Credit ratings and credit risk

Jens Hilscher[*] Mungo Wilson[†]
hilscher@brandeis.edu mungo.wilson@sbs.ox.ac.uk

This version: September 2009

Abstract

This paper investigates the information about credit risk contained in corporate credit ratings. Default probabilities, calculated using a reduced form logit model, are better than ratings at forecasting corporate failure and explain a substantial share of the within-rating variation in CDS spreads. Meanwhile, ratings explain little of the large variation in default probabilities across firms. During and after recessions and financial crises, lower-rated firms become relatively more risky, which implies that credit ratings capture systematic default risk. We use failure beta, the sensitivity of default probability to median default probability, as our measure of systematic default risk. Failure beta is strongly related to credit rating and explains a large share of the variation in CDS risk premia.

JEL Classification: G12, G33

Keywords: credit rating, default risk, forecast accuracy, systematic default risk.

[*]International Business School, Brandeis University, 415 South Street, Waltham MA 02453, USA. Phone +1-781-736-2261.

[†]Oxford University, Saïd Business School.

We would like to thank Robert Jarrow and Don van Deventer of Kamakura Risk Information Services (KRIS) for providing us with data on corporate bankruptcies and failures, and John Campbell, Steve Cecchetti, David Scharfstein, Andrei Shleifer, Robert Whitelaw, Moto Yogo, and seminar participants at Royal Holloway, University of Zürich, LSE, Humboldt Universität zu Berlin, CEMFI, and Brandeis for helpful comments and discussions. Both authors were on leave at the LSE when the first version of this paper was written and would like to thank LSE for its hospitality.

1 Introduction

Despite recent criticism, credit ratings remain the most common and widely used measure of corporate credit quality. Pension funds, banks and insurance companies use credit ratings as investment screens and to allocate regulatory capital. Debt covenants may be contravened and default triggered by a ratings downgrade. More than ever this raises the question of whether ratings are suitable for these purposes. While the apparent failure of credit ratings agencies to correctly value complex credit products has received recent attention (see, for example, Coval, Jurek, and Stafford (2008)), less attention has focused on the accuracy of ratings in what is arguably their core competence: assessing corporate credit risk. In this paper, we examine the information provided by credit ratings, their ability to forecast corporate failure, and their ability to explain variation in market prices.

We begin by assessing the accuracy of credit ratings as a forecast of corporate failure. Recent advances in default prediction call for a reassessment of ratings in this regard. Building on the models in Shumway (2001), Chava and Jarrow (2004), and Campbell, Hilscher and Szilagyi (2008),[1] we evaluate credit ratings' ability to predict default relative to a benchmark in a hazard model.

We find that a straightforward benchmark predictor of default based on accounting data and stock market prices ('failure score') is much better at predicting default at horizons of up to 2 years and that credit rating adds very little information to failure score at horizons up to 5 years. When predicting failure at any time over the next 1 to 5 years, failure score is also more accurate than ratings. Our results are not driven by ratings not capturing time variation in default probabilities or by ratings' discrete ranking, their "broad brush" nature.

We find that CDS spreads reflect information about default probabilities not contained in ratings. Failure score explains an important amount of variation in CDS spreads within identically rated firms, and predicts rating downgrades. By contrast, we find that rating

[1] These papers build on the seminal earlier studies of Beaver (1966), Altman (1968), and Ohlsen (1982). More recent contributions to the long and rich literature on using accounting and market-based measures to forecast failure include Beaver, McNichols, and Rhie (2005), and Duffie, Saita, and Wang (2007).

explains little of the variation in default probability as measured by failure score (virtually none among investment grade firms). There is considerable, almost total, overlap of default probability distributions across investment-grade ratings: the median BBB-rated firm's failure score-based default probability is always lower than the 75th percentile probability for the A-rated and the 90th percentile of AAA-rated firms.

There is important time variation in failure probabilities not captured by ratings. Within a given rating class failure probability varies over time in a way that is not related to migrations across ratings. During and after recessions and financial crises ('bad times') failure probabilities are 30% (investment grade) and 80% (non-investment grade) higher than outside of these times. The ranking of median default probability based on failure score is largely preserved across ratings, but the large overlap across ratings remains. We conclude that credit ratings are a reasonable, if noisy, relative measure of default probability, but a poor absolute measure.

From an investor perspective, credit risk relates to both default probability and systematic risk: the tendency to default in bad states of the world. Given two bonds with identical default probabilities, the bond that is more likely to default in bad times should have a higher yield. If ratings are a single, summary measure of credit risk, they should reflect both concerns to some extent. Ratings agencies themselves appear to conflate default probability and systematic risk: Standard & Poor's website states that a AA rating means that a bond is, in the agency's opinion, "less likely to default than the BBB bond."[2] On the same web-page, the agency states that a speculative-grade rating "factors in greater vulnerability to down business cycles".

The distinction between default probability and systematic risk can also be understood in terms of expected payoff versus risk premium. Expected payoff depends on the probability of default and expected loss given default, while risk premium, which determines the rate at which expected bond payoffs should be discounted, depends only on the systematic risk of the bond.[3]

[2] "... [A] corporate bond that is rated 'AA' is viewed by the rating agency as having a higher credit quality than a corporate bond with a 'BBB' rating. But the 'AA' rating isn't a guarantee that it will not default, only that, in the agency's opinion, it is less likely to default than a 'BBB' bond."

[3] Alternatively, the distinction can be understood as being the difference between objective and risk-neutral default probabilities.

The realized default rate of all firms, and in particular of all rated firms, is much higher during and immediately after recessions and financial crises. Reflecting this fact, our default probability measure appears to contain an important systematic component: the default probability of a typical firm increases strongly in recessions and financial crises. We therefore propose the sensitivity of default probability to median firm default probability, or failure beta, as a straightforward measure of a firm's systematic default risk. We find that failure beta is strongly related to rating. Our results suggest that, in the case of corporate credit risk, credit ratings are at least as informative about systematic risk of default, or bond risk premia, as about probability of default, or expected payoffs. Furthermore, default probabilities and failure betas are each important in explaining variation in CDS spreads. Therefore, it seems that the reason that credit ratings can explain a large share of the variation in CDS spreads is that ratings measure systematic risk.

Our conclusion is that ratings convey information about both default probability (payoff) and systematic risk (risk premium). This conclusion can explain: (1) why ratings are not particularly accurate at forecasting default: to do so is not their sole purpose; (2) why agencies 'rate through the cycle:' If systematic risk equals "vulnerability to down business cycles," it cannot vary over the business cycle, so neither can rating to the extent rating reflects systematic risk; (3) why investors are interested in ratings and why variation in borrowing costs is strongly related to rating; investors care both about expected payoff and about risk premia.

A large early literature evaluates the ability of ratings to predict default, beginning with Hickman (1958). We discuss this literature in the next section. In the context of credit ratings of tranched portfolios secured on pools of underlying fixed-income securities, such as collateralized debt obligations (CDOs), the distinction between default probability and systematic risk has been made by Coval, Jurek, and Stafford (2008) and Brennan, Hein, and Poon (2007). However, both papers assume that ratings relate only to default probability or expected loss and proceed to show how this can lead to mis-pricing. The idea that both default probabilities and risk premia affect bond prices and CDS spreads is well understood, e.g. Elton, Gruber, Agarwal, and Mann (2001). Equivalently, studies have shown that prices depend on both objective and risk-neutral probabilities (Chen (2008), Bhamra and Strebulaev (2007)).

However, none of these papers relate their findings to ratings, other than using ratings as a control. In the context of evaluating Basel II implementations, van Deventer, Li, and Wang (2005) compare accuracy ratios of S&P credit ratings to a reduced form measure of default probability and Cantor and Mann (2003), as well as subsequent quarterly updates of this study, evaluate the ability of Moody's credit ratings to predict bankruptcy relative to various alternatives.

The rest of the paper is organised as follows: the next section describes our data and failure prediction methodology; section 3 presents our main results on credit rating and default probability and then discusses the results of a large range of robustness tests of these main results; section 4 relates ratings to sensitivities to bad times, as opposed to levels of default probabilities, and presents evidence that (i) such sensitivities capture an important component of variation in ratings and (ii) such variation is priced in the cross-section of CDS spreads; the last section concludes.

2 Default probability

2.1 Data description: Corporate failures and explanatory variables

Our method for predicting default is the same as in Campbell, Hilscher, and Szilagyi (2008) and builds on the earlier work of Shumway (2001) and Chava and Jarrow (2004). We use a hazard model to predict corporate failures from January 1963 to December 2008. We use the same failure indicator as in Campbell et al. The data was provided to us by Kamakura Risk Information Services (KRIS). The indicator includes bankruptcy filing (chapter 7 or chapter 11), de-listing for performance-related reasons, D (default) or SD (selective default) rating, and government-led bailout. The broad definition of failure allows us to capture at least some cases in which firms avoid bankruptcy through out-of-court renegotiations or restructurings (Gilson, John, and Lang (1990) and Gilson (1997)), or cases in which firms perform so poorly that they delist, often before subsequently defaulting.

