Predicting Agency Rating Movements with Spread Implied Ratings

Jianming Kou* Simone Varotto*

This version: 22 December 2004

Abstract

Investors in the credit market traditionally rely on credit ratings produced by rating agencies to determine the creditworthiness of debt issues. A major drawback of these agency-produced ratings is the lack of timeliness. In this work we derive the yield spread implied ratings for a large dataset of Eurobonds, taking into account the term structure effect and the time-varying spread level. We then compare the behaviour pattern of the spread implied rating and that of the agency rating. Our statistics suggest that spread implied ratings could be used to predict the future movement of agency ratings.

ISMA Centre, The University of Reading

Whiteknights, PO Box 242, Reading RG6 6BA, UK

Tel: +44 (0)118 378 8239

Fax: +44 (0)118 378 4741

E-mail: j.kou@ismacentre.reading.ac.uk; s.varotto@ismacentre.reading.ac.uk

1. Introduction to credit ratings

Credit ratings are evaluations of the creditworthiness that rating agencies, such as Standard & Poor's, Moody's and Fitch, give to debt issues or issuers.[1] They are judgements about the borrower's ability to meet its obligations. Ratings are reported in discrete categories. For example, S&P's long-term bond ratings include 10 broad categories, AAA, AA, A, BBB, BB, B, CCC, CC, C and D, with AAA representing the highest quality and D being in default. These broad rating categories are further refined by attaching modifiers (+/- sign) to each category. Ratings are usually first assigned to public debt at the time of issuance and periodically reviewed afterwards. An upgrade (downgrade) in the rating reflects the agency's judgement that the borrower's credit quality has improved (deteriorated).

Credit ratings are widely regarded as an important tool to investors in credit markets. They are also an essential input to many credit risk models, such as the pricing model proposed by Jarrow et al. (1997), portfolio credit risk model like JP Morgan's Creditmetrics. Under the US regulation, the ratings of NRSRO[2] are used to assess the value of securities held by securities firms and the amount of capital they must hold. Many institutions such as pension funds can only hold bonds with investment-grade ratings. More recently, the New Basel Accord (Basel 1999) incorporates credit ratings to assess the adequacy of bank's capital. Changes in agency rating have important implications for various market participants, since they can affect the issuer's cost of capital, credit spreads, bond returns, and the prices and hedge ratios of credit derivatives.

Despite their vast popularity, credit ratings have some weaknesses. First, they are frequently criticised for their "stickiness", i.e., inability to provide early warning of potential risk. For example, rating agencies were blamed for failing to predict the emergence of the East Asian Crisis in 1997 (Ferri et al. 1999). Several recent high

---

[1] Although ratings assigned by these agencies all reflect the credit risk of debt issuers, there is a subtle difference in the definition: S&P's issuer rating is an indicator of the default probability alone, while Moody's and Fitch define their debt ratings as opinions about both the likelihood of default and the recovery rate in the event of a default. (See Schuermann & Jafry 2003, Crouhy et al. 2000, Moody's 2004).

[2] NRSRO stands for "Nationally Recognized Statistical Rating Organization". There are four NOSROs in the US. SEC initially granted the title to Standard & Poor's, Moody's and Fitch in 1975. Later on, Dominion Rating Services was also added to the list. See Beaver et al. (2004) for details.

profile default events, such as Enron and Worldcom, also highlight this problem. In Enron's case, the company filed for bankruptcy on 2nd December 2001, but on 1st November 2001, it was still rated as BBB by S&P. Major credit rating agencies did not downgrade it to junk status until 28 November 2001, only 4 days before default.

In reaction to this criticism, rating agencies argue that since rating changes can have substantial economic consequences for a wide variety of debt issuers and investors, their policy is to provide stable measures of relative credit risk, i.e., rating is changed only when the issuer's relative fundamental creditworthiness has changed and the change is unlikely to be reversed in a short period of time. (See Moody's 2003) Inevitably, there is a tradeoff between these two aspects of rating quality: accuracy and stability. And some accuracy with respect to short-term default prediction (so called "early warning power") may be sacrificed. Nevertheless, investors still need an efficient risk indicator to assess the financial risk timely.

Another major criticism often being raised against rating agencies is the potential conflict of interest. After all, it is not unreasonable to worry about the objectivity of agencies' behaviour when it is the issuers, rather than the investors, who pay the fee to the rating agencies. Rating agencies repeatedly claim that reputation is of paramount importance to them, as it is the most important asset that allows them to do business. Yet the reputation risk did not stop Anderson's auditing failure at Enron.

An alternative way to measure credit risk is to derive "implied ratings"[3] from the market price of traded instruments. One type of such indicators is that derived from the equity prices, represented by the Expected Default Frequency (EDF) measures produced by Moody's KMV, which is default probabilities obtained through a Merton-type model. During the past decade, EDF has become an important tool in credit risk assessment. Another type is the bond market-implied rating recently introduced by Barra and Moody's (see Breger et al. 2002 and Moody's 2003). The idea is to derive ratings from the market price of debt instruments or credit derivatives such as credit default swaps (CDS).

[3] To avoid confusion, we will refer to the credit ratings assigned by rating agencies as "agency rating", in contrast to the "implied ratings" derived from the yield spread.

In this work we show that implied ratings inferred from bond yield spreads can be used to predict the future movements of agency ratings and maybe helpful to enhance the responsiveness of agency ratings to changes in a company's credit profile.

2. Yield spread and Spread implied ratings

2.1 Components of the yield spread

The yield spread between a corporate bond and a government bond reflect several factors:

First, investors require higher return as compensation for the risk of default. In fact, the spread is often used as a measure of relative creditworthiness, with reduction in the credit spread reflecting improvement in the issuer's perceived credit quality. Yield spreads are inversely correlated with credit ratings. As bonds with lower credit ratings have higher default probability, their yield spreads are usually higher. Numerous researchers have documented the correlation between agency ratings and yield spreads. (See, for example, Ederington et al. (1987), Kao and Wu (1990), Hand et al. (1992), Altman (1997), Kliger and Sarig (2000), Cunningham et al. (2001), Perraudin and Taylor (2003), etc). The relationship between credit ratings and yield spreads is also explicitly modelled in reduced-form pricing models such as Jarrow et al. (1997) and Lando (1998).

Yield spreads also reflect the relative liquidity of corporate and treasury securities. Perraudin and Taylor (2002) and Houweling et al. (2003) find that liquidity is a major factor in the spread of investment-grade bonds, where the default risk is relatively small. Ericsson and Renault (2002) document the decreasing term structure of liquidity premium and the positive correlation between credit risk and liquidity risk. Diaz and Navarro (2002) also find that the liquidity spread has a downward-sloping term structure. The impact of liquidity on the corporate bond yield spread has also been studied by Jarrow (2001), Janosi et al. (2002), Longstaff et al. (2004), and others.

Other factors affecting the yield spread include the tax premium, which reflect the differential tax treatment on the interest income of corporate bonds and government bonds; and the risk premium, since the return on corporate bonds is riskier than the return on government bonds and a large part of the risk is non-diversifiable. The tax premium and risk premium of corporate bonds have been studied by Elton et al. (2002).

2.2 The yield spread implied rating

Since yield spreads are closely correlated with bonds' ratings, it should be possible to assign "implied ratings" to bonds according to their spread levels. For example, if we observe that over the last two-months AAA bonds typically carry a spread of 0~45 basis points and AA bonds usually carry a spread of 45~70 basis points, any bonds whose spread is less than 45 bps could be assigned an implied-rating of AAA, and those with spread between 45bp and 75bp could be given an implied-rating of BBB.

However, the difficulty with this approach is that the yield spreads of different rating categories often overlap.[4] For example, some single-A issues traded with spreads higher than BBB issues, and it is not uncommon that the spread of some junk bonds is lower than that of certain investment-grade issues. To overcome this difficulty, a criterion is needed to set an appropriate boundary between say, AAA bond spreads and AA bond spreads. Moody's and Barra use different methods to deal with this issue.[5]

To estimate the spread implied rating (abbreviated as SIR hereafter) we adopt the method of Breger et al. (2002). For each rating category, we create a penalty function that depends on the upper and lower spread boundaries. The penalty value will increase when a bond's yield spread is outside the upper or lower boundaries corresponding to its credit rating (i.e., when the bond's implied rating is different from its agency rating). This penalty function is mathematically defined as follows,

[4] This overlapping behaviour is documented, for example, in Perraudin and Taylor (2003).
[5] To derive bond market implied ratings, Moody's (2003) use end-of-month bid price and spread data. They relate option-adjusted spreads to option-adjusted durations on a single day for all straight and callable coupon bonds in the sample and derive a pricing "matrix" which maps the median credit spread by rating category to different option adjusted spreads. A look-up table then allows them to infer the bond market's implied credit rating for any individual bond.

$$P(\mathbf{b}) = \sum_i \left[E\left(s_i - b_i^+ \middle| s_i > b_i^+\right) + E\left(b_i^- - s_i \middle| s_i < b_i^-\right)\right]$$

Where s_i is the spread of a bond with rating *i*, b_i^+ is the upper spread boundary for implied rating *i*, and b_i^- is the lower spread boundary for implied rating *i*. Note that the lower boundary of a higher rating (e.g. rating A) is also the upper boundary of its immediate lower rating (e.g. rating BBB). The optimum boundaries will be the ones that minimize the penalty value jointly for all rating classes.

