Hypothesis 2

Table 4 reports the results of testing Hypothesis 2 on price discovery analysis using quoted prices (transaction prices are too scarce to be included in the analysis). A statistical significance test for the null hypothesis that the estimated price discovery proportions GG_i are equal to 0.5, is also included. The test rejects the null in all cases

[26] Detailed results are available upon request.

with the exception of CMA vs Markit in Europe and Fenics vs Reuters also in Europe. Therefore in these two cases both databases contribute equally to the price discovery process. However, in all other cases the results indicate that there is a leader database and a follower database. CMA is the data source that contributes to a higher extent to the "formation of prices" with newer and more influential information, especially for the total sample and for the US sample, followed by Markit. As mentioned above, for European firms CMA and Markit are almost equally informative in terms of price discovery. The less informative database in this realm seems to be Fenics. The results strongly reject the hypothesis that the price discovery process is evenly spread among data bases, and therefore Hypothesis 2 is not supported by the data.

6. Robustness Tests and Extensions

In this section, we report the results of several checks of the test of Hypothesis 1 presented in Table 2. First, we deal with potential problems of endogeneity and multicollinearity. Second, we repeat the previous analysis for a sub sample of European firms and adding a new data source: JP Morgan. Third, we consider alternative econometric techniques: pooled regressions and Prais-Winsten regressions after filling the missing observations. Finally, we analyze the sensitivity of the results to different data transformations: (i) using as the dependent variable the ratio between the logarithm of the standard deviation of the CDS quotes and the logarithm of mean CDS spread; (ii) excluding first the Reuters EOD quotes and second the Fenics quotes; (iii) limiting the sample period up to December 2009, June 2009 and December 2007;[27] (iv) using single

[27] These alternative sampling periods are used to test whether the results may contain some bias due to the lack of observations in some data sources or due to the effect of the different rules for dealing with the collateral in the CDS contracts. We limit the sample up to December 2009 and June 2009 because the data obtained from Fenics and Markit are available up to such periods, respectively. We limit the sample up to December 2007 to take into account potential differences in terms of the standard underlying collateral which is used for the different data sources.

source datasets constructed without aggregating data (Fenics/GFI and JPMorgan); and (v) grouping the firms by sector.[28]

6.1. Multicollinearity and Endogeneity Tests

In order to deal with potential problems of multicollinearity and endogeneity derived from the use of both the VIX and CDS bid-ask variables, we run a series of panel regressions based on different variations of the baseline regression (1) whose results are reported in Table 2. First, we run an identical panel regression but omitting the VIX Index, the CDS bid-ask spread and both. The results, not presented to save space, are qualitatively very similar to those in Column (1) of Table 2 confirming the significance of the other explanatory variables and suggesting that endogeneity and collinearity are not a serious issue in our case. As expected, the explanatory power of the panel regressions is lower given that we are omitting two powerful explanatory variables: the VIX Index and the CDS bid-ask spread.

As an additional test for potential endogeneity between the standard deviation between the different data sources and the VIX and CDS bid-ask spread variables, we run a regression in which we use a one period (day) lag in both variables. This is a standard procedure to deal with potential endogeneity and we find similar results to the ones reported in Column (1) of Table 2.

Finally, to test whether the results are biased by collinearity reasons due to the high correlation between the bid-ask spread and the VIX we run regression (1) but instead of the VIX Index using as explanatory variable the residual of the regression of the VIX Index onto the bid-ask spread. The residual proxies the VIX net of the illiquidity effect in the CDS market. We also run the regression but instead of the CDS bid-ask spread using as explanatory variable the residual of the regression of the CDS bid-ask spread

[28] All the results of this section are available upon request.

onto the VIX Index. The residual proxies the illiquidity in the CDS market net of the global risk effect. These results are reported in Columns (2) and (3) of Table 5 and they are almost identical to those in Table 2, which are also reported in Column (1) of Table 5 for comparison. In order to estimate the coefficients presented in Columns (2) and (3) of Table 5 we use the bootstrap methodology to correct any potential bias in the standard errors due to the use of generated regressors. The results suggest that collinearity between the two previous variables is not a serious issue in our case.

6.2. Adding a new data source

Our previous analysis is based on five different data sources (GFI, Fenics, CMA, Markit and Reuters EOD). We did not employ the data from J.P. Morgan because data for US firms was not available. However, for the sake of robustness we repeat the previous analysis for the sub sample of European firms adding a new data source: JP Morgan. These data was employed by Mayordomo et al. (2011) in the analysis of arbitrage opportunities in the credit derivatives markets. This new analysis is developed attending to the sample length of JP Morgan, that is, we use observations from January 1, 2005 to August 13, 2009 for the different data sources. It should be remembered that the nature of these data is not exactly the same as the previous ones in the sense that it comes from a single dealer instead of a group of dealers.

First we run regression (1) but including the data for JP Morgan and find similar results to those in Column (2) of Table 2 for the European firms. If we also include variable for the number of contributors, its effect on the dependent variable is also positive. The lower the number of contributors, the lower is the discrepancy among the different data sources. We also find that collinearity and endogeneity are not a serious issue in our case.