Table 1, panel A reports the number of firms and failure events in our data set. The second column counts the number of active firms, which we define to be those firms with some available accounting or equity market data. We report the number of failures over time in column three, together with the failure rate in column four, the percentage of active firms that failed each year. We then repeat this information for those firms with a S&P credit rating in columns five through seven. Since our data on credit ratings begin in 1986 we mainly focus on reporting statistics for the period from 1986 to 2008.

It is apparent that failures are rare events. For the majority of years the failure rate is smaller than 1%, and the average failure rate from 1971 to 2008 is also below 1%. From 1986 to 2008 the failure rate exhibits strong variation over time. It is high in the second half of the 1980s, reaching 1.7% and 1.5% in 1986 and 1988, respectively. This variation over time is at least partly related to recessions and financial crises. The average failure rate during and in the 12 months after NBER recessions is equal to 1.5%. In the 12 months after the October 1987 stock market crash and the September 1998 Russian and LTCM crisis the failure rate is equal to 2%. Both of these are higher than the 0.8% failure rate outside of recessions and crises.

The pattern for rated firms is very similar. The universe of rated firms is much smaller; only 18% of active firms are rated on average. However, rated firms tend to be much larger which means that the average share of liabilities that is rated is equal to 76%. To proxy for this, in the last column we report the share of total book value of liabilities that is owed by firms with an S&P credit rating. The share of rated firms and the share of liabilities rated has increased over time – from 2004 to 2008 23% of firms and 86% of liabilities are rated. The failure rate for rated firms is almost three times higher during and immediately after recessions (2.4%) and crises (2.3%) than it is outside of these times (0.9%).

To our history of failure events we add measures of financial distress. We construct explanatory variables using accounting and equity market data from the daily and monthly CRSP files and add to that quarterly data from Compustat. The explanatory variables we use measure profitability, leverage, past returns, volatility of past returns, firm size, firm cash holdings, and firm valuation. We include the following explanatory variables in our failure prediction model: $NIMTAAVG$, a weighted average of past quarterly ratios of net income to market value of total assets; $TLMTA$, the ratio of book value of total liabilities to market value of total assets; $EXRETAVG$, a weighted average of past monthly log returns relative to the S&P 500 value-weighted return; $RSIZE$, the log ratio of a firm's market capitalization to that of the S&P 500 index; $SIGMA$, the standard deviation of the firm's daily stock return over the previous 3 months; $PRICE$, the firm's log price per share, truncated above at a price of \$15 per share; $CASHMTA$, the ratio of cash to market value of total assets and MB the market-to-book ratio of the firm. Together, these variables, and a constant, make up the vector x_{it}, which we use to predict failure at different horizons. All of these variables as well as the specification and estimation procedure are discussed in Campbell et al. who show that this specification significantly outperforms other standard methods of default prediction.

2.2 Predicting failure in a logit model

We assume the month-t marginal probability of failure in month $t+s$ follows a logistic distribution. We allow the coefficients, the relative weights of the different predictor variables, to depend on the horizon over which we are predicting failure. The conditional probability of failure is given by:

$$P_t(Y_{i,t+s} = 1|x_{it}) = (1 + \exp(-\delta_s' x_{it}))^{-1} \tag{1}$$

where $Y_{i,t+s}$ is an indicator variable that equals one if firm i fails in month $t+s$ conditional on not failing earlier, x_{it} is a vector of our explanatory variables, including a constant, observed at the end of month t, and $\delta_s' x_{it}$ is a linear combination of these explanatory variables. We refer to this linear combination as the 'failure score' of firm i in month t. Failure score and failure probability are then (positively) related by equation (1).

Assuming independence of default in each month, the probability that a firm defaults between month t and month $t+s$ is then one minus the probability of survival for s months:

$$P_t(Z_{i,t,t+s} = 1) = 1 - \Pi_{j=1}^{s} \left(1 - P_t(Y_{i,t+j})\right) \tag{2}$$

where $Z_{i,t,t+s}$ equals one if firm i defaults between month t and month $t+s$.

To predict corporate failure we use the same model as in Campbell et al. Table 2 reports results from estimating a logit model using data from 1963 to 2008. We predict failure over the next month (column (1)) and in 12 months (column (2)). The explanatory variables are related to failure as we would expect. Firms are more likely to fail if they are less profitable, have higher leverage, lower and more volatile past returns, and lower cash holdings. Firms with lower price per share are more likely to fail. At the 1-month horizon, relative size enters with a counterintuitive positive coefficient and is the least significant of all variables, most likely a result of the correlation between size and price. Market-to-book enters with a positive sign. At the 12-month horizon, the results are similar, except that size switches its sign. As measures of fit we report McFadden's pseudo R^2 and the accuracy ratio of the model. The accuracy ratio measures the relative frequency of correct predictions (true positive and true negative) relative to incorrect predictions (false positive and false negative) and is a summary measure of the receiver operating characteristic (ROC) curve. It is a commonly used measure when evaluating a binary response model. We discuss this measure further in the next section. For the 1-month and 12-month models the pseudo R^2 is equal to 31.5% and 11.8%. The sharp drop in prediction accuracy is probably not surprising given the increased horizon of the forecast. It is also reflected in the drop in accuracy ratio from 95.5% to 86.0%.

3 Failure score and credit rating

We now compare our failure score with S&P long-term general corporate credit rating as a predictor of default. Data on monthly S&P credit ratings are from Compustat.[4] To investigate the relative performance of credit rating and failure score, we add rating as an additional explanatory variable in our hazard model. For our first set of results we estimate:

$$P_t(Y_{i,t+s} = 1) = (1 + \exp(-\alpha_s - \gamma_s \left(\delta_s' x_{it}\right) - \phi_s Rating_{it}))^{-1} \tag{3}$$

[4] S&P also supply short-term ratings, but these cover a much smaller sample of issuers. We have checked that our results on prediction accuracy are robust to the inclusion of short-term credit ratings. We have also compared our results to the forecast accuracy of Moody's ratings, as reported in Cantor and Mann (2003), and concluded that our results are not driven by using S&P credit ratings.

We restrict the coefficients δ_s' to equal their estimates obtained when including data for all listed firms, as opposed to only those that are rated. This means that the coefficient vector δ_1 contains the coefficients reported in table 1, column 1. For longer horizons we use the equivalent longer-range estimates. In other words, we estimate a failure score for all listed firms and then estimate how much additional information is contained in ratings regarding the failure prospects of rated firms. This sets the bar for failure score a little higher than just estimating an unrestricted regression with rating as an additional firm characteristic.

S&P credit ratings run from AAA (to which we assign a score of 1), through BBB (the lowest investment grade letter rating, with a score of 9), to C (one notch above default, to which we assign a score of 21). Ratings ranging from AA to CCC are divided into 3 subgroups each with a '+' or a '–' added to the rating (e.g. A+, A, A–). Each reduction in rating receives an additional score of 1, so that our rating variable, like failure score, is positively related to default probability.

3.1 Relative forecast accuracy of credit ratings and failure score

Table 3 reports the results from our estimation of the baseline model in equation (3). Panel A reports pseudo R^2 and accuracy ratios. We report results for specifications with only failure score, only rating, and both failure score and rating. We focus specifically on the ability of different measures to forecast failure at different horizons and consider 1, 3, 6, and 12-month horizons, as well as 2, 3, 4, and 5-year horizons. Panel B reports the coefficient estimates and their associated z-statistics for the specification with both failure score and rating. If rating contains no additional information relative to failure score, or if the additional information in rating is orthogonal to the information in failure score, and if rated and unrated firms are similar, then the coefficient on failure score γ_s should be equal to one, and both the coefficient on rating ϕ_s and the constant α_s should be equal to zero.

Failure score predicts default at horizons of one month with a pseudo R^2 of 39.9% versus 29.2% for rating alone, which means that failure score outperforms rating by 10.7 points. Although both failure score and rating are each statistically significant predictors, adding rating to failure score only increases the pseudo R^2 from 39.9% to 42.5%. Thus, rating appears to contain little additional information about the probability of failure in the immediate future, while failure score significantly outperforms rating. In panel B the coefficient on failure score is estimated to be 0.852.

It should not be too surprising that failure score is a good forecast of default one month into the future: most investors should presumably be aware of impending disaster at such short horizons and equity market data, such as e.g. past returns and volatility, will likely reflect their awareness. However, all the information available to the market is also available to the ratings agencies, so we can conclude from this result that, whatever else ratings may be, they are not optimal forecasts of default in the short run.