After obtaining the optimum spread boundaries, a mapping procedure can be employed to derive implied ratings. Issues within the estimated spread region typical for a certain agency rating will be given a implied rating equal to that agency rating. In other words, an issue traded with a spread level typical for A-rated issues will have an implied-rating of A, even if its actual agency rating is different.

By doing so, we assume that agency ratings are *on average* informative. However, there may be inconsistencies between a bond's agency rating (which reflects rating agencies' judgement of risk) and its market price (which reflects the risk perceived by the market) in which case the implied and agency rating will differ.

As we have pointed out before, the yield spread of corporate bonds is not solely the determined by the default risk premium. However, the tax premium in our sample should be negligible since our data almost entirely consist of bearer securities. Also, we control for the effect of systematic changes in the market-wide liquidity and risk premium by allowing spread boundaries to be time varying.

Several studies have shown that market implied ratings and EDFs react more quickly to the change in credit risk than agency ratings do (See, for example, Breger et al. (2002) and Kealhofer (2003)). However, they commonly use individual event studies to illustrate this effect. In this study, rather than taking individual examples, we use a large dataset to carry out systematic tests on whether implied ratings lead agency ratings.

In a similar study, Hull et al. (2002) examines the relationship between CDS spreads and Moody's rating events. They found that changes in the CDS spread tend to anticipate negative rating announcements, while the results for positive rating events the results are much less significant. Our paper is different from theirs in that we use bond spread implied ratings rather than CDS spreads. In addition, our focus is on the information contained in the level of the ratings, rather than the change in CDS spreads. We find that the spread implied rating significantly leads the agency rating for both upgrade and downgrade, and the leading effect occurs very frequently in cases of downgrade.

3. The Data

The data used in this study are 4183 bond issues (most of them are Eurobonds) collected from Reuters 3000 Fixed Income Database. The data span 11 years from January 1988 to March 1998. These bonds are selected using the following criteria: fixed coupon rate and repaid at par, the principal and coupon payments are in the same currency, neither callable nor convertible, without sinking funds. The information contained in the database includes issue date, dated date,[6] maturity date, coupon rate, seniority, currency, industry, daily price history and rating history. The price data are "Reuters composite" bid prices that correspond to the best bid reported at the close of trading by a market maker from which Reuters have a data feed. These bonds have the following features.

The average initial rating of these bonds is AA. Among all the 4183 bonds, only 148 of them have initial ratings below investment-grade. The average initial maturity is 6.37 years. Most of these bonds are issued by banks or financial service institutions. According to MSCI classification: 1367 of the issuers are banks, 1671 are financial service institutions, and 34 are insurance companies. Together these three industry sectors account for 73.4% of total issuers. Table 1 shows the descriptive statistics of bond maturities, seniorities and currencies.

{Table 1 here}

[6] The dated date is the date from which a bond begins to accrue interest.

4. Methodology to estimate the SIRs

To obtain the SIRs, we first calculate the yield spread of corporate issues and filter out noisy data as described below. After that, we estimate the optimum spread boundaries, which is then used to work out the SIRs.

4.1 Yield spread calculation

First, we use the price and cash flow information to work out the daily yield to maturity for the bonds in the database. Then we calculate the yield spread between them and default-free government bonds. To do this, our approach is the same as Diaz and Navarro (2002). For each bond and each trading day, we create a hypothetical treasury bond that has the same cash flow structure and maturity date as the corporate bond under consideration. We then price the hypothetical treasury using treasury spot curve for the same day. After this step the calculated price and future cash flows are employed to work out the yield to maturity of the hypothetical treasury. The yield spread will be the difference in yield to maturity between the real corporate bond and the hypothetical treasury bond. The major advantage of this approach is that it matches the duration and convexity of corporates and treasuries, thereby avoid the so-called coupon bias in spread calculation. (See Duffee 1998 for a detailed analysis on coupon bias)

The treasury spot rates used to price the hypothetical treasuries are obtained from JP Morgan and Bloomberg. These are rates for 9 currencies and maturities from 1 year to 30 years (with step length of one year between 1-10 years, and 5 years between 10-30). For the non-integer maturities, we use linear interpolation of spot rates.

4.2 The problem with the yield and spread data

A major problem with the corporate bond data is the matrix price. When a bond does not trade frequently, dealers may price an issue by using simplistic algorithms or matrices. This problem is also reflected in our dataset.[7] In some cases, dirty prices

[7] Although Reuters do not populate price series with matrix data, it might well be the case that the dealers quote matrix price themselves and feed it into the system.

(clean price plus accrued interest) are higher than the face value plus the coupon payment at maturity, which implies that the yield to maturity is negative. Also at times prices tend to stay at the same level for a very long period. These phenomena usually occur when bonds are close to maturity,[8] which is probably because bonds usually become very illiquid during this period.[9] For this reason, we eliminate observations with time to maturity less than one year.

In addition, the following criteria are used to filter out possible errors and outliers in the price/yield observation. Observations are ignored whenever (a) the corporate yield is negative; (b) the yield spread is negative; (c) the issue is very illiquid: no price within 7 days before and after the current date; (d) Outliers occur in the spread time series;[10] (e) incorrect entries in credit rating history: rating changes but reverts back to its previous level within 5 trading days.

After the filtering process, we are left with around 3 million spread observations, covering the period from January 1988 to April 1998. In the following graph, we take a sample of bonds with maturity between 2 to 10 years from January 1994 to April 1994 and illustrate the distribution of spreads by agency rating categories.

{Graph 1 here}

As one might expect, average spreads increase with declining credit quality. In this sample, the mean yield spreads for AAA, AA, A, BBB and junk (BB and below) ratings are 44, 63, 84, 128, 236 basis points respectively. Unsurprisingly, spreads of lower rating categories exhibit higher volatility. For example, the standard deviation for AAA spreads is 21 bps, while the standard deviation for junk bond spreads is 100 bps. Another important feature is that spreads of different rating groups are clearly overlapping. For example, the spread of some AAA issues are higher than that of BBB issues, and quite a few junk issues traded with spread lower than that of investment grade bonds.

---

[8] When a bond is near maturity date, any tiny change in the price will be translated into huge variation in yield to maturity. Therefore, if the price feed is not very accurate, the yield to maturity can be very questionable.

[9] Amihud & Mendelson (1991) documented that when bonds approach maturity they have already been locked away in investors' portfolios and a large part of each issue is not readily available for trading.

[10] There is probably an error in data entry if the spread level of any trading day is higher (lower) than both the previous trading day and the next trading day by mean daily spread changes plus (minus) 2 standard deviations of daily spread changes of this bond.

4.3 Boundary setting

Our next step is to estimate spread boundaries between different rating categories. Since spreads for different rating categories often overlap, we use the optimization procedure described before to seek the optimum boundaries.

There are two practical issues related to the choice of boundaries. First, empirical evidence indicates that market spread levels fluctuate with the business cycle. (See, for example, Van Horne (1998), Huang and Kong (2003), etc.) During times of recession, credit spreads are expected to increase as firms experience difficulty in cash flow generating and investors become more risk averse. Yield spreads on junk bonds are particularly sensitive to the business cycle effect. Graph 2 plots the time-series of median spreads by agency ratings. For example, a 100 bp spread is equivalent to AAA rating in 1991, while in 1995 this is closer to a BBB. Naturally, the spread boundary should also reflect this time-varying behavior of spread levels.

{Graph 2 here}

Another issue is the term structure effect of credit spreads.[11] As shown in Graph 3, the spread of a 2-year BBB issue can be very different from the spread of a 10-year BBB issue, even they have the same agency rating. In addition, this term structure of credit spreads is also time varying: In June 1992, the term structure for BBB and junk issues are strongly downward sloping, with very high short-term spreads; while in April 1997 they behave like higher grade spreads with slightly upward sloping structure.

{Graph 3 here}

Taking these two issues into account, we estimate the boundary matrices in the following way: For each trading day, we collect yield spreads during the last two months for all traded bonds. We then divide the sample into 5 groups based on their time to maturity, 1 to 2 years, 2 to 3 years, 3 to 5 years, 5 to 10 years, and 10 to 30 years. Within each maturity band, we pool the spread data into different categories according to their agency ratings. We then use an optimization algorithm to find the

[11] Documents on the term structure of credit spreads include Longstaff & Schwartz (1995), He et al. (2002), Duffie & Singleton (1999), etc.

boundaries between adjacent rating categories within each maturity band.[12] After this, we obtain a matrix of optimum spread boundaries from which the spread implied ratings can be inferred. An example of the boundary matrix is illustrated in Table 2.