Finally, we repeat the price discovery analysis for the six data sources and find that the CMA database leads the price discovery process with respect all other databases, including JP Morgan. The second more efficient data source is Markit which reflects credit risk more efficiently than JP Morgan, Fenics and Reuters EOD. The latter are all equally efficient.

6.3. Using other econometric methodologies

As a robustness test we repeat the previous analysis using alternative econometric techniques: pooled OLS regressions and Prais-Winsten regressions after filling the missing observations.

To test whether the assumption of firm fixed effects affect significantly the results, we pool all the data and run a pooled OLS regression.[29] The results imply that, the assumption of firm fixed effects does not have a major effect on the results.

Our data form an unbalanced panel and so, we also run a Prais-Winsten regression with correlated panels, corrected standard errors (PCSEs) and robust to heteroskedasticity, contemporaneous correlation across panels and serial autocorrelation within panels. The correlation within panels is treated as a first-order autocorrelation AR(1) and the coefficient of this process common to all the panels.[30,31] The only difference with respect to the baseline results is that the interaction variable is not significant.

6.4. Testing the robustness of the results to data transformations

---

[29] Detailed results are available on request

[30] Each element in the covariance matrix of the disturbances is computed with all available observations that are common to the two panels contributing to the covariance.

[31] The panel is unbalanced because we do not have information on some variables from the beginning of the sample. However, there are no missing values once we include the first realization of the series. There were some missing observations in the VIX Index across the 90 firms due to the US holidays (i. e.: third Monday in January and February, Last Monday in May, July 4, First Monday of September, Fourth Thursday in November, etc.). However, we exclude these days from our analysis. There were some missing values in the market capitalization variable which are related with holidays in the corresponding country. Nevertheless, due to the low variability in this variable, we substitute the missing data with the first previous day's data available.

The dependent variable that we employ in the previous analysis is defined in logs in order to limit the effect of potential outliers which could appear in the quoted spreads due to any mistake in the contributed prices. By using the logs we also limit potential problems derived from a skewed distribution given that the value of the mean is almost four times the value of the median. We repeat regression (1) using as the dependent variable the ratio between the logarithm of the standard deviation among the CDS quotes and the logarithm of mean across the CDS quotes. The results are almost identical to the ones reported in Table 2.

The data obtained from Reuters EOD are available from December 2007 whereas the remaining data sources have information starting from January 2004. To avoid any potential bias due to the different length of the sample period covered by the different data sources we repeat the previous analysis without including the Reuters EOD quotes. We do not report these results to save space but they are almost identical to the ones reported in Table 2.

The data obtained from Fenics and Markit are available up to June 2009 and December 2009, respectively. To test if the results are biased by the lack of date in a given data source after a given date, we estimate equation (1) using data first up to June 2009, and after up to December 2009. The sign and magnitude of the coefficients are very similar to the ones reported in the first column of Table 2 and are available upon request.

All the data sources but Fenics are based on the traders or dealers prices. As was mentioned in Section 3, Fenics data can be actual trades or mid prices calculated from the bid/offer quotes. If none of these are available, GFI, which is the responsible of the Fenics quotes, calculates the CDS spread using the Hull and White methodology to ensure a credit curve always exists for each reference entity. Thus, we repeat regression (1) using as the dependent variable the logarithm of the standard deviation among the

CMA, Markit, Reuters EOD and JPMorgan quotes (excluding Fenics) on the corresponding explanatory variables.[32] Results are consistent with the ones obtained when we include Fenics in our analysis.

Since the beginning of the financial crisis counterparty risk in the CDS contracts has been partially mitigated through the use of collateralization. Actually, full collateralization of CDS liabilities has become the market standard. The ISDA Margin Survey 2009 reports that 74 percent of CDS contracts executed during 2008 were subject to collateral agreements. In order to limit any potential difference in the use of this collateral by the CDS data source we repeat the same analysis using a sub sample which spans up to December 2007 given that the use of the collateral was more limited before 2008.[33] The results do not materially differ from the ones reported in the first column of Table 2. The only significant difference is that the coefficient of interaction of the number of days without a trade one day before the current date and the trade dummy, although with a positive sign is not significant now (p-value = 0.23).

One problem of using composite prices is that we do not know who the contributors are and nor how these prices are obtained. We only observe the final price which is obtained by averaging different dealers quoted and traded prices. However, for a sample of European firms we observe JPM quoted and traded prices and Fenics constructed, quoted and traded (by means of GFI) prices although in any of the two cases we cannot distinguish between traded and quoted/constructed spreads. Both JPM and Fenics CDS spreads are obtained from single sources and not by aggregating data. As an additional robustness analysis, we test if the previous results are maintained when we compare

[32] We restrict our analysis to the European subsample in which we have information on JPM given that the use of the whole sample imply that the standard deviations across quotes calculated in the period before the crisis is obtained with just two contributors (CMA and Markit).