In the medium term, at horizons of 3, 6, 12, and 24 months, failure score continues to outperform rating substantially. The difference in pseudo R^2 between using failure score only and rating only is equal to 14.1 and 12.1 points at the 3 and 6 month horizons, which means that it is even larger than at the 1 month horizon. One reason may be that the rating of a failing firm is often downgraded very close to the actual failure but that the rating does not accurately reflect the risk of failure at intermediate horizons. At 12 and 24 months failure score outperforms rating by 8.1 and 0.8 points respectively. For horizons of 3 to 12 months the coefficient on failure score is statistically indistinguishable from 1 at the 1% confidence level, and is equal to 0.82 at 24 months.

The lower part of panel A in table 3 reports accuracy ratios for the same regressions. Accuracy ratio measures the tendency for the default predictor to be higher when default actually subsequently occurs (true positive) and lower when default subsequently does not occur (true negative). It is a useful non-parametric measure of model performance and varies from 100% (perfect model) to 50% (random model).[5] We find that using accuracy ratio as

[5]Specifically, the accuracy ratio is the area under the ROC curve. For each level of the predictor variable, e.g. failure score, the ROC curve plots the true positive rate against the false positive rate. For example, if all of the highest 340 one-month failure scores were observed one month before the failures of the relevant firms, for any cutoff level lower than the sample failure rate, there would be no false positives. As a result, the accuracy ratio would be 100%. In contrast, for a random model, the true positive and false positive rates are equal for any cutoff value; in the example, if the cutoff value is the 90th percentile of the failure score, both the true postive and the false positive rates are equal to 1/9 and the accuracy ratio is equal to 50%. For a more detailed discussion of accuracy ratio in the context of default prediction see Vassalou and Xing (2004).

a measure of predictive performance results in the same pattern across prediction horizons and across the three models (failure score only, rating only, both), except that failure score outperforms rating at horizons up to 36 months, versus 24 months for the pseudo R^2 criterion.

The ability of either failure score, rating, or both to forecast failure declines monotonically with the forecast horizon. Using both measures, the pseudo R^2 declines from 42.5% at the 1-month horizon to 5.6% at the 60-month horizon. The same is true for using either failure score only or rating only. For longer term prediction, the accuracy of all measures is lower, but also rating becomes relatively more important. At 36 months both measures have close to the same forecast accuracy; the pseudo R^2 is 0.3 points higher for rating only while the accuracy ratio is 0.5 points higher for failure score only. For the 4 and 5-year horizons rating only is a more accurate predictor than failure score only. Nevertheless, failure score remains a statistically significant predictor of default and using both measures is always better in terms of accuracy than using only one of the measures. The significance levels of credit rating and failure score, when both are included, reflect the relative performance of the individual measures. Both are statistically significant at all horizons, but failure score is much more significant up to about 2 years, significance levels are the same at 3 years, and rating is more significant at 4 and 5 years. Thus, at horizons up to 2-3 years, failure score performs better as a default predictor than rating but rating performs slightly better at longer horizons.

The superior performance of rating at predicting failures at horizons of more than 3 years, conditional on no earlier failure, is not enough to make up for the much greater predictive power of failure score earlier on when evaluating each as a predictor of default between date t and date $t + s$. Figure 1 plots the pseudo R^2 for all horizons from our baseline model for failure score only, rating only and both together. The area under each line can be thought of as an estimate of the ability to forecast default over time, rather than at some future point. The area under the 'both' line is not much greater than under the line for failure score alone, while it is clearly larger than the area under the line for rating alone.

We also consider the relative accuracy of failure score and rating more formally. For each January we construct cumulative failure events for the following 1, 2, 3, 4, and 5 years. We then use rating and 12-month failure score as predictors of default. Pseudo R^2 and accuracy ratios decline monotonically with the horizon but are always higher for failure score only than for rating only. Panel C of table 3 reports the pseudo R^2 measures. At horizons of one year, failure score predicts 40.9% correctly versus 24.1% for rating, a difference of 16.8 points. Adding rating to failure score increases pseudo R^2 from 40.9% to 42.3%. At 5-year horizons, failure score predicts correctly cumulative failure events 23.6% of the time versus 18.0% for rating only, a difference of 5.6%, and outperforms at all intervening horizons. Adding rating to failure score increases pseudo R^2 from 23.6% to 26.9% at 5-year horizons.[6] We conclude that by ignoring or not responding to publicly available early warning signals of default at horizons of up to 3 years, rating fails as an optimal forecast of default at horizons of up to 5 years.

3.2 Robustness of relative forecast accuracy

We now investigate the robustness of our conclusions to a range of other possibilities in investigating relative forecast accuracy. We briefly discuss the reason for each robustness test as well as the results of performing it.

First, we check if our results are driven by look-ahead bias and consider the ability of the model to predict failure out-of-sample. Since we estimate failure score using data from 1963 to 2008 and compare rating and failure score from 1986 to 2008, there is a large overlap in the sample period. We perform two checks: First, we consider estimating coefficients on failure score for the same period as in Campbell et al. (1963 to 2003) and then test relative out-of-sample performance using data on failure events only from 2004 to 2008. In doing so the data used to construct the independent variable (the estimate of the coefficients for the vector δ_s) and the data used for the dependent variable (the failure indicator) do not overlap. Thus this is a genuine out-of-sample test (as opposed to a pseudo out-of-sample test) of the ability of the model to predict corporate failure, given the earlier results in Campbell et al. We find that the relative difference between failure score and rating is larger during this period than for the full sample used in table 3. Next, we compare failure score, estimated recursively, to credit rating. We re-estimate the model each year from 1986 to 2007, updating the estimates of δ_s, and

[6] The pattern in accuracy ratios, which we do not include in the table, is similar.

use those coefficients to predict failure during the following year. We then compare forecast accuracy of failure score only and rating only. Our results are not significantly affected by this alternative procedure. We conclude that failure score is a superior predictor of corporate failure both in and out-of-sample.

Second, the superior performance of failure score could be due to the discrete nature of credit rating, and our comparing it, perhaps unfairly, to a continuous failure score. To address this possibility we discretize our failure score measure and compare its performance with rating using the same procedure as we used for the continuous version. We choose our discretization so that the size of a group with a common (discrete) failure score accounts for the same proportion of the rated sample as the group with a common rating. For example, the number of observations of firms rated AAA corresponds to the size of the group with the lowest failure score. We then assign scores of 1 to 21 to these groups. The discretized failure score predicts default at a similar level of accuracy as continuous failure score and consequently performs as well relative to ratings.

Third, our results may be driven by the inability of ratings to capture variation in aggregate default rates. From the results in table 1 we know that there are significant differences in the failure rate over time. However, there is no corresponding change in ratings, given that ratings 'rate through the cycle' (Amato and Furfine (2004), Löffler (2004)).[7] It is possible, therefore, that the forecast accuracy of ratings would improve if we were to allow predicted average default rates to vary over time. We investigate this hypothesis in three ways: First, we include separate dummy variables for recessions and financial crises and compare relative performance. Second, we include median failure score together with rating. If failure score reflects time variation but ratings do not, adding median failure score to rating should reduce this disadvantage. Third, we include time dummies together with ratings and failure score. Since there are several years with only very few events, we include two-year dummies for estimation purposes. None of these alternative specifications significantly affects the results in table 3.

[7] We discuss this point more in section 4 of the paper where we also present additional evidence that ratings alone do not capture time variation in default rates.

Fourth, our results may be driven by not accounting for possible non-linearities in the relationship between rating and observed failures. We include rating linearly in the logit model and using a different functional form may lead to an increase in forecast accuracy. Although such a change may increase the pseudo R^2, it will not affect the accuracy ratio of the predictor since any monotonic transformation of rating will lead to the same classification of predicted failures and therefore have the same accuracy ratio. To investigate whether or not the pseudo R^2 is affected we include rating dummies instead of rating score.

From an estimation point of view it is not possible to include a different dummy variable for each rating. Some ratings have very low frequencies of failures, and some have no observed events. It is therefore necessary to group observations together. Grouping ratings also helps with the possibility that the relationship between rating and failure may not be monotonic. For example, it may be that in the data B- rated firms are more likely to default than CCC+ rated firms. We group firms into 10 groups by rating and estimate a logit model allowing the coefficient on the dummy variables to vary freely. Again, we find that failure score outperforms rating by a substantial margin in predicting default.

We conclude that our results are robust to look-ahead bias and out-of-sample evaluation, discretization, non-linearities in the effect of rating, and time effects.

3.3 Variation of default probability by rating

Having established that failure score is superior to ratings as a predictor of default and corporate failure, we now investigate further the information contained in ratings relevant to assessing future default. We do this by comparing fitted default probabilities across credit ratings. In table 4 we calculate default probability for seven rating groups, grouping ratings by letter ratings and combining firms rated CCC and below. We report the share of observations by rating group, and the mean and standard deviation of the annualized 12-month default probability. There is substantial variation in average default probability across ratings and the ranking of fitted default probabilities matches the ranking of ratings: The 1.5% of observations rated AAA have a mean annualized default probability of 0.4% while the 1.6% of observations rated CCC or below have a mean default probability of 8.2%. In between, mean default probability increases monotonically going from AAA to CCC. There is also substantial within-rating variation, which is larger in magnitude to the mean default probability of the rating. For investment grade firms the mean default probability is equal to 0.43% (AA), 0.49% (A), and 0.60% (BBB), while the standard deviations are equal to 0.65% (AA), 0.54% (A) and 0.73% (BBB). This large within rating dispersion suggests that ratings do not clearly separate firms into categories by default probability.