{Table 2 here}

The boundary matrices are estimated daily from 31 January 1989 to 3 April 1998. In graph 4, we illustrate the term structure of boundaries between 1993 and 1998. In this graph, the boundary between any two adjacent rating categories is represented by a surface. The term structures of boundaries between 4 investment grade categories are generally upward sloping, while the slope of the boundary between BBB and Junk are highly uncertain.

{Graph 4 here}

Using the spread boundaries obtained above, we can infer the history of implied ratings from the yield spread time series. Graph 5 gives an example of the implied ratings.

{Graph 5 here}

As we can see from Graph 5, spread implied ratings are very volatile. Moody's (2003) documented that market implied ratings are much more volatile than Moody's (agency) ratings, even though their data frequency is lower than ours.[13] Unsurprisingly, by using daily data, the volatility of our spread-implied ratings is even higher.

In most cases, SIR fluctuates around the agency rating (abbreviated as AR hereafter). However, if the SIR is persistently different from the AR, one might argue that the credit risk of the issue as perceived by the market is different from the judgement of rating agencies. In this case, since implied ratings are forward looking by nature, they might be able to predict the future movement of agency ratings. In the next section, we use our data to study the lead-lag relationship between these two types of ratings.

[12] As we have very few observations for ratings below BBB, the convergence of the optimisation procedure is difficult and very unstable. For this reason, we combine all the categories below BBB together as a JUNK category.

[13] Moody's (2003) use end-of-month bid price and spread data.

5. Test the lead-lag relationship between SIR and AR

Since agency ratings do not change often, they may not reflect credit quality changes in a timely manner. In contrast, market prices change continuously and anticipate changes in credit qualities (See Grier and Katz (1976), Weinstein (1977), Pinches and Singleton (1978), Hite and Warga (1997)). It is not uncommon to observe that an agency rating downgrade or upgrade does not produce the expected adjustment in the price of the affected securities, as the market has already anticipated the rating change.

There are many studies comparing the lead-lag relationship between agency ratings and market implied measures. For instance, Moody's KMV occasionally publish some case studies illustrating the early-warning power of EDFs relative to agency ratings using some high-profile default events (see, for example, Kealhofer 2003). Breger et al. (2002) also compares the relative power of market implied ratings and agency ratings using Xerox, Enron and Arhold as examples. However, none of them did a comprehensive study of the lead-lag relationship of agency rating and market-based measures.

Since we have the history of agency ratings for each bond issue in the sample and have derived the time-series of spread-implied ratings, we are able to compare their behaviour over time.

As described before, the spread boundaries are built upon the basis of bond's agency ratings. In equilibrium, agency ratings and spread implied ratings should be consistent with each other. Any inconsistency would suggest that the investors in the market and rating agencies have different opinion towards the future prospects of the issue. However, if both parties did a good job (which means market is efficient and AR is consistent), over time this inconsistency tends to disappear, and new equilibrium will form. (Although in practice, like in many macroeconomic models, a static equilibrium may never occur).

During the process towards the new equilibrium, we may observe four general patterns of SIR and AR movements: (a) SIR leads; (b) AR leads; (c) Convergence;[14] and (d) divergence[15].

5.1 Forward-tracking analysis

Based on the reasoning above, we carry out a forward-tracking statistics. We construct non-overlapping 22-trading-day (equivalent to one calendar month) intervals for each bond issue. Intervals during which agency rating changes are eliminated. For each interval we calculate the average difference between spread implied ratings and agency ratings. If there is an adjustment in agency rating in the next 22-day interval, we compute the average value of AR and SIR for the 5 trading days after the AR adjustment (including the day when AR changes) and classify the events as one of the four patterns mentioned above.[16] To ensure the consistency of the comparison, we ignore observations where the agency rating is "non-rated" or "withdrawn".

In out dataset, 2850 bonds have rating histories by Standard & Poor's, and 4005 have rating histories by Moody's. We carry out the tracking process using S&P and Moody's data separately. The descriptive statistics are presented in Table 3a and Table 3b.

{Table 3 here}

Table 3a shows the result for S&P data. Intervals in the first column of the table denotes the average difference between an issue's spread implied rating and agency rating. Since we transform letter ratings into numerical values (AAA is 1, AA is 2, A is 3, BBB is 4, BB and below is 5), negative (positive) values mean the spread implied rating is better (worse) than the agency rating. The remaining columns are the

[14] During the process of convergence, spreads often overreact because they are more dynamic (especially in extreme market conditions). This might also happen with agency ratings, but should be less frequent.

[15] The situation of "Divergence" can arise because, for example, if the inconsistence is due to factors other than credit risk, such as liquidity, rating will not necessarily change as spread suggests.

[16] The exact rule is: We transform the letter ratings into numerical values (AAA to 1, AA to 2, A to 3, BBB to 4, BB and below to 5) and calculate the average difference between SIR and AR for the 22-day interval. In the case when the initial 22-day average SIR minus AR is positive (i.e., the implied rating is worse than the agency rating), if there is a downgrade in AR in the following interval and SIR remains unchanged or worse than before, we count it as an SIR-lead; if there is a downgrade in AR in the following interval and SIR is higher than before, we count it as a convergence; if there is a upgrade in AR in the following interval and SIR is higher than before, we count it as an AR-lead; if there is a upgrade in AR in the following interval and SIR remains unchanged or lower than before, we count it as a divergence. The same logic applies when the initial average SIR minus AR is positive.

frequencies of each pattern occurring given that the average difference between SIR and AR during the previous 22-day interval located in a particular interval.

From the whole sample we observe 431 changes in S&P ratings. As displayed in the last row of panel A, on average, the spread implied rating leads the S&P rating in 50% of time (214 observations), while the reverse is true only in 13% of time (55 observations). Convergence occurs in 29% of cases (123 observations) and the frequency of divergence pattern is 9% only (39 observations). Three points are worth noting.

First, the SIR-lead pattern clearly dominates when the spread implied rating is worse than the S&P rating initially (positive interval). When the spread implied rating is better than the S&P rating initially (negative interval), if the initial difference is within one notch the S&P rating tends to lead more frequently. If the initial difference is larger than one notch, SIR-lead pattern again dominates. Second, the larger the initial difference, the more likely that SIR will lead. Third, the convergence pattern can also be interpreted as a type of SIR-lead, since when convergence occurs AR actually moves towards the direction where the SIR has predicted previously. Although at the same time the SIR also moves towards its counterparty, it is largely due to the fact that SIRs are more volatile. Unsurprisingly, the larger the initial difference between SIR and AR, the more likely that convergence will occur.

We repeat the statistics for observation intervals of 44 trading days and 66 trading days and the conclusion remain unchanged. The results are presented in Panel B and Panel C of Table 3a.

Table 3b shows the result for Moody's ratings. For the 22-day window, there are 801 observations in total. Similar to S&P's case, spread implied ratings lead the Moody's ratings in 53% of time, while the reverse occurs in only 13% of time. This is true for all initial difference ranges except when the initial average difference between the SIR and Moody's rating is between –1 and 0. Convergence occurs in 26% of cases and the frequency of divergence pattern is 8% only. Again, the conclusion remains unchanged for the 44-day and 66-day observation windows.

These distribution patterns for various initial gaps are further illustrated in Graph 6a (S&P) and Graph 6b (Moody's).

{Graph 6 here}

In both S&P and Moody's case, the SIR-lead pattern is striking when the initial difference between the SIR and the AR is positive: the chance that agency ratings lead spread implied ratings is at most 5%, while the combined frequency of SIR-lead and convergence is above 90%. This observation is particularly important to investors ---- People should be very cautious when the spread implied rating persistently stays below the agency rating, because it is very likely that the rating agencies will follow up.

The forward-tracking analysis is interesting in that it gives us a picture about various lead patterns and the frequency of their occurrence. However, it does not reveal the statistical significance of the SIR's leading power. In addition, it would be interesting to know the extent to which that SIR can provide the early warning, in other words, the time lag between SIRs and ARs. A backward-tracking analysis will provide answers to both these two questions.

5.2 Backward-tracking analysis

The second part of the empirical analysis is a backward-tracking analysis. The idea is, whenever we observe a change in agency rating, we look back to see whether this change has already been anticipated in the market by comparing the past history of SIR and AR. For each agency rating events, we calculate the average difference between the SIR and AR for various intervals prior to the AR change and check whether the mean of the average differences is significantly greater (less) than zero for AR downgrades (upgrades).