[33] One of the key developments in restoring market confidence was Intercontinental Exchange's (ICE) introduction of CDS clearing in March 2009. Looking forward, we aim to test what are the implications of the central clearing on the issues raised in this paper and to verify whether we still need to worry about data quality issued from different sources.

prices obtained from two single (not composite) sources: the most active inter-dealer broker (GFI/Fenics) and the most active broker (JP Morgan). We focus our analysis on the results reported in Table 2 and run a regression of the difference between the 5-year JPM and Fenics CDS spreads both in absolute and relative terms on the same explanatory variables that are employed in equation (1). The difference between JPM and Fenics CDS spread in relative terms is obtained as the absolute difference between both data sources divided by the mean between JPM and Fenics spreads. Since there is a high correlation between the CDS bid-ask spread and the VIX Index (0.714) for the cases in which we have observations on both JPM and Fenics we include only one of these variable in the regression. Results are shown in Table 6. Even when we compare data sources which are formed individually without attending to a conglomerate of traders, the differences persist and can be explained by the same variables as the baseline case in Table 2, independently of whether the difference between the two quotes are reported in absolute (Columns (1) and (2)) or relative terms (Columns (3) and (4)).

7. Conclusions

We study the consistency of the six most widely used CDS data bases: GFI, Fenics, Reuters EOD, CMA, Markit and JP Morgan, for the period from 2004 to 2010 using the most liquid single name 5-year CDS of the components of the leading market indexes, iTraxx (European firms) and CDX (US firms). We find that there are significant differences among them in several dimensions.

Our main empirical findings are:

1) When timely information on traded prices is available, the different price sources largely agree among them in general terms. However as the information on transaction prices become scarcer, prices from different sources tend to diverge from the

common trend. The most extreme disagreements are in the case of American reference entities during the crisis, where very few transaction prices are available in the GFI database.

2) Deviations (in absolute value) from the common trend among the different CDS quoted spreads are not purely random but are related to idiosyncratic factors such as firm size and also to liquidity, global risk and trading factors. Prices tend to diverge more from the common trend in the case of for smaller firms. Increases in market illiquidity, idiosyncratic stock market volatility and global volatility increase the divergence from the common trend among prices coming from different data bases.

3) CMA quoted CDS spreads led the credit risk price discovery process with respect to the quotes provided by the other databases.

Extensive robustness tests support these results. Since our analysis is based on the most liquid CDS prices, we would expect that the differences we find for these prices in the different databases would be even larger for less liquid CDSs not included in our study.

Our analysis has important implications for research studies and industry participants. First, for US names with low trade frequency, no reliable information exists because there are almost no recorded trade prices in the GFI platform. Second, in studies of price discovery of the CDS market with respect other markets, given that there is a data source (CMA) leading the others, empirical results may change depending on the database employed. Third, the smaller the firm, the higher is the volatility of the deviations from the common trend of the quoted prices provided by the different CDS data sources and therefore the less reliable and comparable research results might be. Fourth, in times of high illiquidity or increased stock market volatility CDS prices from different databases will tend to substantially diverge from the common trend making it more difficult for agents to disentangle the CDS fair value from the different prices they receive from the databases and for researchers using the data to decide what database gives the market prices' most reliable account.

Looking forward, the analysis of how the discrepancy among the different CDS contributors may affect the relation between the CDS and corporate Bond spreads is a topic worth studying. Also the consequences of using different CDS sources on testing the degree of informational efficiency of the different markets where credit risk is traded is also an interesting avenue for future research.

References

Acharya, V. and Johnson, T. (2007), "Insider Trading in Credit Derivatives". *Journal of Financial Economics* 84, 110-141.

Arora, N., Gandhi, P., and Longstaff, F. (2009), "Counterparty Credit Risk and the Credit Default Swap Market". Working Paper, UCLA.

Aunon-Nerin, D., Cossin, D., Hricko, T. and Huang, Z. (2002), "Exploring for the Determinants of Credit Risk in Credit Default Swap Transaction Data: Is Fixed-Income Markets' Information Sufficient to Evaluate Credit Risk?". Working paper, HEC-University of Lausanne and FAME.

Bennin, R. (1980) "Error rates in CRSP and Compustat: a second look". *Journal of Finance*, 35, 1267-1271.

Blanco, R., Brennan, S. and Marsh, I. W. (2005), "An Empirical Analysis of the Dynamic Relationship between Investment Grade Bonds and Credit Default Swaps". *Journal of Finance* 60, 2255-2281.

Cao, C., Yu, F., and Zhong, Z. (2010), "The information content of option-implied volatility for credit default swap valuation". *Journal of Financial Markets*, Forthcoming.

Chen, R.-R., Cheng, X., Liu, B. (2008), "Estimation and evaluation of the term structure of credit default swaps: An empirical study". *Insurance: Mathematics and Economics*, 43, 339-349.

Christoffersen, P. F., Ericsson, J., Jacobs, K., and Jin, X. (2009), "Exploring Dynamic Default Dependence". Available at SSRN: http://ssrn.com/abstract=1400427.

Fulop, A. and Lescourret, L. (2007), "An Analysis of Intra-Daily Patterns on the CDS Market". Working Paper, ESSEC Business School.

Gonzalo, J. and Granger, C.W.J. (1995), "Estimation of Common Long-Memory Components in Cointegrated Systems". *Journal of Business and Economics Statistics* 13, 27-35.

Hull, J., Predescu, M. and White, A. (2004), "The Relationship between Credit Default Swap Spreads, Bond Yields, and Credit Rating Announcements". *Journal of Banking and Finance* 28, 2789-2811.