Furthermore, ratings do not capture variation in default probability over the business cycle. For all rating groups the default probability is much higher during and after recessions and financial crises (bad times) relative to the average values outside of those times (good times). In addition to the overall increase in default probability, differences across ratings become larger during bad times: for example the difference in mean default probability between AA and BBB rated firms is 0.13% in good times and 0.24% in bad times.

We also look at the distribution of failure probabilities across credit ratings. Figure 2 presents box plots of default probability by rating. Each box plot is a vertical line showing the 10th percentile, median, and 90th percentiles as horizontal bars, with the interquartile range as a grey box. The highest-rated firms are closest to the origin and have the lowest default probabilities. Specifically, we plot the base-ten logarithm of the annualized 12-month failure probability for firm-months with a given rating. To facilitate comparison across time, we have subtracted from every failure score the annual median across all rated firms. This way the variation in default probability by rating is not driven by common variation in default probability over time, which we discuss more in section 3.6.

Three obvious inferences can be made from Figure 2. First, the ranking of distributions by rating is broadly correct: all points of the distribution, more or less, increase monotonically as rating increases from AAA to CC. Second, however, there is considerable overlap across ratings. For example, the median default probability for any rating is always lower than the 75th percentile for the next higher rating, or even for that two notches higher. Third, the overlap in distributions is much more obvious for investment grade issuers: there appears to be almost total overlap for issuers rated between 3 (AA) and 10 (BBB-). So much overlap, in fact, that for some adjacent ratings or even ratings two notches apart, we are unable to reject the hypothesis that their mean default probabilities are the same. In fact, the 75th percentile AA-rated issuer, two notches below AAA, is more likely to default than the median BBB-rated issuer, two notches above junk. Therefore, the decline in distribution is mainly for non-investment grade issuers.

It appears that, especially for investment-grade issuers, credit ratings are not strongly related to our default probability estimate, even though the latter is significantly better at predicting corporate failure, both in and out-of-sample. We can also capture the ability of rating to explain variation in default probability by running a regression of log default probability on rating. For non-investment grade issuers, rating only explains 19% of the variation in default probability. Although the R^2 is low, rating and failure probability are clearly related. However, for investment grade issuers, the relationship is even weaker: Credit rating only explains 4% of the variation in default probability.

Of course, our failure score is a noisy estimate of the true propensity to fail, even though statistically and economically superior to rating as a predictor of default. For this noise to explain our results, however, it must be the case that failure scores are noisier for investment grade issuers, where the spread of the distribution, and therefore the dispersion of estimates, is smaller. To investigate the variation in default probability further we now relate default probability to CDS spreads.

3.4 Default probability and CDS spreads

If variation in default probability is viewed by market participants as being informative it should be reflected in market prices. We now examine this hypothesis by considering the ability of default probability to explain variation in credit default swap (CDS) spreads. We specifically focus on variation across issuers within the same rating.

We use monthly 5-year CDS spreads from 2001 to 2007, obtained from Markit Partners. Our sample consists of all rated firms for which we are able to construct a failure probability resulting in a sample of over 38,000 firm-months. CDS spreads can be thought of as the spread over a default-free bond of equivalent maturity that a given issuer must pay. Intuitively, the spread should reflect both compensation for expected loss (probability of default times expected loss given default) and a risk premium. At this point we are only interested in the extent to which variation in spreads across issuers and over time can be attributed to objective failure probability. We return to the second issue in the next section

Table 5 investigates the ability of log 12-month failure probability to explain log spread variation within rating groups. We again group ratings into 7 groups (there are sometimes too few firms with a given rating to do otherwise): AAA, AA, A, BBB, BB, B and CCC and below. We regress log spread on log failure probability and credit rating. We control for rating within rating group, but assume a linear relation between rating and spread. As before, we compare overall model fit for three cases: failure probability only, credit rating only, and both.

Log 12-month failure probability explains a significant amount of variation in CDS spreads within these ratings groups, varying from about 19% for the 253 A-rated issuers in our sample to about 34% for the 47 AA issuers, and 30-47% for BBB, BB, or B rated issuers. In contrast, using rating to explain the variation results in R^2 of 0.01% (AA) to 13% (BBB).

We also find that failure probability is significantly related to spread for all ratings groups except for CCC and below, for which it is significant at the 10% level.[8] For AAA issuers, a 1% increase in the 12-month failure probability is associated with a 0.86% increase in spread. Coefficients on failure probability tend to decline with rating but are still broadly similar in magnitude across ratings. The coefficients for the ratings with the most observations (A, BBB, BB, and B) range from 0.52 to 0.67.

The relative stability of the coefficients across ratings suggests that we can pool observations across different credit ratings into one regression. This means that we are assuming a linear relationship between log spread and log default probability.[9] In the last column of the table we estimate pooled results for the entire panel. Default probability remains significant in the pooled regression and the coefficient magnitude is comparable to the coefficients in the rating

[8] Standard errors are clustered by year to take into account the possibility of persistent cross-sectional correlation. Our results are robust to clustering by firm instead.

[9] The assumption of linearity is consistent with an earlier study by Berndt, Douglas, Duffie, Ferguson and Schranz (2008). They show that log CDS spreads and log of Moody's EDF, a distance-to-default-based measure of default probability, are strongly linearly related for the broadcasting and entertainment, healthcare, and oil and gas industries.

group regressions. In this context we are combining both variation within and across ratings which means that we cannot evaluate the ability of default probability to explain within rating variation by considering the regression R^2. To do this we next run panel regressions.

Table 6 presents pooled regression results that include different sets of fixed effects. We again regress log spreads on log failure probabilities. Rating fixed effects alone explain 64.5% of total variation in log spread in our panel. Comparing this to the pooled regression in table 5 we notice that it is only slightly higher than a specification that forces the relationship between spread and rating to be linear, which explains 62.8% of all variation. It seems, therefore, that the relationship between log spreads and ratings is close to linear. We discuss the implications of this point further in the next section.

Controlling for rating fixed effects, log 12-month failure probabilities are highly economically and statistically significant determinants of CDS spreads. A 1% increase in 12-month failure probability is associated with a 0.59% increase in spread. The overall R^2 increases from 64.5% to 75.4% when log 12-month failure probability is added to a model containing just rating fixed effects.

Within firms with the same rating, 12-month log failure probability explains an average of 30.7% of variation in log default spreads, and 80.2% in variation of the average log spread across ratings groups. Even when year and firm fixed effects are added (and 87% of overall variation is captured by these three fixed effects), log failure score is still economically and statistically significant. The coefficient in the final column is 0.454 and there is still a non-negligible increase in overall R^2 from 87% to 89.9%.

We conclude that our estimates of default probability, although undoubtedly noisy, explain significant variation in default spreads and are economically related to spreads.

3.5 Default probability and downgrades

Our estimates of default probability also predict ratings downgrades. We run the same logit regressions as before, but now the dependent indicator variable equals one if a rated firm is downgraded during the next month. We use 12-month failure score to predict such events. We run the regressions separately for each letter ratings group. Table 7 reports the results:

a higher failure score predicts an increased probability of a downgrade and all the coefficients are statistically significant. The pseudo R^2 is 1.8% for AAA, 5.1% for AA, 3.4% for A, and then increases monotonically as rating declines, to about 15% for B and CCC. The accuracy ratio is about 65% for A-rated issuers or above, increasing to 73% for BBB, 75.5% for BB and over 80% for lower-rated issuers.

3.6 Variation in default probability over time

Having established that our measure of default probability is better at forecasting corporate failure, that it is reflected in market prices, and that it forecasts downgrades, we next use our measure of default probability to examine variation in default probability by rating over time. We already know that both average default probability for all ratings and realized failure rates for rated firms are higher during recessions and financial crises (tables 1 and 4). We now examine time variation in default probability further.

Figure 3 plots median annualized 12-month failure probabilities (roughly, the probability of default in the next 12 months) over time, since 1986, for the 5 different ratings categories – the 5 letter ratings with the most available observations (AA, A, BBB, BB, and B). Although the ranking of failure probability by rating is preserved over time, the probability of failure of a typical firm in a given rating class rises dramatically in recessions and financial crises. If rating corresponded to probability of failure, the lines in figure 3 should be roughly flat and parallel. The lines also appear to spread out in recessions and crises, suggesting rating partly reflects sensitivity of credit risk to bad times.