To test the significance of this mean, if the sample size is large, then according to the Central Limit Theorem the sample mean $\overline{X}$ is normally distributed with $\mu_{\overline{X}} = \mu$ and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$. We can use the standard t-statistic

$$t = \frac{\mu}{\sigma / \sqrt{n}}$$

to determine whether the sample mean is significantly different from zero. However, if the sample size is small, this standard t-test becomes inappropriate. For example, in S&P case, for upgrading and downgrading by more than 2 notches, we only have 7 and 15 observations respectively. In this case, we use the bootstrap technique documented in Hull et al. (2002).

Here the null hypothesis is that the mean of the average SIR-AR difference is zero. For a sample of average SIR-AR differences X_1, $X_2 \ldots X_n$, we form a group of n adjusted observations $X_i - \overline{X}$ (As a result, this group of adjusted observations has a mean zero). We then repeat the following a large number of times: sampling n times with replacement from the adjusted group and calculate the t-statistic of the sample mean. This provides an empirical distribution for the t-statistic under the null hypothesis that the mean of average SIR-AR difference is zero. If the t-statistic for the original (i.e., unadjusted) sample mean exceeds the 99th percentile of this empirical distribution, we reject the null hypothesis. The results are presented in Table 4.

Table 4 here

Table 4a shows the result of the backward-tracking process using S&P data. Panel A summarizes the statistics for AR upgrade and downgrade across the whole sample. The first column describes the direction and magnitude of agency rating adjustment. The remaining columns show the average difference between spread implied ratings and agency ratings during various time intervals prior to the agency rating adjustment. As we have transformed the letter ratings into numerical values, negative (positive) values mean that the average spread implied rating is better (worse) than the agency rating. Numbers in the parenthesis underneath the mean differences are the t-statistic of the mean difference, bootstrapped 99th percentile[17] of the t-statistic under the null hypothesis that the mean difference is zero, and the number of observations respectively. Since we have very few observations for the S&P rating adjustments more than one notch, the bootstrapped percentile is the preferred choice in determining the significance of sample values.

[17] We provide the bootstrapped percentiles for all cases. However, for AR downgrade or upgrade by 1 notch where we have a large number of observations, the bootstrapped percentiles are close to 2.33, the 99% percentile under the normal distribution assumption. This is consistent with the prediction of the central limit theorem.

The table shows that the agency rating adjustments are anticipated by spread implied ratings. First, the signs of the mean difference between SIR and AR are all consistent with the SIR-lead pattern. For AR upgrades, the mean difference are negative for all time intervals up to 126 trading days (half a calendar year) prior to it, which means that SIR has risen above AR before the AR upgrading well in advance. For AR downgrades, the mean difference between SIR and AR are positive for all time intervals prior to it, which means that SIR has dropped below AR well ahead of the AR downgrade. In the case of AR upgrade/downgrade by one notch where we have a lot of observations, the magnitude of these mean differences clearly increases when the date of AR adjustment is approaching, indicating that the increasing discrepancies between the market's opinion and the current agency rating signal the pending agency rating adjustments. Second, all these mean differences are statistically significantly at 99^{th} confidence level. Third, the absolute values of mean differences are larger for bigger agency rating adjustments, implying that larger disparity between SIR and AR are followed by stronger adjustment by the rating agency.

Since many portfolio managers' investments are restricted to the investment-grade securities, it is interesting to look at the agency ratings' behaviour around the boundary between investment-grade and speculative-grade. We use the popular definition of "fallen angels" (issuers whose ratings fell from investment grade to speculative-grade) and "rising stars" (issuers whose ratings rose from speculative grade to investment grade) and examine the relationship between SIR and AR prior to their "rising" or "falling". As shown in Panel B of Table 4a, the result remains qualitatively the same as before. For "rising stars" the mean differences are negative for all time intervals and are significant at 99^{th} confidence level. For "fallen angels" they are significantly positive for all intervals up to 44 days (2 calendar months) ahead of the AR downgrade. It appears that the spread implied rating can predict the Standard & Poor's action around the boundary between the investment-grade and speculative-grade, especially for "rising stars".

Using Moody's rating data, we obtain a similar result. In panel A of Table 4b, nearly all of the observations are for the upgrading/downgrading by one notch, where the mean differences are correctly signed and highly significant for all intervals. For downgrading more than one notch the mean difference is positive, contrary to the prediction of SIR lead patterns. However, this might be due to the small sample bias since we only have 7 observations here.[18] In the statistics for "rising stars" and "fallen angels" in panel B, the mean differences for rising stars are negative and statistically significant, which is consistent with the SIR lead pattern. The mean differences for fallen angels are mixed and insignificant.

Overall, the result of the backward-tracking analysis corroborates the SIR's leading power. For the AR move by one notch where most observations lie, the leading effect is highly significant. The results for various intervals show that the SIR can provide warning as early as six months ahead of the AR change. In addition, it appears that SIRs are able to predict the rating agency's action around the boundary between the investment-grade and speculative-grade, especially for "rising stars".

6. Conclusion

We derive the implied ratings for a large dataset of Eurobonds based on their yield spreads over risk-free hypothetical treasury bonds, taking into account the term structure effect and the time-varying spread levels. Compared with agency ratings, these spread-implied ratings are forward-looking and more dynamic by nature.

We compare the behaviour of spread implied ratings and agency ratings and find that spread implied ratings are able to predict the future movements of agency ratings. Our analysis reveals that when the spread implied rating is persistently different from the agency rating, it is very likely that the latter will be adjusted towards the direction that the former has indicated, conditional on there is an action by the agencies. For larger differences, they are also likely to converge. Both upgrade and downgrade are anticipated by spread-implied ratings well ahead of the events. We also show that spread-implied ratings have predicting power for the agency's action around the boundary between the investment-grade and speculative-grade, which might be useful to portfolio managers subject to investment constraints in credit ratings (e.g. Pension funds).

[18] The result for upgrading by two or more notches should be ignored, since we only have one observation here.

Overall, spread-implied ratings can be regarded as a valuable addition to the information provided by agency ratings.

Reference

Altman, Edward I. and Duen Li Kao, November 1991, The Implications of Corporate Bond Rating Drift, NYU Salomon Center's Working Paper S-91-51,

Altman, Edward I., 1997, Rating migration of corporate bonds: comparative results and investor/lender implications

Bank for International Settlements, Basel Committee on Banking Supervision, 1999, A new capital adequacy framework, Consultative paper, June

Beaver, William H., Catherine Shakespeare and Mark T. Soliman, 2004, Differential properties in the ratings of certified vs. non-certified bond rating agencies, Stanford University Working Paper

Breger, Ludovic., Lisa Goldberg and Oren Cheyette, 2002, Market Implied Ratings, Horizon, The Barra Newsletter, Autumn 2002

Cantor, P. and F. Packer, 1997, Differences of opinion and selection bias in the credit rating industry, Journal of Banking and Finance 21, 1395-1417

Crouhy, Michel, Dan Galai and Robert Mark, 2000, A comparative analysis of current credit risk models, Journal of Banking and Finance, 24, 59-117

Cunningham, Alastair., Liz Dixon and Simon Hayes, 2001, Analysing yield spreads on emerging market sovereign bonds, Financial Stability Review December 2001, Bank of England

Diaz, Antonio., Eliseo Navarro, 2002, Yield spread and term to maturity: default vs. liquidity, Financial Analysts Journal

Duffie, D. and K.J. Singleton, 1999, Modeling term structures of defaultable bonds, Review of Financial Studies, 12 (4) 687-720

Duffee, Gregory R., 1998, The relation between treasury yields and corporate bond yield spreads, Journal of Finance, December 1998, 2225-2241

Ederington, H. Louis, Jess B. Yawitz and Brian E. Roberts, 1987, The information content of bond ratings, Journal of Financial Research 10, 211-226

Elton, Edwin J., Martin J. Gruber, Deepak Agrawal, and Christopher Mann, 2001, Explaining the rate spread on corporate bonds, Journal of Finance 56, 247-277

Ferri, G., L.-G. Liu and J.E. Stiglitz, 1999, The Procyclical role of rating agencies: Evidence from the east asian crisis, Economic Notes 3-1999, 335-355

Grier, Paul and Steven Katz, 1976, The differential effects of bond rating changes among industrial and public utility bonds by maturity, Journal of Business 49, 226-239

He, Jia., Wenwei Hu and Larry H. P. Lang, 2000, Credit spread curves and credit ratings, Working paper, Chinese University of Hong Kong

Hite, Galen & Arthur Warga, 1997, The effect of bond-rating changes on bond price performance, Financial Analysts Journal, May/June, 35-47

Moody's, 2003, Measuring The Performance of Corporate Bond Ratings, April

Moody's, 2004, Default & Recovery Rates of Corporate Bond Issuers A Statistical Review of Moody's Ratings Performance 1920-2003