Jorion, P., and Zhang, G. (2007), "Good and bad credit contagion: Evidence from credit default swaps". *Journal of Financial Economics*, Volume 84, Issue 3, June 2007, Pages 860-883.

Jorion, P., and Zhang, G. (2009), "Credit Contagion from Counterparty Risk". *Journal of Finance*, Volume 64, Number 5, pp. 2053-2087(35).

Leland , H. (2009), "Structural Models and the Credit Crisis". Presented at the China International Conference of Finance. July 8, 2009. http://haas.berkeley.edu/groups/finance/CHINA7.pdf

Longstaff, F. A, Mithal, S. and Neis, E. (2005), "Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market". *Journal of Finance* 60, 2213-2253.

Mayordomo, S., Peña, J.I., and Romo, J. (2009): "A New Test of Statistical Arbitrage with Applications to Credit Derivatives Markets". Working Paper CNMV.

Micu, M., Remolona, E., and Wooldridge, P. D. (2004) "The Pricing Impact of Rating Announcements: Evidence from the Credit Default Swap Market". *BIS Quarterly Review.*

Nashikkar, A. J., Subrahmanyam, M. G. and Mahanti, S. (2009), "Limited Arbitrage and Liquidity in the Market for Credit Risk". NYU WP No. FIN-08-011.

Nashikkar, A. J. and Subrahmanyam, M. G. (2007), "Latent Liquidity and Corporate Bond Yield Spreads". NYU Working Paper No. FIN-07-013.

Predescu, M. (2006), "The Performance of Structural Models of Default for Firms with Liquid CDS Spreads". Working Paper, Rothman School of Management, University of Toronto.

Rosenberg, B. and Houglet, M. (1974), "Error Rates in CRSP and COMPUSTAT Data Bases and Their Implications", *Journal of Finance*, 29, 1303-1310.

Schoar, A. S. (2002), "Effects of corporate diversification on productivity", *Journal of Finance*, 57, pp. 2379-2403.

Saita, L., (2006), "The Puzzling Price of Corporate Default Risk". Working paper, Standford Graduate School of Business.

Villalonga, B. (2004), "Diversification discount or premium? New evidence from BITS establishment-level data", *Journal of Finance*, 59, pp. 479-506.

Zhang, B. Y., Zhou, H., and Zhu, H. (2009) "Explaining Credit Default Swap Spreads with the Equity Volatility and Jump Risks of Individual Firms". *Review of Financial Studies* 2009 22(12):5099-5131.

Zhu, H. (2006), "An Empirical Comparison of Credit Spreads Between the Bond Market and the Credit Default Swap Market". *Journal of Financial Services Research* 29, 211-235.

Table 1: Firm Names by Sector and CDS Index (iTraxx and CDX)

This table shows the descriptive statistics for the single name 5-year CDS. Panel A shows the names classified by index and sector. We use European and American firms included in the iTraxx and the CDX indexes, respectively, over the whole sample period. Panel B provides aggregate CDS descriptive statistics for the European and American firms over the five data sources (GFI, CMA, Markit, Fenics, and Reuters). The information is divided before and during the crisis. As the actual sample size of the different data sources differ (because of missing values and slightly different periods covered), we report the summary statistics for the cases in which we have common observations (trades and quotes) in all the data sources in Panel C. Panel D reports the summary statistics for the cases in which there is no trade but there are quotes in all the data sources.