Several previous studies find that default probabilities vary counter-cyclically. See e.g. Fons (1991), Blume and Keim (1991), Jonsson and Fridson (1996), McDonald and Van de Gucht (1999), Hillegeist, Keating, Cram, and Lundstedt (2004), Chava and Jarrow (2004), and Vassalou and Xing (2004).

It is well established that ratings are meant to 'rate through the cycle' (Amato and Furfine (2004), Löffler (2004)), which means that the share of ratings in a particular rating group will not vary directly with business conditions. In other words, an increase of the overall default probability during bad times is not necessarily associated with increased number of downgrades relative to upgrades. To check that this is the case in our data, we plot the share of firms rated AA, A, BBB, BB, and B in figure 4. Although there is a clear decline over time in the share of firms rated AA and A (also see Blume, Lim, and MacKinlay (1998)), there is no clear tendency of the share of lower-rated issuers to increase during and after recessions and financial crises. This means that although 'rating through the cycle' implies that ratings cannot measure variation in default probabilities through time, figure 3 demonstrates and quantifies the inability of rating to reflect fluctuations in default probability over time.

4 Systematic risk and failure beta

We now consider whether or not, in addition to default probability, rating also reflects systematic risk. The previous section presented evidence that a simple linear combination of variables based on publicly available accounting magnitudes and equity prices is a significantly better predictor of corporate default and failure than credit rating at horizons of up to 5 years. Failure score also explains substantial variation in CDS spreads within firms of the same rating. Furthermore, although rating appears to imply the correct ranking of distributions of failure scores, the overlap is so large, especially for investment grade issuers, that rating does not appear to be particularly informative about the probability a given firm will fail. Put differently, the evidence suggests that rating does not reflect information about default probability contained in measures that market participants and studies in the academic literature have used.

This raises the possibility that ratings measure something other than the objective probability of failure over some time-span. When determining market prices a bondholder cares not only about default probability (expected payoff) but also about systematic risk (discount rate). In fact, S&P on its website suggests that its rating reflects both things: AA is described as having a "very strong capacity to meet financial commitments" while BBB is described as "Adequate capacity to meet financial commitments, but more subject to adverse economic conditions."

We now examine if rating measures systematic risk. That is, we ask if rating is related to the tendency to fail in bad times. The evidence in table 4 and figure 3 – the tendency of lower-rated issuers' default probability to increase by more in bad times – is consistent with this hypothesis.

4.1 Failure beta

We measure systematic default risk as the tendency of default probability to increase during bad times. Our preferred measure of systematic risk is failure beta: the sensitivity of default probability to median default probability. Intuitively, firms that tend to default in bad times, when the stochastic discount factor is above its mean, have a higher systematic risk, holding constant their expected loss. In figure 5 we graph median and value-weighted mean failure probability over time, together with annual realized failure rates, reported each January. The figure shows that median failure probability has a clear tendency to increase in recessions and financial crises, when economic theory suggests the stochastic discount factor should be high. Furthermore, table 1 and figure 5 show that realized failure rates are higher during and immediately following bad times. Therefore, a company whose own failure probability tends to increase by a larger amount when the median failure probability increases has higher systematic risk.

Mean and median failure probability are clearly related to the realized failure rate and times of high failure probability are almost always also times of high realized failure rates. The one exception to this relationship is the spike in failure rates in 2001, after the end of the technology bull market of the late 1990s, which is not associated with a large increase in default probabilities. The reason is visible in figure 3: most of the sharp increase in failures were accounted for by B-grade issuers (junk), whose median default probability did increase in advance. However, these issuers were not close to the median issuer and did not account for a large proportion of total rated corporate issuance.

Our hypothesis is that failure beta is larger for lower-rated issuers, which means that the sensitivity to aggregate fluctuations in default probability increases in credit rating. To test this hypothesis we estimate failure beta by rating. We use 12-month annualized median failure probability as our measure of aggregate default risk conditions. We use the 12-month measure since it will not be focused excessively on short-term determinants of failure. In addition, the 12-month measure is viewed by the CDS market as an important determinant of within rating spread variation (table 5). Specifically, for firm i, with cumulative failure probability P_{it} defined as in equation (2), and with credit rating CR we estimate:

$$P_{it} = \alpha_{CR} + \beta_{CR} median(P_{it}) + \varepsilon_{it}. \tag{4}$$

This specification constrains all firms with the same rating to have the same failure beta, and the resulting estimate is the average firm failure beta, equal-weighted across all firm-months in the panel. Like stock market beta, failure beta must be estimated and such estimates are subject to error. Pooling the regression by rating, therefore, has the additional benefit of reducing measurement error of beta. In order to ensure a sufficient number of observations for each rating, we combine all observations rated CCC and below together. This specification does not mechanically require failure beta to be monotonically related to rating. However, as a robustness check, we also estimate failure betas firm-by-firm, sort into groups by failure beta, re-estimate the failure betas for each group, and compare with the mean rating for that group. Our results are not materially different if we use this alternative method, so we conclude that grouping the data by rating is not what drives our results.

Figure 6 plots estimates of failure beta by rating, together with 95% confidence intervals, using OLS standard errors, clustered by month. Failure beta is strongly related to credit rating: highly-rated issuers have low failure betas while low-rated issuers have high failure betas. In fact, failure beta is almost monotonic in rating. The relationship is strongly economically and statistically significant over the whole ratings spectrum and over the investment-grade spectrum alone. We also find that the relationship is still strongly significant when controlling for average default probabilities.

The result shown in figure 6 and associated tests establish that ratings are informative about the tendency of an issuer's propensity to default to worsen in bad times, controlling for the issuer's unconditional tendency to default. Ratings, therefore, are measuring at least two distinct forms of credit risk: default probability and systematic risk. The relative importance of systematic risk appears to be larger for investment grade bonds: a BBB bond is very little more likely to default in a given time span than a AA bond, and many BBB bonds are always less likely to default than many AA bonds (figure 2). However, the BBB bonds become much more likely to default in bad times than the AA bonds: mean default probabilities increase between good and bad times from 0.38 to 0.54 (an increase of 0.16) for AA and from 0.51 to 0.78 (increase of 0.27) for BBB (see table 4) and the respective failure betas are equal to 0.59 and 1.05. Since bad times are not easy to anticipate, BBB bonds are therefore much riskier than AA bonds to a diversified investor. By contrast, default probabilities are more strongly associated with rating for non-investment grade bonds.

4.2 Failure beta and CDS risk premia

We now examine whether or not variation in failure beta can explain variation in CDS risk premia. We use both default probability and failure beta to explain variation in CDS spreads.

A risky bond has an expected payoff equal to 1 unless it defaults. If it defaults with probability P_t, its expected payoff equals the expected recovery rate rec. If the expected payoff is discounted at a rate that allows for a risk premium RP and risk-free rate R_f, then today's price satisfies:

$$price = \frac{1 - P_t(1 - rec)}{1 + R_f + RP}. \tag{5}$$

Holding default probability P_t constant, a higher failure beta, which appears to be reflected in a lower rating, should result in a higher risk premium. Thus our results suggest that ratings do not exclusively relate to the numerator, the expected payoff, but are also important in identifying the denominator, the expected return. We now evaluate this idea more directly.

Spreads are defined, roughly speaking, as

$$price = \frac{1}{1 + R_f + spread}. \tag{6}$$

Therefore

$$RP = (1 + R_f + spread)(1 + P_t(1 - rec)) - (1 + R_f). \tag{7}$$

Using CDS spreads and recovery rates from Markit, together with our estimated failure probabilities, we use this relationship to calculate implied risk premia for all issuers in our CDS sample. Figure 7 plots median log risk premia for 5-year CDS contracts from 2001 to 2007 against log failure betas estimated as above by rating group. Each point is a particular rating in a particular year with the highest rated issuers (and lowest failure beta) a vertical group of points closest to the y-axis. (The row is vertical because we only estimate one failure beta per rating, but 7 different annual risk premia.) Because we have no explicit model of what determines CDS risk premia, we cannot specify what the precise form of the relationship should be between risk premia and failure beta. However, there is clearly a strong relationship in figure 7. To quantify the extent to which risk premia are related to failure beta we run a regression of risk premia on failure beta. We find a strongly statistically significant relationship that is steeper for investment grade issuers.

Table 8 reports these results. We regress median log risk premia for each year and each ratings group, calculated as described above from CDS spreads, our estimated default probabilities, and Markit average recovery rates, on log failure beta and log failure beta interacted with an investment grade dummy. Column one shows that log failure beta is statistically significantly related to CDS risk premia and explains 85% of variation in risk premia across the 111 ratings-year groups. Column two shows that this result is robust to the inclusion of year fixed effects. Column three shows that the relationship is significantly stronger for investment grade ratings.

The next three columns include log default probability as an additional control. We find that the relationship between risk premia and failure beta is robust to this control. Including default probability only, we find a significant effect (column (4)), but this seems to be driven by time variation in default probability: once we include year effects (column (5)), default probability enters with a counterintuitive negative sign. In the final column we find that, when we allow the effect of default probability to vary across investment grade groups, the effects of failure beta continue to be robust: failure beta and its interaction with he investment grade indicator are still strongly economically and statistically significant.