Hand, J., R. Holthausen and R. Leftwich, 1992, The effect of bond rating agency announcements on bond and stock prices, Journal of Finance 47, 733-752

Houweling, Patrick., Albert Mentink and Ton Vorst, 2003, How to measure corporate bond liquidity? Erasmus University Rotterdam working paper

Hull, John., Mirela Predescu-Vasvari and Alan White, 2002, The relationship between credit default swap spreads, bond yields, and credit rating announcements, University of Toronto working paper

Huang, Jingzhi. and Weipeng Kong, 2003, Explaining credit spread changes: Some new evidence from option-adjusted spreads of bond indexes, Penn State University working paper

Janosi, Tibor., Robert Jarrow and Yildiray Yildirim, 2002, Estimating expected losses and liquidity discounts implicit in debt prices, working paper

Jarrow, Robert A., David Lando and Stuart M. Turnbull, 1997, A Markov Model for the Term Structure of Credit Risk Spreads, The Review of Financial Studies, Vol. 10, No. 2, 481-523

Jarrow, Robert., 2001, Default parameter estimation using market prices, Financial Analysts Journal, Sep/Oct 2001

JP Morgan, CreditMetrics™, 1997

Kao, C. and C. Wu, 1990, Two-step estimation of linear models with ordinal unobserved variables: The case of corporate bonds, Journal of Business and Economic Statistics 8, 317-325

Kealhofer, Stephen, 2003, Quantifying Credit Risk I: Default Prediction, Financial Analysts Journal

Kliger, Doron and Oded Sarig, 2000, The information value of bond ratings, Journal of Finance, Vol LV, No.6, Dec 2000, 2879-2902

Lando, David, 1998, On cox process and credit risky securities, Review of Derivatives Research 2 (2), 99-120

Lando, David and Torben M. Skodeberg, 2002, Analyzing rating transitions and rating drift with continuous observations, Journal of Banking and Finance 26, 423-444

Longstaff, Francis A., and E.S. Schwartz, 1995, A simple approach to valuing risky fixed and floating rate debt, Journal of Finance, 50, pp. 789-819

Longstaff, Francis A., Sanjay Mithal and Eric Neis, 2004, Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market, UCLA working paper

Pinches, G.E. & J.C. Singleton, 1978, The adjustment of stock prices to bond rating changes, Journal of Finance 23, 29-44

Perraudin, William., Alex P. Taylor, 2002, Liquidity and Bond Market Spreads, Bank of England Working paper

Perraudin, William., Alex P. Taylor, 2003, On the Consistency of Ratings and Bond Market yields, Bank of England Working paper

Schuermann, Til, Yusuf Jafry, 2003, Measurement and Estimation of Credit Migration Matrices

Weinstein, Mark, 1977, The effect of a rating change announcement on bond price, Journal of Financial Economics t, 329-350

Table 1 Summary statistics of the bond data

Panel A. Maturity of bond issues

Maturity range (years)	Number	Maturity range (years)	Number
(0,2]	11	(10,11]	820
(2,3]	194	(11,12]	13
(3,4]	609	(12,13]	39
(4,5]	328	(13,14]	5
(5,6]	1353	(14,15]	2
(6,7]	181	(15,20]	55
(7,8]	390	(20,25]	18
(8,9]	91	(25,30]	9
(9,10]	56	(30,50)	9

Panel B. Seniority of bond issues

Seniority type	Number	Seniority type	Number
Secured	54	Guaranteed	1021
Collateralised	9	Senior unsecured	425
Mortgaged	17	Unsecured	2399
Government guaranteed	165	Subordinated	93

Panel A. Currency of bond issues

Currency	Number	Currency	Number
US Dollar	1390	French Franc	305
German Mark	463	Japanese Yen	276
Canadian Dollar	456	Australian Dollar	274
Swiss Franc	418	Netherlands Guilder	210
UK Pound	391		

Table 2 The boundary matrix for 2nd December 1992

Maturity range	Rating category			
	BBB/BB	A/BBB	AA/A	AAA/AA
1 to 2 years	324	190	119	75
2 to 3 years	250	184	114	85
3 to 5 years	295	195	130	96
5 to 10 years	268	178	126	94
10 to 30 years	268	178	146	67

Table 3a Forward tracking statistics of SIR/AR behaviour patterns (Standard & Poor's)

Panel A. 22-day interval

SIR minus AR	No. of observations	Spread lead	Rating lead	Divergence	Convergence
[-4, -3]	**4**	75%	0%	0%	25%
(-3, -2]	**13**	69%	0%	0%	31%
(-2, -1]	**62**	42%	15%	19%	24%
(-1, 0]	**78**	14%	47%	22%	17%
(0, 1]	**179**	61%	4%	5%	30%
(1, 2]	**73**	63%	3%	1%	33%
(2, 3]	**21**	38%	0%	0%	62%
(3, 4]	**1**	100%	0%	0%	0%
Overall	**431**	**50%**	**13%**	**9%**	**29%**

Panel B. 44-day interval

SIR minus AR	No. of observations	Spread lead	Rating lead	Divergence	Convergence
[-4, -3]	**3**	67%	0%	0%	33%
(-3, -2]	**13**	69%	0%	0%	31%
(-2, -1]	**50**	44%	20%	16%	20%
(-1, 0]	**85**	18%	47%	21%	14%
(0, 1]	**174**	56%	5%	6%	33%
(1, 2]	**75**	51%	3%	1%	45%
(2, 3]	**19**	16%	0%	0%	84%
(3, 4]	**1**	100%	0%	0%	0%
Overall	**420**	**45%**	**15%**	**9%**	**32%**

Panel C. 66-day interval

SIR minus AR	No. of observations	Spread lead	Rating lead	Divergence	Convergence
[-4, -3]	**2**	100%	0%	0%	0%
(-3, -2]	**8**	75%	0%	0%	25%
(-2, -1]	**40**	43%	25%	8%	25%
(-1, 0]	**102**	24%	43%	21%	13%
(0, 1]	**172**	57%	7%	6%	30%
(1, 2]	**69**	51%	1%	1%	46%
(2, 3]	**17**	41%	0%	0%	59%
(3, 4]	**1**	100%	0%	0%	0%
Overall	**411**	**46%**	**16%**	**9%**	**29%**

Note: The first column displays the average spread implied rating minus the agency rating, negative number means spread implied rating is better than agency rating, and vice versa. For example, the interval (-2, -1) means on average SIR is better than AR by between one and two notches.

Table 3b Forward tracking statistics of SIR/AR behaviour patterns (Moody's)

Panel A. 22-day interval

SIR minus AR	No. of observations	Spread lead	Rating lead	Divergence	Convergence
[-4, -3]	**6**	67%	17%	0%	17%
(-3, -2]	**30**	67%	10%	13%	10%
(-2, -1]	**133**	45%	17%	20%	19%
(-1, 0]	**132**	25%	47%	12%	16%
(0, 1]	**333**	59%	5%	5%	31%
(1, 2]	**135**	67%	2%	0%	30%
(2, 3]	**31**	58%	0%	0%	42%
(3, 4]	**1**	100%	0%	0%	0%
Overall	**801**	**53%**	**13%**	**8%**	**26%**

Panel B. 44-day interval

SIR minus AR	No. of observations	Spread lead	Rating lead	Divergence	Convergence
[-4, -3]	**4**	50%	0%	0%	50%
(-3, -2]	**27**	41%	22%	7%	30%
(-2, -1]	**112**	46%	14%	17%	22%
(-1, 0]	**166**	25%	48%	11%	16%
(0, 1]	**314**	59%	4%	4%	32%
(1, 2]	**137**	59%	3%	0%	38%
(2, 3]	**27**	37%	0%	0%	63%
(3, 4]	**3**	67%	0%	0%	33%
Overall	**790**	**49%**	**15%**	**7%**	**29%**

Panel C. 66-day interval

SIR minus AR	No. of observations	Spread lead	Rating lead	Divergence	Convergence
[-4, -3]	**3**	67%	0%	0%	33%
(-3, -2]	**21**	48%	19%	5%	29%
(-2, -1]	**101**	44%	20%	12%	25%
(-1, 0]	**189**	26%	43%	13%	18%
(0, 1]	**287**	59%	5%	5%	31%
(1, 2]	**138**	59%	1%	0%	40%
(2, 3]	**20**	40%	0%	0%	60%
(3, 4]	**2**	50%	0%	0%	50%
Overall	**761**	**48%**	**16%**	**7%**	**29%**

Note: The first column displays the average spread implied rating minus the agency rating, negative number means spread implied rating is better than agency rating, and vice versa. For example, the interval (-2, -1) means on average SIR is better than AR by between one and two notches.