Panel A

iTraxx Firm Name	Ticker	Sector	CDX Firm Name	Ticker	Sector
AKZO Nobel NV	AKZO	Auto/Indust.	Alcoa Inc.	AA	Auto/Indust.
Bayer Aktiengesellschaft	BAYG	Auto/Indust.	Carnival Corporation	CCL	Auto/Indust.
Bayerische Motoren Werke AG	BMWG	Auto/Indust.	CSX Corporation	CSX	Auto/Indust.
Compagnie de Saint-Gobain	SGOB	Auto/Indust.	The Dow Chemical Company	DOW	Auto/Indust.
EADS NV	AERM	Auto/Indust.	Eastman Chemical Company	EMN	Auto/Indust.
Siemens Aktiengesellschaft	SIEG	Auto/Indust.	Honeywell International Inc	HON	Auto/Indust.
Volkswagen Aktiengesellschaft	VOWG	Auto/Indust.	Union Pacific Corporation	UNP	Auto/Indust.
Aktiebolaget Volvo	VOLV	Auto/Indust.			
Accor	ACCP	Consumers	Altria Group, Inc.	MO	Consumers
British American Tobacco PLC	BATS	Consumers	AutoZone, Inc.	AZO	Consumers
Carrefour	CARR	Consumers	Baxter International Inc.	BAX	Consumers
Marks and Spencer PLC	MKSA	Consumers	Bristol-Myers Squibb Company	BMY	Consumers
LVMH Moet Hennessy Louis Vuitton	LVMH	Consumers	Campbell Soup Company	CPB	Consumers
Metro AG	METB	Consumers	Cardinal Health, Inc.	CAH	Consumers
Koninklijke Philips Electronics NV	PHG	Consumers	Loews Corporation	LTR	Consumers
PPR	PRTP	Consumers	Safeway Inc.	SWY	Consumers
Sodexho Alliance	SODE	Consumers	Southwest Airlines Co.	LUV	Consumers
Unilever NV	UN	Consumers	The Walt Disney Company	DIS	Consumers
			Whirlpool Corporation	WHR	Consumers
Edison SPA	EDN	Energy	Anadarko Petroleum Corporation	APC	Energy
Electricite de France	EDF	Energy	Arrow Electronics, Inc.	ARW	Energy
EnBW Energie Baden-Wuerttemberg	EBKG	Energy	ConocoPhillips	COP	Energy
Enel SPA	ENEI	Energy	Constellation Energy Group, Inc.	CEG	Energy
EDP - Energias de Portugal SA	EDP	Energy	Devon Energy Corporation	DVN	Energy
E.ON AG	EONG	Energy	Dominion Resources, Inc.	D	Energy
Fortum Oyj	FUMC	Energy	Progress Energy, Inc.	PGN	Energy
Iberdrola SA	IBE	Energy	Sempra Energy	SRE	Energy
Repsol YPF SA	REP	Energy	Transocean Inc.	RIG	Energy
RWE Aktiengesellschaft	RWEG	Energy	Valero Energy Corporation	VLO	Energy
GDF Suez	GDF	Energy			
Veolia Environnement	VIE	Energy			
Aegon NV	AEGN	Financials	Ace Limited	ACE	Financials
AXA	AXAF	Financials	American Express Company	AXP	Financials
Barclays Bank PLC	BCSB	Financials	American International Group, Inc.	AIG	Financials
Commerzbank Aktiengesellschaft	CBKG	Financials	Boeing Capital Corporation	BA	Financials
Deutsche Bank Aktiengesellschaft	DB	Financials	Cigna Corporation	CI	Financials
Hannover Rueckversicherung AG	HNRG	Financials	General Electric Capital Corporation	GE	Financials
Banca Monte Dei Paschi Di Siena Spa	BMPS	Financials	Marsh & McLennan, Inc.	MMC	Financials
Muenchener Rueckversicherung	MUVG	Financials	Simon Property Group, L.P.	SPG	Financials
Swiss Reinsurance Company	RUKN	Financials	Wells Fargo & Company	WFC	Financials
			XL Capital Ltd.	XL	Financials
Bertelsmann AG	BTGG	TMT	AT&T Inc.	T	TMT
Deutsche Telekom AG	DTA	TMT	CenturyTel, Inc.	CTL	TMT
France Telecom	FTE	TMT	Comcast Cable Communications, LLC	CMCC	TMT
Hellenic Telecommunications	OTE	TMT	Omnicom Group Inc.	OMC	TMT
Koninklijke KPN NV	KPN	TMT	Time Warner Inc.	TWX	TMT
Telecom Italia SPA	TLIT	TMT			
Telefonica SA	TEF	TMT			
Vodafone Group PLC	VOD	TMT			

Panel B: Using all the observations

Europe	Before 9th August 2007: Average Number of Quotes or Trades	Mean	S.D.	Median	After 9th August 2007: Average Number of Quotes or Trades	Mean	S.D.	Median
GFI	331	35	20	32	104	95	71	85
CMA	905	31	20	27	664	105	89	78
Markit	869	30	18	27	575	110	94	80
Fenics	874	31	20	27	462	111	101	78
Reuters					527	118	93	89
US	**Average Number of Quotes or Trades**	**Mean**	**S.D.**	**Median**	**Average Number of Quotes or Trades**	**Mean**	**S.D.**	**Median**
GFI	140	50	33	47	14	99	119	67
CMA	905	40	27	34	664	128	186	76
Markit	884	38	23	34	585	134	195	79
Fenics	877	39	24	35	473	120	209	74
Reuters					533	148	203	83

Panel C: Using the observations in the days in which there is a trade and quotes in all the data sources

Europe	Before 9th August 2007: Average Number of Trades and Quotes	Mean	S.D.	Median	After 9th August 2007: Average Number of Trades and Quotes	Mean	S.D.	Median
GFI	306	36	17	33	51	128	82	105
CMA	306	35	16	32	51	128	82	104
Markit	306	35	16	32	51	128	81	104
Fenics	306	36	16	32	51	128	82	105
Reuters					51	127	81	103
US	**Average Number of Trades and Quotes**	**Mean**	**S.D.**	**Median**	**Average Number of Trades and Quotes**	**Mean**	**S.D.**	**Median**
GFI	122	54	36	44	11	131	129	95
CMA	122	52	36	44	11	149	124	111
Markit	122	52	36	44	11	148	122	113
Fenics	122	51	37	44	11	130	115	92
Reuters					11	147	123	109

Panel D: Using the observations in the days in which there is not a trade but quotes in all the data sources

Europe	Before 9th August 2007: Average Number Quotes	Mean	S.D.	Median	After 9th August 2007: Average Number Quotes	Mean	S.D.	Median
CMA	553	28	19	25	264	135	114	96
Markit	553	28	19	25	264	135	113	96
Fenics	553	28	19	25	264	136	114	97
Reuters					264	135	113	96
US	**Average Number Quotes**	**Mean**	**S.D.**	**Median**	**Average Number Quotes**	**Mean**	**S.D.**	**Median**
CMA	749	36	20	33	339	151	224	91
Markit	749	36	20	33	339	153	224	93
Fenics	749	37	21	34	339	140	239	88
Reuters					339	147	230	86