Our conclusion is that even our simple measure of systematic default risk is strongly related to rating and CDS risk premia. This relationship is robust to controlling for time effects and default probabilities. Moreover, default probabilities are not related to risk premia when controlling for failure beta. Therefore, ratings appear to measure, at least partly, the tendency for credit quality to deteriorate in bad times, as well as the raw tendency to default.

4.3 Failure beta and alternative measures of systematic risk

Obvious alternative measures of issuer systematic risk should also be related to rating, if rating measures systematic risk. However, of all these only failure beta is an explicit measure of the tendency of default probability to increase in bad times. We find that stock market betas of portfolios sorted by rating, although strongly related to rating overall, are much more weakly related for investment grade issuers than are failure betas. We conjecture that this is due to the greater ability of failure beta to identify off of bad times, such as severe crises and recessions.

Table 9 investigates our hypothesis with regard to investment grade issuers only. The left half of panel A presents evidence on how rating is related to four measures of systematic risk: failure beta, standard equity CAPM beta, and up and down beta. Up and down beta are simply estimates of CAPM beta that condition respectively on positive or negative equity index excess returns. If stock returns are non-linearly related to stock market returns, then up and down beta can be different and risk premia may be more strongly related to downside risk as measured by down beta (Ang, Chen and Xing (2006)). All of our estimates of beta are pooled by ratings group, as with our main estimate of failure beta. The right-hand half of panel A relates our estimates of CDS risk premia to these same betas.

If our conjecture about the greater ability of failure beta to identify sensitivity to bad times, we would expect down beta to be related more closely to rating than up beta. This is indeed what we find; in fact up beta is not statistically significantly related to rating. When relating up beta to CDS risk premium the coefficient has a counterintuitive negative sign. In contrast, down beta is strongly statistically and economically related to both rating and CDS risk premium. This result suggests that what rating is primarily measuring is sensitivity to bad times. Failure beta and CAPM beta are also related to rating and risk premium, but failure beta is more strongly related than down beta which is in turn more strongly related than CAPM beta. Failure beta explains 93% of variation in rating versus 86% for down beta and 68% for CAPM beta. The same pattern is reflected when relating measures of systematic risk to CDS risk premia: Failure beta explains 84% of variation in CDS risk premia versus 72% for down beta and 52% for CAPM beta beta. These results suggest that failure beta measures somewhat more extreme downside risk than down beta. Since a negative market return of ordinary magnitude should not increase the default probability of a typical investment grade issuer by very much, it is likely that failure beta measures exposure to a more drastic event than an average negative market return month and that this is the reason for its greater ability to explain variation in risk premia.

Failure beta does a good job at measuring the systematic risk of sensitivity to bad times and it is related to rating and risk premia. But what exactly determines failure beta? Empirically, variation in which determinant(s) of default probability is driving variation in failure beta? To answer this question we investigate which components of failure probability are most sensitive to bad times. We find that these variables are volatility, leverage, and past returns. We next investigate differences across rating in the sensitivity of these explanatory variables to bad times. Although volatility of return increases markedly in bad times, it does so fairly uniformly across firms, and therefore cannot account for cross-sectional variation in failure beta. By contrast the sensitivity of leverage and 12-month return to bad times exhibits wide variation across issuers grouped by rating. Panel B of table 9 shows that sensitivities of leverage and past 12 month return to median failure score explain about the same degree of cross-sectional variation in rating and CDS risk premia as does failure beta itself (leverage sensitivities explains slightly more). Companies are more likely to default if their leverage increases and they suffer a long series of low returns. If these indicators deteriorate by more in bad times, then these companies are more sensitive to bad times. Such companies are obviously more risky to a diversified investor, who should require a higher risk premium for holding them.

We conclude that failure beta measures the tendency of issuers' credit quality to deteriorate in bad times. It captures firms' tendency to have higher leverage and lower returns in bad times and is in particular more closely related to the risk of deteriorating returns in down markets. These are the risks that diversified investors care about and ratings seem to reflect this. For investment grade issuers, rating is strongly related to failure beta and, consequently, to CDS risk premia.

5 Conclusion

In this paper we investigate the information in corporate credit ratings relevant to investors concerned about credit risk, noting that credit risk can mean both individual firm probability of default at a given horizon and systematic default risk: the tendency to default in bad times. We argue the latter should be of more concern to a diversified investor and propose failure beta as a natural measure of the tendency of firm failure probabilities to covary across firms.

Using the same reduced-form model as in Campbell, Hilscher, and Szilagyi to model failure probability, we first investigate the ability of rating and failure probability to forecast corporate bankruptcy and failure. We assess the ability of this most recent state-of-the-art default prediction model to forecast failure and compare its ability to do so with credit ratings at different horizons.

We find that failure score, a linear combination of easily available accounting and market-based measures of financial distress, is better able to predict failure at short and medium horizons, up to 2 years (conditional on no earlier default having taken place), than S&P credit ratings. It is also better at forecasting the cumulative failure probability at horizons up to 5 years. However, ratings still provide useful incremental information for predicting cumulative default at all horizons up to five years. At one year horizons, adding rating to failure score increases the 1-year cumulative pseudo R^2 from 41% to 42% and at 5 years from 24% to 27%. Thus, publicly available information does not entirely drive out ratings a predictor of default, at least in our specification.

We also relate estimated reduced form default probabilities to credit ratings and establish a series of new facts. (1) There is substantial variation in default probability both within and across ratings. (2) For investment grade firms, there is an extremely high degree of overlap in the distribution of default probabilities across ratings. (3) Default probabilities are strongly economically and statistically related to credit spreads, controlling for rating, explaining an average of 31% of variation of CDS spreads within a given rating. (4) Default probabilities predict ratings downgrades well for all ratings, but better for lower-rated firms. (5) Default probabilities vary substantially across the business cycle in a way not captured by rating, consistent with rating through the cycle. (6) The default probabilities of firms with lower ratings rise by more during bad times, suggesting that those firms have higher systematic risk.

Taken together, these findings establish that an investor concerned about firm-level default at any horizon up to five years ahead should not rely only on corporate credit rating, but should also pay attention to information in share prices and accounts. However, they also show that such an investor should not ignore ratings altogether, while raising the intriguing possibility that ratings do not relate only to the probability of default.

We therefore proceed to relate credit rating to systematic default risk: the tendency to default in bad times. This measure of credit risk is of more importance to a diversified investor, who, other things equal, would prefer to buy a bond that is not more likely to default in bad times, when share prices are low, unemployment is high, and other bonds are more likely to default: Median default probability, and realized failure rates, show a strong tendency to rise in bad times, such as recessions and financial crises.

We propose an intuitive measure of systematic default risk, sensitivity to median default risk, or failure beta, and show that it is strongly related to rating and CDS risk premia even controlling for default probability. This finding is especially true for investment-grade issuers, whose default probabilities are much closer together: failure beta is more strongly related to CDS risk premia for investment grade issuers. This finding implies that ratings convey more information about systematic risk, and consequently less about default probability, for investment grade issuers.

Our conclusion is that ratings are a single summary measure that relate to two somewhat different aspects of credit risk: firm-level default probability and systematic default risk. This can explain a number of previously puzzling aspects of corporate credit ratings: First, why ratings are so easy to beat at predicting default using simple default predictors based on publicly available information; second, why agencies rate through the cycle and respond sluggishly to new information; third, why investors pay so much attention to ratings and why they draw such a distinction between investment-grade and non-investment-grade issuers and why ratings are strongly related to bond risk premia.

References

[1] Altman, Edward I., 1968, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", *Journal of Finance*, 23, 589-609.

[2] Amato, Jeffrey D., and Craig H. Furfine, 2004, "Are Credit Ratings Procyclical?" *Journal of Banking and Finance* 28, 11, 2641-2677.

[3] Ang, Andrew, Joseph Chen and Yuhang Xing, 2006, "Downside Risk," *Review of Financial Studies*, 2006 19(4), 1191-1239.

[4] Beaver, B., 1966, "Financial ratios as predictors of failure", *Journal of Accounting Research*, 4, 91-101.

[5] Beaver William H., McNichols M. and J.-W. Rhie, 2005, "Have financial statements become less informative?", *Review of Accounting Studies*, 10, 93-122.

[6] Berndt, Antje, Rohan Douglas, Darrell Duffie, Mark Ferguson, and David Schranz, 2008, "Measuring Default Risk Premia from Default Swap Rates and EDFs," unpublished paper, Tepper School of Business, Carnegie Mellon University and Graduate School of Business, Stanford University.

[7] Bhamra, Harjoat S., Lars-Alexander Kuehn and Ilya A. Strebulaev, 2007, "The levered equity risk premium and credit spreads: a unified framework", working paper, UBC.