Table 4a Mean differences between SIR and AR for various time intervals prior to an AR change (Standard & Poor's)

Panel A. Statistics for the whole sample

Time interval	[-1,-5]	[-1,-22]	[-23,-44]	[-45,-66]	[-67,-126]
Upgrade by two notches or more	-1.09 (-1.76, -1.21*, 7)	-1.13 (-1.87, -1.13*, 7)	-1.33 (-2.00, -1.29*, 6)	-1.33 (-1.98, -1.31*, 6)	-1.37 (-1.70, -1.86*, 6)
Upgrade by one notch	-0.77 (-8.14, -2.37, 118)	-0.70 (-7.65, -2.47, 115)	-0.61 (-5.94, -2.29, 112)	-0.54 (-5.09, -2.35, 108)	-0.55 (-5.88, -2.30, 104)
Downgrade by one notch	0.60 (12.72, 2.33, 386)	0.56 (11.92, 2.37, 380)	0.51 (9.89, 2.42, 361)	0.45 (8.33, 2.35, 351)	0.25 (3.86, 2.42, 325)
Downgrade by two notches or more	1.68 (7.01, 2.52, 15)	1.81 (8.38, 2.19, 15)	1.84 (7.84, 2.09, 15)	1.52 (7.02, 2.38, 15)	1.23 (3.88, 2.83, 15)

Panel B. Statistics for rising stars and fallen angels

Time interval	[-1,-5]	[-1,-22]	[-23,-44]	[-45,-66]	[-67,-126]
Rising stars	-1.06 (-6.58, -1.87, 30)	-1.00 (-6.53, -1.76, 30)	-0.99 (-5.92, -1.65, 29)	-1.03 (-6.80, -1.70, 28)	-0.90 (-6.23, -1.52, 27)
Fallen angels	0.88 (4.68, 2.99, 15)	0.85 (4.41, 3.26, 15)	1.02 (7.25, 2.30, 14)	0.51 (1.95, 3.62, 14)	-0.59 (-1.32, 2.85, 14)

Note:

1. The time intervals denote the various periods (in trading days) before an agency rating change. For example, [-23, -44] refers to the period from 44 days before the agency rating change to 23 days before the agency rating change.
2. In the main body of the table are the mean difference between SIR and AR for different time intervals prior to a AR change, negative values mean that SIR are higher than AR, and positive values mean that SIR are lower than AR. These mean differences are presented together with the t-statistics within the parenthesis underneath, which are: t-statistic of the mean difference, the bootstrapped 99th percentile of the t-statistic under the null hypothesis that the mean difference is zero, and the number of observations.
3. The number with * are the bootstrapped 95^{th} percentile.
4. The central limit theorem indicates that when the size of a sample is large, the distribution of its mean converges to normal distribution. In the table, the 99^{th} percentile of t-statistics for AR upgrade/downgrade by one notch is very close to 2.33, the 99^{th} critical value of the standard normal distribution.

Table 4b Mean differences between SIR and AR for various time intervals prior to an AR change (Moody's)

Panel A. Statistics for the whole sample

Time interval	[-1,-5]	[-1,-22]	[-23,-44]	[-45,-66]	[-67,-126]
Upgrade by two notches or more	0.00	0.00	0.00	0.00	0.00
	(-, -, 1)	(-, -, 1)	(-, -, 1)	(-, -, 1)	(-, -, 1)
Upgrade by one notch	-0.77	-0.75	-0.74	-0.76	-0.61
	(-12.04, -2.33, 239)	(-11.97, -2.36, 234)	(-11.78, -2.37, 228)	(-12.11, -2.25, 225)	(-11.84,-2.37,212)
Downgrade by one notch	0.55	0.51	0.39	0.40	0.27
	(16.07, 2.31, 764)	(14.81, 2.40, 755)	(9.90, 2.36, 730)	(9.71, 2.46, 708)	(6.12, 2.35, 642)
Downgrade by two notches or more	1.00	0.99	1.62	1.76	1.69
	(1.85, 7.49, 7)	(1.83, 6.43, 7)	(6.91, 1.10, 5)	(7.38, 0.00, 4)	(5.49, 0.00, 4)

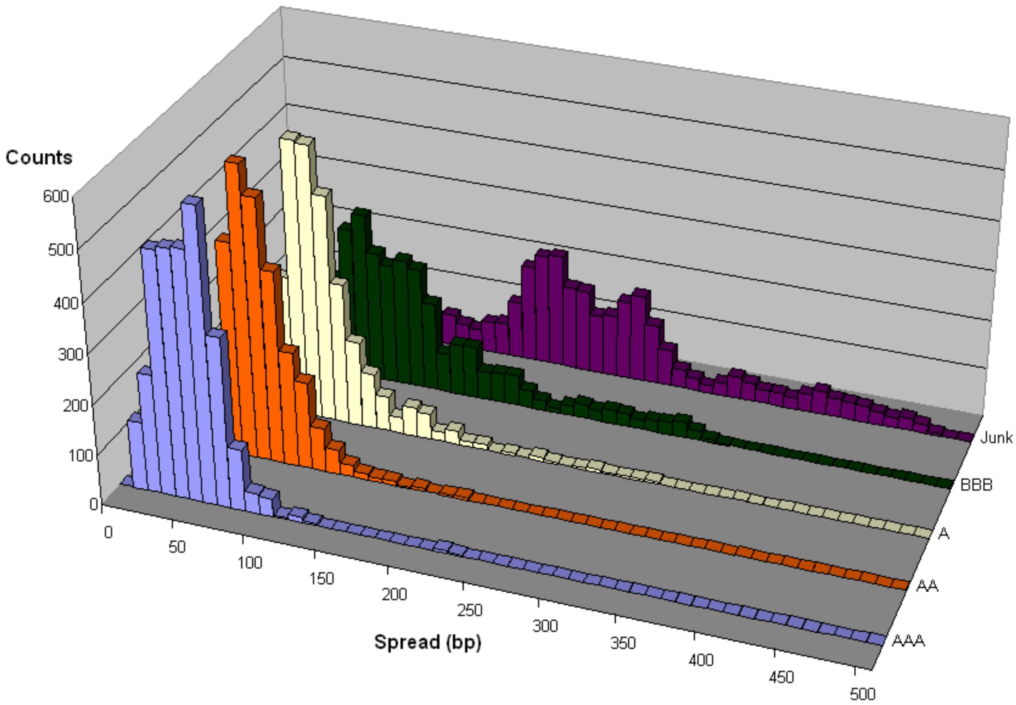
Panel B. Statistics for rising stars and fallen angels

Time interval	[-1,-5]	[-1,-22]	[-23,-44]	[-45,-66]	[-67,-126]
Rising stars	-1.06	-1.05	-1.01	-0.87	-0.67
	(-9.68, -1.71, 47)	(-10.30, -1.64, 46)	(-10.48, -1.78, 46)	(-8.42, -2.05, 45)	(-9.13, -2.51, 44)
Fallen angels	0.10	-0.01	-0.22	-0.29	-0.16
	(0.41, 3.12, 28)	(-0.04, 3.06, 28)	(-0.75, 2.77, 24)	(-0.72, 2.52, 22)	(-0.38, 3.18, 20)

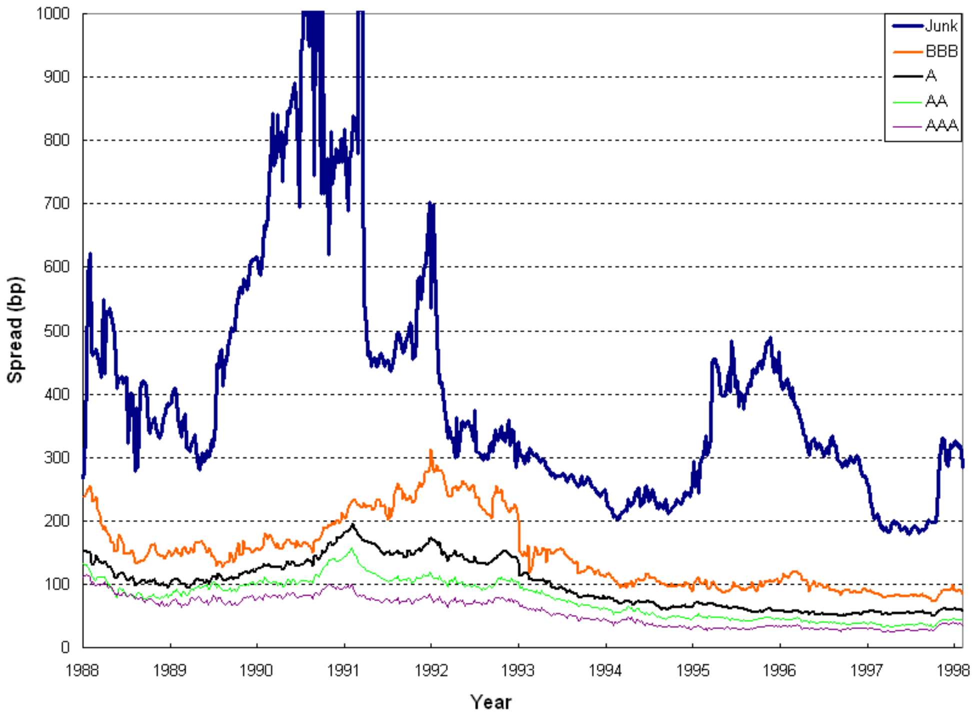
Note:

1. The time intervals denote the various periods (in trading days) before an agency rating change. For example, [-23, -44] refers to the period from 44 days before the agency rating change to 23 days before the agency rating change.
2. In the main body of the table are the mean difference between SIR and AR for different time intervals prior to a AR change, negative values mean that SIR are higher than AR, and positive values mean that SIR are lower than AR. These mean differences are presented together with the t-statistics within the parenthesis underneath, which are: t-statistic of the mean difference, the bootstrapped 99th percentile of the t-statistic under the null hypothesis that the mean difference is zero, and the number of observations.
3. The central limit theorem indicates that when the size of a sample is large, the distribution of its mean converges to normal distribution. In the table, the 99^{th} percentile of t-statistics for AR upgrade/downgrade by one notch is very close to 2.33, the 99^{th} critical value of the standard normal distribution.

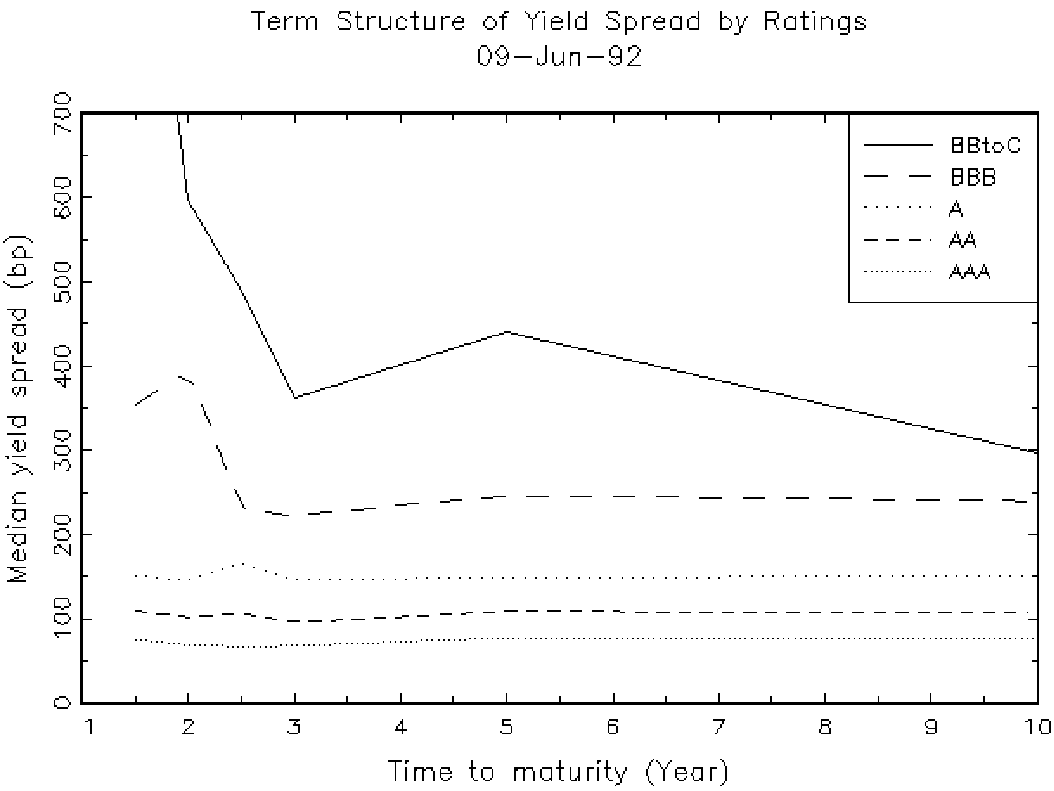
Graph 1 Distribution of yield spreads for different rating categories

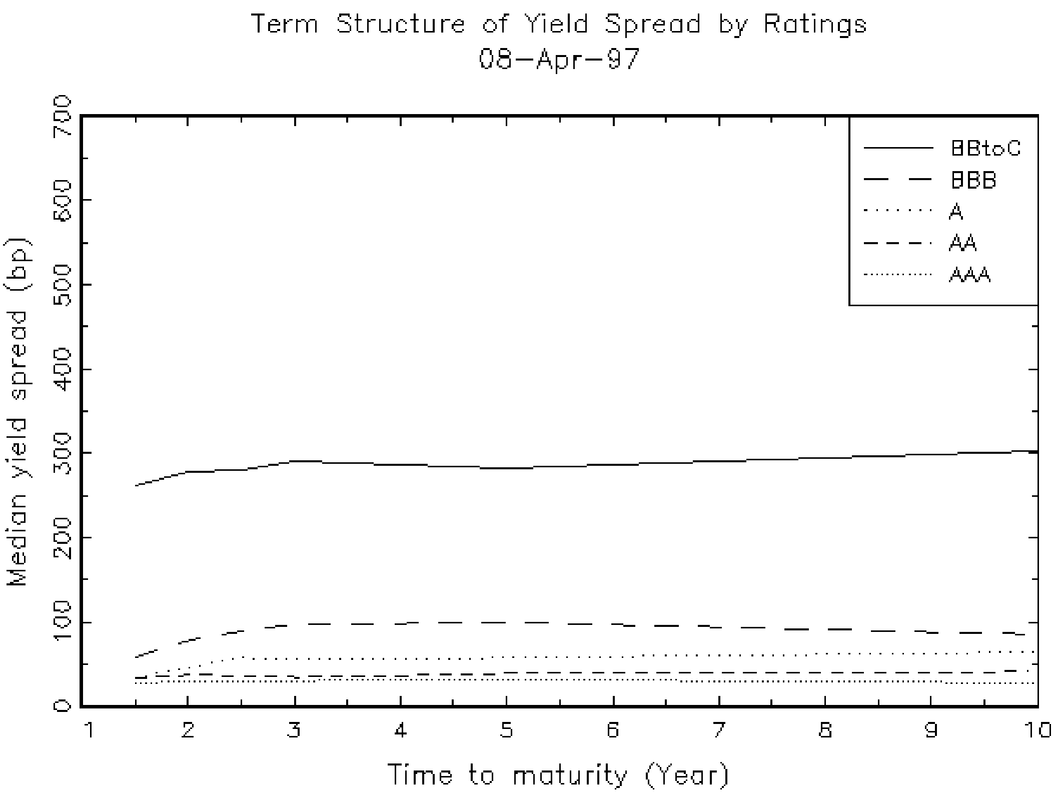


Graph 2 Time series of median spread by agency ratings



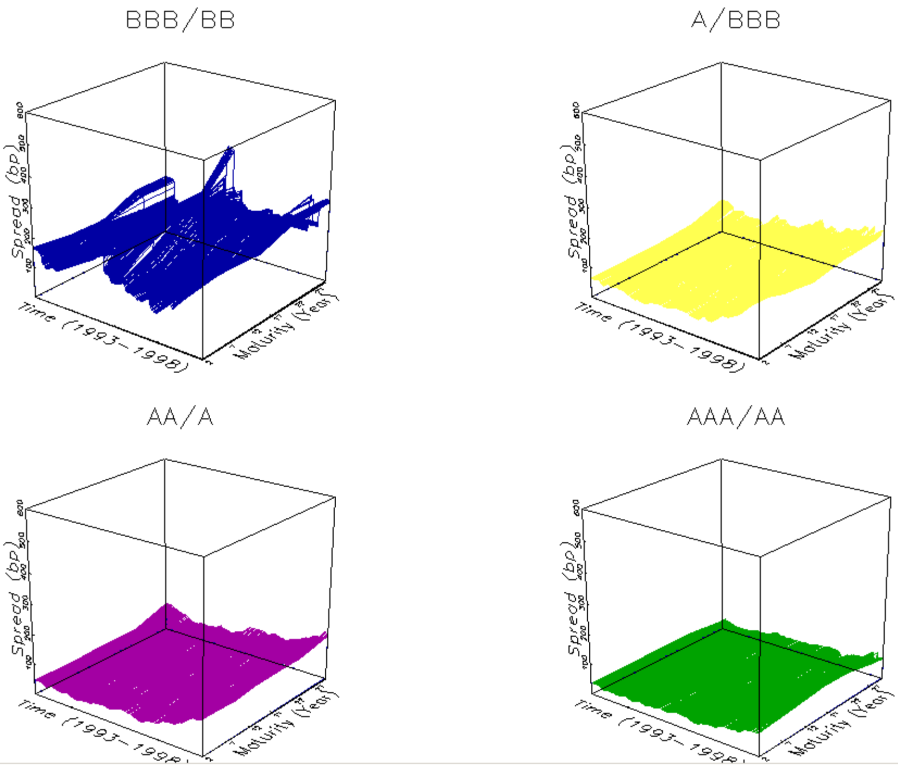
Graph 3 Term structure of median yield spread by ratings