Table 2: Determinants of the standard deviation among the CDS data sources

This table reports the regression coefficients of the unbalanced panel regressions. The dependent variable is the standard deviation among the different CDS data sources (CMA, Markit, Fenics, Reuters EOD). The database includes ninety European and US firms (47 of the firms are European and the rest are American) which are the most liquid CDSs included in either the Itraxx or the CDX Index since the launching of the indexes, from January 2004 to April 2010. The estimation uses a fixed-effects model robust to heteroskedasticity. Column (1) reports the results for the whole sample of firms, Column (2) reports the results for the subsample of European firms, and Column (3) reports the results for the subsample of American firms. The *t-statistics* are reported between brackets.

		(1)	(2)	(3)
Log(Mkt. Cap.)		-0.095	-0.048	-0.187
		(-10.53)	(-5.91)	(-12.17)
Trade		-0.079	-0.061	-0.078
		(-10.01)	(-7.11)	(-4.95)
Days w/o a trade		0.0003	0.0000	0.0004
		(10.80)	(1.05)	(13.21)
Interaction Trade and Days w/o trade		-0.0006	0.0003	-0.0013
		(-1.97)	(0.61)	(-3.47)
CDS Bid-Ask Spread		0.041	0.075	0.030
		(14.76)	(34.77)	(11.39)
VIX Index		0.045	0.032	0.051
		(64.85)	(49.05)	(77.25)
Max Quotes		0.290	0.080	0.565
		(32.86)	(7.86)	(37.42)
Constant		0.938	-0.370	3.327
		(4.30)	(-1.89)	(9.00)
R-squared		0.481	0.371	0.458
Number of observations		138653	71605	67048
Number of groups		90	47	43
Observations per group	Minimum	940	940	1150
	Average	1541	1524	1559
	Maximum	1569	1569	1569
F-statistic		6922.270	4315.440	3337.690
Prob. > F-statistic		0	0	0
Condition Index		6.64	7.60	6.19

Table 3: Determinants of the standard deviation among the CDS data sources using the number of contributors as an explanatory variable

This table reports the regression coefficients of the unbalanced panel regressions. The dependent variable is the standard deviation among the different CDS data sources (CMA, Markit, Fenics, Reuters EOD). The database includes ninety European and US firms (47 of the firms are European and the rest are American) which are the most liquid CDSs included in either the Itraxx or the CDX Index since the launching of the indexes, from January 2004 to April 2010. The estimation uses a fixed-effects model robust to heteroskedasticity. Column (1) reports the baseline regression's results for the whole sample of firms without using the number of contributors as an explanatory variable. Column (2) reports the results obtained by adding the number of contributors as an additional explanatory variable to the ones in Column (1). Column (3) reports the results obtained using as an additional explanatory variable a generated regressor which is obtained after regressing the number of contributors on the VIX Index and the CDS bid-ask spread and then using the residual to proxy the number of contributors net of the global risk and the illiquidity effect in the CDS market. In order to estimate the coefficients presented in Column (3) of this table we use the bootsptrap methodology to correct any potential bias in the standard errors due to the use of a generated regressor. The *t-statistics* are reported between brackets.

		(1)	(2)	(3)
Log(Mkt. Cap.)		-0.095	-0.105	-0.105
		(-10.53)	(-10.33)	(-10.27)
Trade		-0.079	-0.134	-0.134
		(-10.01)	(-16.75)	(-15.95)
Days w/o a trade		0.0003	0.0004	0.0004
		(10.80)	(14.61)	(14.73)
Interaction Trade and Days w/o trade		-0.001	-0.0003	-0.0003
		(-1.97)	(-1.09)	(-1.10)
CDS Bid-Ask Spread		0.041	0.041	0.041
		(14.76)	(14.33)	(13.70)
VIX Index		0.045	0.048	0.046
		(64.85)	(66.66)	(63.80)
Max Quotes		0.290	0.338	0.338
		(32.86)	(29.74)	(30.23)
Number of Contributors			0.014	0.014
			(21.33)	(19.86)
Constant		0.938	0.854	1.101
		(4.30)	(3.49)	(4.46)
R-squared		0.481	0.495	0.495
Number of observations		138653	128179	128179
Number of groups		90	90	90
Observations per group	Minimum	940	332	332
	Average	1541	1424	1424
	Maximum	1569	1492	1492
F-statistic		6922.270	5870.500	
Prob. > F-statistic		0	0	
Wald Chi2				47673.88
Prob. > Wald Chi2				0
Condition Index		6.64	10.67	10.67

Table 4: Price Discovery Analysis by Pairs of CDS spreads

This table reports the results of the price discovery analysis. First, we estimate the Gonzalo-Granger (GG) price discovery metrics for different pairs of 5-year single name CDS spreads using different data sources. The estimations are based on a VECM in which the VAR-length is selected according to the Schwarz information criteria. Then we calculate the average Gonzalo-Granger metric for all the firms, the European and the American firms for the different pairs of data sources. When the price discovery metric is higher than 0.5, the corresponding data source leads the price discovery process. The symbols ***, **, and * (^^^, ^^, and ^) summarize the statistical significance test and indicate that the average price discovery metric (GG) corresponding to a given data source is significantly higher (lower) than 0.5 at a significance level of 99, 95 and 90%, respectively.