[8] Blume, Marshall E. and Donald Keim, 1991, "Realized returns and volatily of low-grade bonds: 1977-1989", *Journal of Finance*, 46, 49-74.

[9] Blume, Marshall E., Felix Lim and A. Craig MacKinlay, 1998, "The declining credit quality of U.S. corporate debt: myth or reality?", *Journal of Finance*, vol. 53, 4, 1389-1413.

[10] Brennan, Michael J., Julia Hain and Ser-Huang Poon, 2007, "Tranching and rating," working paper, UCLA.

[11] Campbell John Y., Jens Hilscher and Jan Szilagyi, 2008, "In search of distress risk", *Journal of Finance*, Vol. 63, 6, 2899-2939.

[12] Cantor, Richard, and Christopher Mann, 2003, "Measuring the Performance of Corporate Bond Ratings" Moody's special comment (April).

[13] Chava, Sudheer, and Robert A. Jarrow, 2004, "Bankruptcy prediction with industry effects," *Review of Finance* 8, 537–569.

[14] Chen, Hui, 2008, "Macroeconomic conditions and the puzzles of credit spreads and cpital structure", working paper, MIT.

[15] Coval, Joshua, Jakub Jurek, and Erik Stafford, 2008, "Economic catastrophe bonds," *American Economic Review*, forthcoming.

[16] van Deventer, Donald R., Li Li, and Xiaoming Wang, 2005, "Advanced credit model performance testing to meet Basel requirements," *The Basel Handbook: A Guide for Financial Practitioners*, Michael K. Ong, editor, Risk Publications.

[17] Duffie, Darrell, Leandro Saita and Ke Wang, 2007, "Multi-period corporate default prediction with stochastic covariates", *Journal of Financial Economics*, 83, 635-665.

[18] Elton, Edwin J., Martin J. Gruber, Deepak Agarwal and Christopher Mann, 2001, "Explaining the rate spread on corporate bonds", *Journal of Finance*, Vol. 56, 1, 247-277.

[19] Fons, Jerome, 1991, "An approach to forecasting default rates", Moody's special paper.

[20] Gilson, Stuart C., 1997, "Transaction costs and capital structure choice: Evidence from financially distressed firms", *Journal of Finance*, 61, 921-956.

[21] Gilson, Stuart C., Kose John and Larry Lang, 1990, "Troubled debt restructurings: An empirical study of private reorganisations of firms in default", *Journal of Financial Economics*, 27, 315-353.

[22] Hickman, W. Braddock, 1958, *Corporate Bond Quality and Investor Experience*, Princeton University Press, Princeton New Jersey.

[23] Hillegeist S. A., E. K. Keating, D. P. Cram and K. G. Lundstedt, 2004, "Assessing the probability of bankruptcy", *Review of Accounting Studies*, 9, 5-34.

[24] Jonsson J. and M. Fridson, 1996, "Forecasting default rates on high-yield bonds", *Journal of Fixed Income*, June, 69-77.

[25] Löffler, Gunter, 2004, "An Anatomy of Rating through the Cycle," *Journal of Banking and Finance*, 28, 3, 695-720.

[26] McDonald C. G. and L. M. Van de Gucht, 1999, High-yield bond default and call risks, *Review of Economics and Statistics*, 81, 409-419.

[27] Ohlson, James A., 1980, "Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research*, 18, 109-131.

[28] Shumway, Tyler, 2001, "Forecasting Bankruptcy more Accurately: A Simple Hazard Model," *Journal of Business* 74, 101–124.

[29] Vassalou, Maria and Yuhang Xing, 2004, "Default risk in equity returns", *Journal of Finance*, vol. 59, 2, 831-868.

Table 1: Failures over time - all firms and rated firms

Panel A lists the number of firms and failures for all active firms (1971 to 2008) and for all firms with a S&P credit rating (1986 to 2008). Failure is defined as the first of bankruptcy (chapter 7, chapter 11), de-listing for performance related reason, default (D) or selective default (SD) rating, and government-led bailout. The number of firms is the average number of firms in a given year or over a given range of years. Firms are included as active if they have either available accounting or equity market data. Share (column (7)) is the share of rated firms' total liabilities (book value) in overall total universe. Panel B reports failure rates during and 12 months after NBER recessions and financial crises (October 1987, September 1998).

Panel A: Failures over time

	all firms			rated firms			
year	**firms**	**failures**	**rate (%)**	**firms**	**failures**	**rate (%)**	**share**
1971-1980	3673	121	0.32	.	.	.	
1981-1985	5172	259	0.97	.	.	.	
1986	5896	101	1.71	907	20	2.20	0.68
1987	6331	58	0.92	999	9	0.90	0.70
1988	6445	94	1.46	974	30	3.08	0.72
1989	6346	77	1.21	932	11	1.18	0.72
1990	6269	80	1.28	889	8	0.90	0.70
1991	6291	72	1.14	873	15	1.72	0.68
1992	6622	49	0.74	913	5	0.55	0.68
1993	7149	40	0.56	988	7	0.71	0.69
1994	7793	33	0.42	1082	4	0.37	0.68
1995	8099	43	0.53	1156	5	0.43	0.69
1996	8474	34	0.40	1288	5	0.39	0.73
1997	9273	61	0.66	1482	9	0.61	0.74
1998	9572	148	1.55	1647	10	0.61	0.73
1999	9270	207	2.23	1741	39	2.24	0.78
2000	9018	167	1.85	1719	34	1.98	0.81
2001	8379	333	3.97	1691	67	3.96	0.82
2002	7757	229	2.95	1646	50	3.04	0.84
2003	7334	165	2.25	1597	32	2.00	0.85
2004	6777	38	0.56	1633	14	0.86	0.87
2005	6781	36	0.53	1631	16	0.98	0.86
2006	6786	18	0.27	1613	6	0.37	0.85
2007	6919	24	0.35	1544	6	0.39	0.83
2008	6896	50	0.73	1454	27	1.86	0.85

Panel B: Failures during and after recessions and crises

	all firms			rated firms		
	months	**failures**	**rate (%)**	**months**	**failures**	**rate (%)**
normal	287	1167	0.80	183	175	0.89
recession	143	1030	1.49	67	187	2.41
crisis	26	340	1.98	26	67	2.27

Table 2: Failure prediction in a logit model

In panel A we report results from logit regressions of the failure indicator for all active firms including unrated firms on a set of monthly explanatory variables defined as follows: weighted average of net income over market value of total assets over the previous 12 months (NIMTAAVG), total liabilities over market value of total assets (TLMTA), weighted average of annualized log of gross excess return over value weighted S&P 500 return over previous 12 months (EXRETAVG), log of firm's market equity over the total valuation of S&P 500 (RSIZE), square root of a sum of squared firm stock returns over a three-month period, annualized (SIGMA), stock of cash and short term investments over the market value of total assets (CASHMTA), market-to-book value of the firm (MB) and log of price per share winsorized above $15 (PRICE). Market value of total assets is the sum of market value of firm equity and its total liabilities. Z-statistics are reported in parentheses. * denotes significant at 5%, ** denotes significant at 1%.

Panel A: Logit model (1963 to 2008)		
	(1)	(2)
Lag (months)	0	12
NIMTAAVG	-28.92	-21.67
	(22.74)**	(19.60)**
TLMTA	3.436	1.65
	(29.36)**	(19.60)**
EXRETAVG	-7.791	-8.50
	(15.00)**	(17.35)**
SIGMA	2.09	1.69
	(19.42)**	(20.53)**
RSIZE	0.064	-0.058
	(2.32)*	(3.07)**
CASHMTA	-2.53	-2.35
	(9.64)**	(9.94)**
MB	0.048	0.074
	(4.48)**	(6.48)**
PRICE	-0.416	0.096
	(15.57)**	(3.41)**
Constant	-10.41	-9.93
	(30.28)**	(40.53)**
Observations	2,022,562	1,870,481
Failures	1756	2159

Panel B: Forecast accuracy		
In-sample (1963 to 2008)		
Pseudo R sq	0.315	0.118
Accuracy ratio	0.955	0.860

Table 3: Failure score vs. S&P credit ratings (1986 to 2008)

This table reports results from monthly logit regressions of our failure indicator on failure score (computed using data for the full sample and corresponds to the methodology used in Table 2) and S&P long-term credit rating (AAA=1, AA+=2, …, CCC-=19, CC=20, C=21). We estimate logit specifications using data lagged 1, 3, 6, 12, 24, 36, 48, and 60 months. The sample period is from 1986 to 2008 and contains rated companies only, while failure score is estimated using the full sample of all active firms. Panel A reports McFadden's pseudo R-squared and accuracy ratio as measures of model performance. Panel B reports coefficients on failure score and rating when both are included. Z-statistics are reported in parentheses. * denotes significant at 5%, ** denotes significant at 1%. Panel C reports performance measures when forecasting failure events every January over the next 1, 2, 3, 4, and 5 years using 12-month failure score (from Table 2).