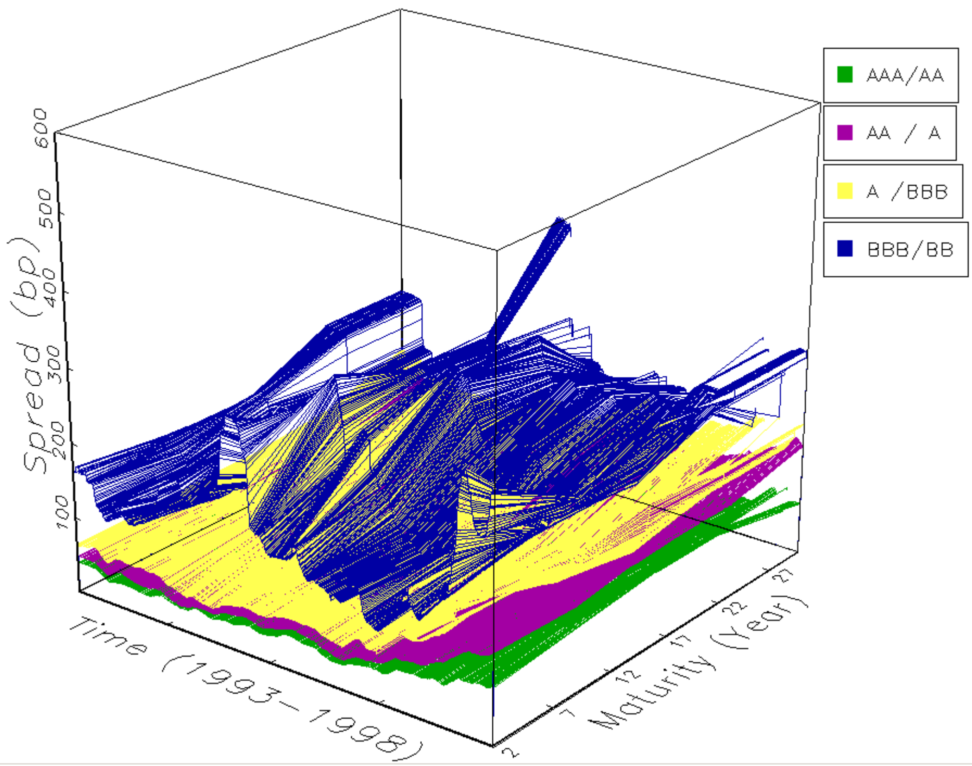


Graph 4 Term structure of Spread boundaries 1993-1998

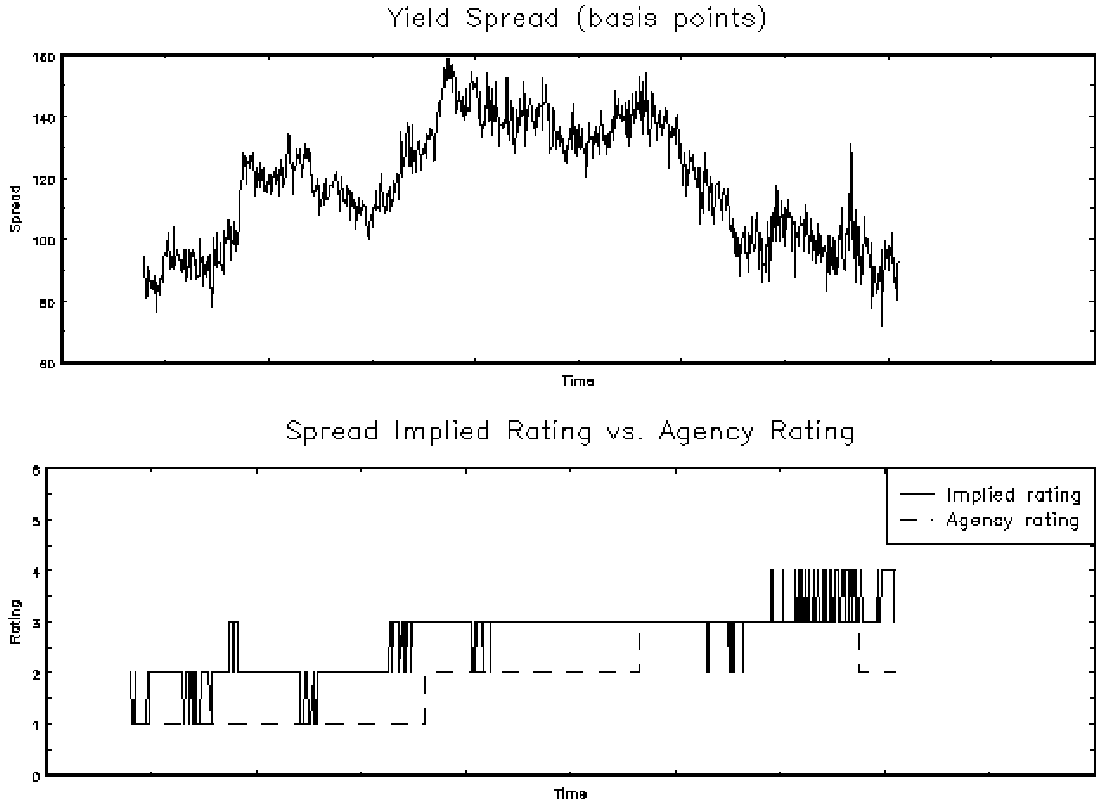
Panel A Boundaries shown separately



Panel B: Boundaries shown together

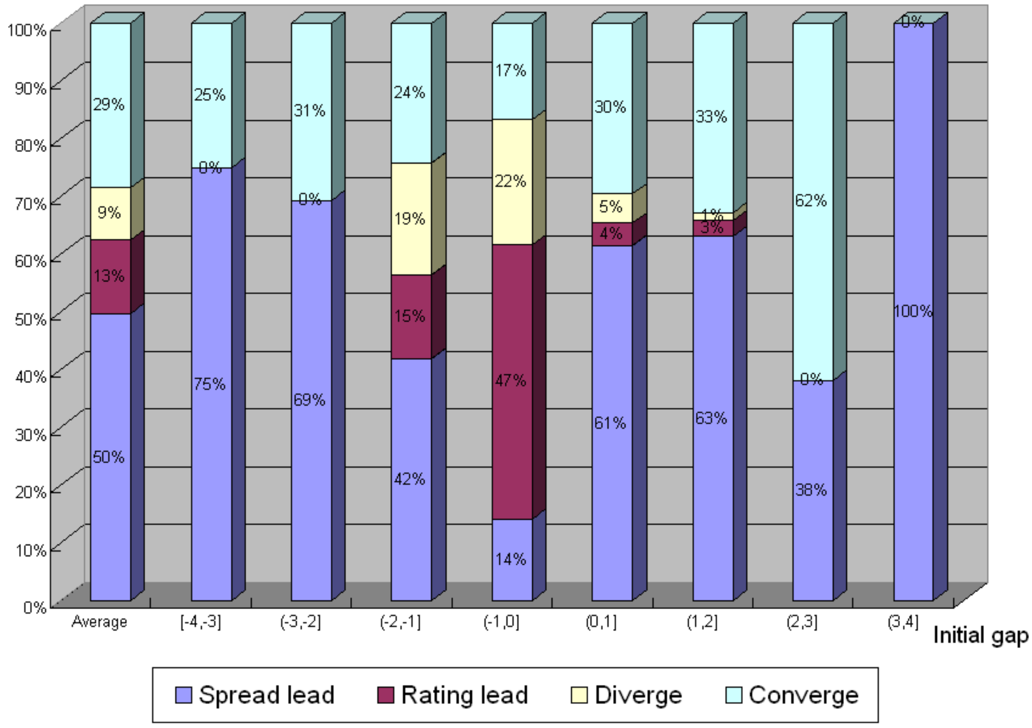


Graph 5 Yield spreads, spread implied rating and agency rating



Note: In the scale for rating, 1 stands for AAA, 2 for AA, 3 for A, 4 for BBB, and 5 for BB and below.

Graph 6a Behaviour patterns of S&P ratings and spread implied ratings conditional on their initial difference



Graph 6b Behaviour patterns of Moody's ratings and spread implied ratings Conditional on their initial difference