CMA versus Markit GG Price Discovery Metrics

	Total	Europe	US
CMA	0.574**	0.502	0.660***
Markit	0.426^^	0.498	0.340^^^

CMA versus Fenics GG Price Discovery Metrics

	Total	Europe	US
CMA	0.734***	0.754***	0.708***
Fenics	0.266^^^	0.246^^^	0.292^^^

CMA versus Reuters EOD GG Price Discovery Metrics

	Total	Europe	US
CMA	0.798***	0.859***	0.718***
Reuters EOD	0.202^^^	0.141^^^	0.282^^^

Markit versus Fenics GG Price Discovery Metrics

	Total	Europe	US
Markit	0.771***	0.800***	0.735***
Fenics	0.229^^^	0.200^^^	0.265^^^

Markit versus Reuters EOD GG Price Discovery Metrics

	Total	Europe	US
Markit	0.783***	0.893***	0.644***
Reuters EOD	0.217^^^	0.107^^^	0.356^^^

Fenics versus Reuters EOD GG Price Discovery Metrics

	Total	Europe	US
Fenics	0.398^^^	0.461	0.318^^^
Reuters EOD	0.602***	0.539	0.682***

Table 5: Determinants of the standard deviation among the CDS data sources using proxies for the VIX Index and the CDS illiquidity measure

This table reports the regression coefficients of the unbalanced panel regressions. The dependent variable is the standard deviation among the different CDS data sources (CMA, Markit, Fenics, Reuters EOD). The database includes ninety European and US firms (47 of the firms are European and the rest are American) which are the most liquid CDSs included in either the Itraxx or the CDX Index since the launching of the indexes, from January 2004 to April 2010. The estimation uses a fixed-effects model robust to heteroskedasticity. Column (1) reports the baseline regression results which are the same as in Column (1) of Table 2. Column (2) provides the results obtained when we use as an explanatory variable a generated regressor which is obtained as the residual of a regression in which the VIX Index is regressed on the CDS bid-ask spread. Column (3) reports the results obtained when we use as an explanatory variable a generated regressor which is obtained as the residual of a regression in which the CDS bid-ask spread is regressed on the VIX Index. In order to estimate the coefficients presented in Columns (2) and (3) we use the bootstrap methodology to correct any potential bias in the standard errors due to the use of generated regressors. The *t-statistics* are reported between brackets.

		(1)	(2)	(3)
Log(Mkt. Cap.)		-0.095	-0.095	-0.095
		(-10.53)	(-10.18)	(-10.15)
Trade		-0.079	-0.079	-0.079
		(-10.01)	(-10.05)	(-11.27)
Days w/o a trade		0.0003	0.0003	0.0003
		(10.80)	(11.54)	(11.29)
Interaction Trade and Days w/o trade		-0.0006	-0.0006	-0.0006
		(-1.97)	(-2.12)	(-1.81)
CDS Bid-Ask Spread		0.041	0.082	
		(14.76)	(38.96)	
VIX Index		0.045		0.057
		(64.85)		(206.26)
VIX Index net of the CDS Bid-Ask Spread effect			0.045	
			(64.96)	
CDS Bid-Ask Spread net of the VIX Index effect				0.041
				(13.54)
Max Quotes		0.290	0.290	0.290
		(32.86)	(33.99)	(31.30)
Constant		0.938	1.642	0.928
		(4.30)	(7.09)	(4.11)
R-squared		0.481	0.482	0.482
Number of observations		138653	138653	138653
Number of groups		90	90	90
Observations per group	Minimum	940	940	940
	Average	1541	1540.6	1540.6
	Maximum	1569	1569	1569
F-statistic		6922.270		
Prob. > F-statistic		0.000		
Wald Chi2			40905.47	60080.36
Prob. > Wald Chi2			0.000	0.000
Condition Index		6.64	5.06	6.1

Table 6: Determinants of the differences between JPM and Fenics CDS spreads

This table reports the regression coefficients of the unbalanced panel regressions. The database includes ninety European and US firms (47 of the firms are European and the rest are American) which are the most liquid CDSs included in either the Itraxx or the CDX Index since the launching of the indexes, from January 2004 to April 2010. The estimation uses a fixed-effects model robust to heteroskedasticity. Columns (1) and (2) report the coefficients of the determinants of the JPM and Fenics CDS spreads difference in absolute terms when we exclude the CDS Bid-Ask Spread and the VIX Index variables, respectively. Columns (3) and (4) provide the coefficients of the determinants of the JPM and Fenics CDS difference in relative terms when we exclude the CDS Bid-Ask Spread and the VIX Index variables, respectively. The difference in relative terms is obtained as the ratio between the difference in absolute terms and the mean between JPM and Fenics spreads. The *t-statistics* are reported between brackets.

