markit™

CDS Data Cleaning Process

November 2011

Contents

Overview

Markit's CDS pricing service provides an independent source of pricing data covering global-corporate, municipal, and sovereign single name CDS. Daily historical data is usually available going back to January 2001.

Markit's CDS pricing service aggregates contributor data to provide high quality data that meets the needs of market participants such as buy-side, sell-side, regulators, or alliance partners.

The result is an independent and rigorous data set, appropriate for (without limitation) research, analytics, risk and marking-to-market.

Daily Pricing

Markit receives contributed pricing from many market makers (which can include the G14 banks) and publishes daily consensus pricing on approximately 2650 individual entities and 3000 entity-tiers. Markit applies rigorous data cleaning tests to each curve in order to provide an independent, accurate and reliable data set.

Markit calculates a mathematical average of all contributed prices and spread data for a given instrument type, entity, tier, maturity, currency, and restructuring type (the "composite price"). The composite price is the average of the prices provided to Markit by its contributors once those prices failing any one of the data quality tests have been excluded.

In order to form a composite, Markit requires at least 3 distinct contributors submitting curves of which at least 2 pass all data cleaning tests.

Inputs

Markit currently receives marked par spread curves and recovery rates from its contributor network. These par spread curves represent the true replacement value of the CDS (taking into account curve shape and marked recovery).

Markit converts the par spreads to points upfront using the contributed recovery rate for each curve so the cleaning tests apply towards a common unit. Data cleaning tests are applied to curves in upfront price type. Composites are initially formed in points upfront, which are then converted into par spreads using the composite recovery rate. The composite recovery rate is the average of the recovery rates of the curves which are used to form the composite.

Markit uses the ISDA CDS Standard Model to convert between points upfront and par spreads. This conversion uses a hazard rate term structure that is consistent with the upfront and par spread curve. Also, the conversion uses the full curve and real recovery rates rather than a flat hazard rate and the assumed recovery rate.

Upfront composites are also converted in to conventional spreads (also known as flat/ISDA/quoted spread). Similar to the conversion from upfront to par spread, this conversion also follows the ISDA CDS Standard Model but uses a flat hazard rate and an assumed recovery rate based upon the standard transaction type's assumed recovery.

More information about standard setup can be found at http://www.cdsmodel.com/cdsmodel/fee-computations.page. Additional conversion details can be found in the Appendix.

Data Cleaning Tests

All contributed data is subjected to a range of rigorous tests with the intention of ensuring that only the highest quality data is used to form composite prices. Data that fails any of the tests is excluded from the calculation of the composite. An indication of the stringency of these tests is that Markit regularly rejects in excess of 50% of the data received.

The four primary tests applied to all contributed CDS data are combined into two main categories:

1. Logic Tests ensure that each contributor's data follow standard pricing mechanisms:
    a. Curve Buildability Test; and
    b. Backwardation Test.

2. Relative Tests compare data across contributors to create a robust consensus pricing level:
    a. Stale Data Test; and
    b. Outlier Test.

Curve buildability test
The curve buildability test checks for valid survival probabilities along the contributed curves using the bootstrapping method in the ISDA CDS Standard Model. During this process, the yield curve (interest rate curve) that corresponds to the default currency of the reference entity and the par spread curve (with the respective recovery rate) are used to find the survival probability at each tenor. Curves which result in unreasonable survival probabilities are rejected.

Backwardation test
For a given contributor, the backwardation test checks the expected relationship across restructuring types (docclause) for submissions on each entity-tier. Specifically, it is intended to validate that the following inequality holds true for a given contributor's spreads:

CR >= MM >= MR >= XR

If a contributor sends curves that do not conform to this rule, the non-standard docclause curves are rejected. The curve that is retained is the one that matches the docclause believed to be most commonly traded for the given issuer (the standard/default curve). In addition, the test also checks that the basis between a curve in the standard docclause and curves in other docclauses for the same currency are not too wide relative to observed average levels.

Stale data test
One challenge encountered with respect to incoming data is that it can relate to credits in which an institution previously had, but no longer has, a material position. As the levels in these curves have minimal or no effect on an institution's profit or loss, traders tend to update them infrequently or not at all. Therefore the data becomes stale. Such stale data should be excluded.

For each contributed curve, Markit calculates the number of days the 5Y CDS spread has been unchanged. Curves from all dealers are ranked in terms of their 5Y staledays and Markit generally passes the top 50% of curves for illiquid entities and top 67% of curves for liquid entities. An entity is defined to be illiquid if it receives contributions from 13 or fewer contributing banks; otherwise it is defined to be liquid. The remaining curves fail the test and are not used in the composite calculation.

Note that the stale data test is relative. The following are a few important points on the stale data test:

- The stale test always passes at least three curves, given there is a minimum of 3 curves tested;
- Whole curves can either pass or fail the stale test; and
- All curves deemed stale are excluded from the composite calculation on that particular date.

Outlier Test
Outlying prices may occur for many reasons, including:

- A trader has not updated her curve recently and the market has moved;
- A curve is incorrectly mapped to the Markit entity name; and
- A trader genuinely has a different opinion of the value of the CDS.

The outlier test works by first calculating a provisional median curve based on all contributions that passed the previous tests i.e. curve buildability, backwardation, and stale data tests. The outlier test then computes a weighted sum of squared deviations across tenors and recovery rate of each contributed curve from the provisional median curve. It then ranks the curves and rejects the ones with highest deviations.

Note that a non-default curve will pass the outlier test if the default curve belonging to the same contributor passes, provided it does not deviate from the default curve more than a certain tolerance level. Furthermore, if the default curve fails the outlier test, then the non-default curve from the same dealer also fails.

Calculating Composites

Composites are calculated only for entity-tiers that have 3 submitted and 2 passed contributor curves (from distinct dealers). Composites calculation has three major steps:

1. Calculate composites for the default docclause and currency group;
2. Calculate docclause and currency group adjustment factors from the default curve to non-default curves, provided there are at least two dealers who have contributed and passed for both curves. An adjustment factor is the average (across dealers) of the basis observed between default and non-default curves. Note that a zero basis will be ignored when calculating the adjustment factors; only those curves of contributors who have marked a non-zero basis will be used; and
3. Calculate all non default curves by taking the default curve and applying the adjustment factors. Non default curves are only calculated for docclause-currency combinations that have a certain number of contributor curves submitted for them.

Further Information

For further information, please contact Markit via email at support@markit.com or telephone:

- Europe - 00800 6275 4800
- North America - +1 877 762 7548
- Asia - +65 6499 0079
- Japan - +81 3 6402 0127

Appendix

Par Spread-Upfront Conversion

Each contributed par spread curve must have a corresponding recovery rate so that it can be fed into data cleaning tests. This is because these tests are carried out in the upfront space and a recovery rate is required for par spread-upfront conversion.

Markit uses the bootstrapping method to obtain the survival probability for each maturity using the par spreads, recovery rate and interest rate curve as inputs into the following equation:

$$PV(fee\ leg) = PV(default\ leg) \tag{1}$$

Both sides of this equation are discounted (using the interest rate curve) and weighted by the survival probability of the underlying reference entity at each tenor. Using the inputs mentioned above, the survival probability for each tenor is derived iteratively. The bootstrapping method begins by using the *6M* par spread to solve the equation above which will result in the hazard rate for this tenor. Then the survival probability is found using:

$$Prob(surviving\ period\ [0, 6M]) = 1 - hazard\ rate(6M)$$

Next, the *1Y* spread and *6M* hazard rate are fed into equation *(1)* to calculate the *1Y* hazard rate. This is then used in the equation below to calculate the survival probability for *1Y*:

$$Prob(surviving\ period\ [0, 1Y]) = (1 - hazard\ rate(6M))(1 - hazard\ rate(1Y))$$

One can generalise this equation to obtain the survival probability for any tenor *T*:

$$Prob(surviving\ [0, T]) = \big(1 - hazard\ rate(6M)\big)\big(1 - hazard\ rate(1Y)\big) \ldots (1 - hazard\ rate(T)) \tag{2}$$

This method is repeated until the survival probability of all tenors is obtained. The shape of this survival probability curve is checked in the ‘Curve Buildability’ test. Once the survival probability curve passes this test, it is used to find the upfront for each one of the tenors.

The upfront payment in a CDS contract is simply the present value of the difference between the payments made using a running spread (par spread) and fixed coupon over the life of the deal. One should take the present value of this difference since the upfront is a one off payment at inception. Markit uses the following equation to convert from par spread to upfront at each tenor:

$$Upfront(tenor\ T) = PV(Par\ Spread(T) - fixed\ coupon)$$

Alternatively, one could take this difference and multiply it by the expected life of the CDS contract, namely *PV01*, to get the upfront:

$$Upfront(tenor\ T) = PV01(Par\ Spread(T) - fixed\ coupon) \tag{3}$$

The interest rate and the survival probability curves can be used to find *PV01* at each tenor. The upfront at tenor *T* can then simply be found by feeding *PV01* (at tenor *T*), par spread (at *T)* and fixed coupon into equation *(3)*.

Upfront-Par Spread Conversion

A method similar to the Par Spread-Upfront Conversion above can be used to convert from upfronts to par spreads. As mentioned above, composites are initially formed in terms of upfronts and then converted into par spreads. This conversion uses the composite recovery rate.

Starting with the upfront curve, the bootstrapping method can be used to find the survival probabilities at each tenor with the following equation:

$$Upfront + PV(fixed\ coupon\ payments) = PV(default\ leg) \tag{4}$$

The upfronts, fixed coupon, recovery rate and interest rate curve will be used as inputs into equation *(4)* which would then yield the hazard rate at each tenor. These hazard rates can be fed into equation *(2)* to obtain the survival probabilities. Once the survival probability at each tenor is found, it can be fed into the equation below along with the upfront, interest rate curve and fixed coupon. This will yield the par spread at each tenor *T*:

$$Par\ Spread(tenor\ T) = \frac{Upfront(T)}{PV01} + fixed\ coupon \tag{5}$$

Note that the *PV01* which appears in equation *(5)* is specific for tenor *T*.

Disclaimer:

Opinions, estimates and projections in this report constitute the current judgment of the author(s) at the time of writing. They do not necessarily reflect the opinions of Markit Group Limited. Markit Group Limited has no obligation to update, modify or amend this report or to otherwise notify a reader thereof in the event that any matter stated herein, or any opinion, projection, forecast or estimate set forth herein, changes or subsequently becomes inaccurate.

Markit Group makes no warranty, expressed or implied, as to its accuracy, completeness or timeliness, or as to the results to be obtained by recipients, and shall not in any way be liable to any recipient for any inaccuracies, errors or omissions herein.

Without limiting the foregoing, Markit Group shall have no liability whatsoever to a recipient of this report, whether in contract (including under an indemnity), in tort (including negligence), under a warranty, under statute or otherwise, in respect of any loss or damage suffered by such recipient as a result of or in connection with any opinions, recommendations, forecasts, judgments, or any other conclusions, or any course of action determined, by it or any third party, whether or not based on the content, information or materials contained herein.