		(1)	(2)	(3)	(4)
Log(Mkt. Cap.)		-2.325	0.029	-0.012	-0.009
		(-18.31)	(0.26)	(-10.41)	(-7.37)
Trade		-0.634	-0.019	-0.006	-0.005
		(-9.05)	(-0.25)	(-9.67)	(-7.96)
Days w/o a trade		0.016	0.012	0.000	0.000
		(14.04)	(11.14)	(13.55)	(13.14)
Interaction Trade and Days w/o trade		-0.004	-0.012	0.00002	0.00001
		(-0.68)	(-1.08)	(0.74)	(0.26)
VIX Index		0.3171		0.0003	
		(42.29)		(11.38)	
CDS Bid-Ask Spread			1.0290		0.0013
			(32.18)		(13.36)
Constant		53.738	-2.658	0.330	0.252
		(17.74)	(-0.98)	(12.00)	(8.91)
R-squared		0.203	0.320	0.108	0.115
Number of observations		46772	46772	46772	46772
Number of groups		43	43	43	43
Observations per group	Minimum	891	891	891	891
	Average	1088	1088	1088	1088
	Maximum	1149	1149	1149	1149
F-statistic		369.480	250.080	114.520	109.030
Prob. > F-statistic		0	0	0	0
Condition Index		4.40	2.99	4.40	2.99

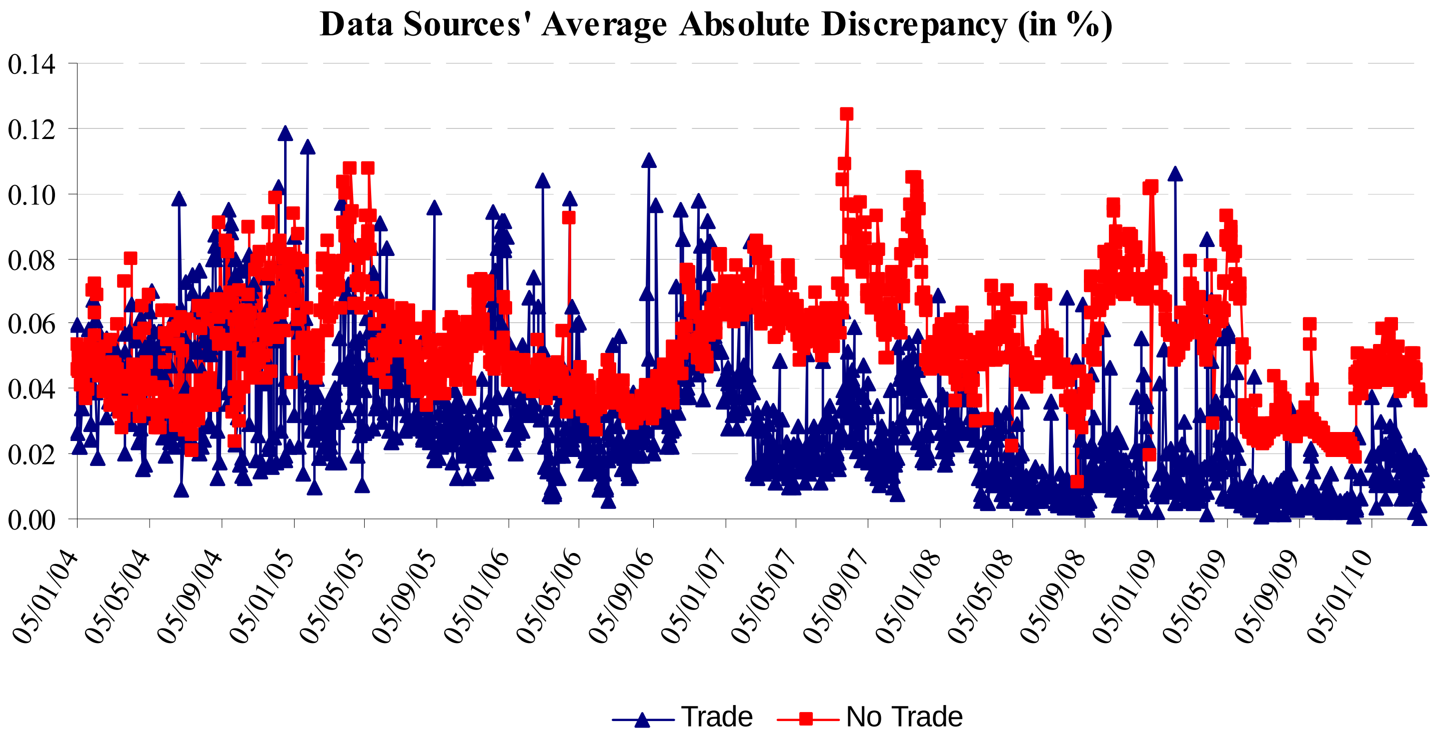

Figure 1. Data Sources' Average Absolute Discrepancies (AAD). This figure shows the cross-sectional deviations across data sources over time. The series is computed as the absolute value of the average difference across pairs of data sources (CMA - Markit, CMA - Fenics, CMA - Reuters, Markit - Fenics, Markit - Reuters, Fenics – Reuters) divided by the average CDS spread across the four previous data sources (CMA, Markit, Fenics, and Reuters) for each firm. Then, we calculate the average of the previous series date by date across the total number of entities.