Panel A: Accuracy of failure prediction

prediction month	1	3	6	12	24	36	48	60
Pseudo R-squared								
failure score only	39.9%	34.7%	28.1%	20.1%	9.2%	6.9%	5.0%	3.2%
credit rating only	29.2%	20.6%	16.0%	12.0%	8.4%	7.2%	6.4%	5.4%
both	42.5%	35.9%	29.3%	21.4%	11.3%	8.8%	7.2%	5.6%
Accuracy Ratio								
failure score only	97.0%	96.1%	94.5%	91.8%	83.3%	79.6%	76.8%	72.1%
credit rating only	93.2%	90.3%	88.0%	84.7%	80.8%	79.1%	78.2%	76.2%
both	97.5%	96.5%	95.1%	92.2%	84.9%	82.0%	80.1%	76.7%

Panel B: Coefficients on failure predictors in the 'both' specification

prediction month	1	3	6	12	24	36	48	60
failure score	0.852	0.970	0.994	1.033	0.820	0.798	0.666	0.410
	(24.13)**	(28.57)**	(28.71)**	(25.36)**	(13.31)**	(8.97)**	(5.72)**	(2.43)*
credit rating	0.311	0.186	0.171	0.170	0.207	0.200	0.210	0.219
	(11.08)**	(8.10)**	(8.39)**	(8.91)**	(10.45)**	(8.99)**	(8.77)**	(8.22)**
# of observations	341,287	336,994	329,946	313,911	278,381	245,150	215,796	189,557
# of failures	340	401	429	437	370	282	231	186

Panel C: Cumulative prediction accuracy

years	1	2	3	4	5
Pseudo R-squared					
failure score only	40.9%	33.3%	28.3%	25.8%	23.6%
credit rating only	24.1%	21.4%	19.6%	18.7%	18.0%
both	42.3%	35.5%	31.0%	28.8%	26.9%

Table 4: Fitted failure probability by credit ratings

This table reports number of observations and annualized 12-month default probability for the full sample, in good times and in bad times, as well as the full sample standard deviation. Bad times are defined as NBER recessions, the October 1987 and September 1998 financial crises, and include 12 months after these events. The definition is the same as that used in table 1. Ratings are grouped by letter rating.

rating	# obs	share (%)	PD	PD (good)	PD (bad)	stdev(PD)
AAA	5,011	1.5	0.36	0.32	0.43	0.46
AA	23,100	6.8	0.43	0.38	0.54	0.65
A	79,454	23.3	0.49	0.44	0.61	0.54
BBB	97,925	28.7	0.60	0.51	0.78	0.73
BB	75,481	22.1	1.05	0.80	1.52	1.85
B	54,810	16.1	2.40	1.64	3.72	4.27
CCC, below	5,592	1.6	8.18	6.32	10.61	10.34

Table 5: Explaining CDS spreads with failure probability and credit rating

This table reports results from regressions of log(CDS spread) on log(12-month failure probability) and S&P credit rating. CDS spread data is for the 5-year CDS contract; data is from Markit. * denotes significant at 5%, ** denotes significant at 1%. Spread data are from Markit and are end of month 5 year CDS spreads (modified restructuring). The sample period is 2001 to 2007. All firms with CDS spread, ratings and available default probability in the CRSP-COMPUSTAT universe are included. Standard errors are robust and are clustered by year.

	AAA	AA	A	BBB	BB	B	CCC,CC	All
log(failure probability)	0.86	0.78	0.52	0.67	0.54	0.51	0.24	0.58
	(7.71)**	(13.95)**	(5.56)**	(6.23)**	(9.04)**	(11.27)**	(1.74)	(7.58)**
Credit rating		-0.03	0.20	0.30	0.20	0.14	0.19	0.23
		(0.48)	(11.82)**	(14.89)**	(5.15)**	(4.74)**	(1.27)	(13.58)**
Constant	9.89	9.52	6.54	6.93	7.15	7.58	4.95	6.86
	(10.32)**	(22.19)**	(7.65)**	(6.47)**	(7.91)**	(16.17)**	(1.59)	(8.12)**
R-squared								
log(failure probability) only	34.9%	34.1%	18.7%	30.3%	42.6%	46.5%	11.1%	37.9%
rating only		0.01%	5.8%	12.7%	10.6%	9.1%	7.6%	62.8%
both		34.2%	24.7%	40.7%	47.9%	49.1%	16.3%	73.9%
Number of firms	11	47	253	420	249	140	27	797
Number of observations: 38,569								

Table 6: Explaining variation in CDS spreads

This table reports results from regressions of log(5 year CDS spread) on log(failure probability) and a set of dummy variables: S&P credit rating, year, and firm. Standard errors are clustered by year and are reported in parentheses. ** denotes significant at 1% level. The sample corresponds to the sample in Table 5.

	(1)	(3)	(5)	(6)	(8)	(9)
R-squared (overall)	0.645	0.754	0.777	0.828	0.870	0.899
R-squared (within)		0.307				
R-squared (between)		0.802				
log(12 month failure probability)		0.590		0.427		0.454
		(7.48)**		(10.80)**		(12.86)**
Rating fixed effects	X	X	X	X	X	X
Year fixed effects			X	X	X	X
Firm fixed effects					X	X

Number of observations: 38,569

Table 7: Failure score and downgrades

This table reports results from logit regressions of downgrades (an increas in S&P credit rating, e.g. from AA to AA-) on 12-month failure score estimated using the model in Table 2. ** denotes significant at 1% level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rating	AAA	AA	A	BBB	BB	B	CCC
Failure score (12 months)	0.655	1.051	0.912	1.175	0.965	0.991	0.933
	(3.02)**	(14.38)**	(21.07)**	(36.71)**	(38.24)**	(40.86)**	(16.68)**
Constant	0.284	4.068	2.804	4.503	2.728	2.337	1.826
	(0.16)	(7.18)**	(8.45)**	(19.37)**	(16.15)**	(16.96)**	(7.12)**
Pseudo R-squared	1.8%	5.1%	3.4%	8.5%	10.3%	15.7%	14.9%
Accuracy Ratio	63.9%	67.9%	64.0%	72.6%	75.5%	82.1%	80.3%
# of observations	5011	23059	79299	97705	75059	54676	5687
# of downgrades	31	319	1035	1221	1218	1145	308

Table 8: Regression of risk premia on failuire beta

This table reports results from regressions of log risk premium (calculated using 1-year cumulative default probability using our logit model and recovery rates from Markit) on failure beta and log default probability. ** denotes significant at 1% level. Standard errors are robust and clustered by year. We include median annual observations for 16 rating groups.

	(1)	(2)	(3)	(4)	(5)	(6)
log(failure beta)	1.834	1.844	1.571	1.468	2.109	1.132
	(25.59)**	(26.04)**	(22.64)**	(10.51)**	(16.49)**	(7.07)**
log(failure beta)*inv. grade			0.948			1.175
			(5.22)**			(8.31)**
log(default probability)				0.160	-0.115	-0.006
				(4.18)**	(2.50)*	(0.10)
log(def. prob.)*inv. grade						0.050
						(5.65)**
Year fixed effects		X	X		X	X
R-squared (overall)	85.0%	93.7%	95.0%	86.8%	94.2%	95.8%
R-squared (within)		93.1%	94.6%		93.7%	95.4%
Number of observations: 111						

Table 9: Determinants of failure beta

This table reports results from regressions of different measure of systematic risk (failure beta, CAPM equity beta, up beta, down beta) on rating as well as regressions of CDS risk premia on these measures of systematic risk. Results are for investment grade issuers (rated AAA to BBB-). For risk premia calculations we include median annual risk premia by rating (subsample of the sample used in Table 8), and include year fixed effects in the regression. Panel A reports results using failure beta, CAPM equity beta (estimated using equally-weighted equity portfolio returns), up beta, and down beta, estimated by allowing the coefficient on the market factor RM (from Ken French's website) to vary for the factor being below and above 0. Panel B reports results using senstivities of leverage and average returns to median failure score.

Panel A: Different measures of systematic risk				
	Regression of beta on rating		**Regression of risk premia on beta**	
	Coefficient	R-squared	Coefficient	R-squared
failure beta (log)	0.051	92.7%	2.362	83.8%
	(10.07)**		(16.88)**	
CAPM beta	0.023	67.9%	3.567	52.4%
	(4.11)**		(7.78)**	
Up beta	-0.009	17.3%	-2.193	7.6%
	(1.29)		(2.12)*	
Down beta	0.046	86.3%	2.609	71.8%
	(7.08)**		(11.84)**	
Panel B: Sensitivities of explanatory variables to failure score				
TLMTA beta	0.027	93.5%	4.697	90.3%
	(10.69)**		(22.57)**	
EXRETAVG beta	0.026	92.6%	4.662	82.9%
	(10.00)**		(16.33)**	

Figure 1: Predicting corporate failure using failure score and S&P credit rating

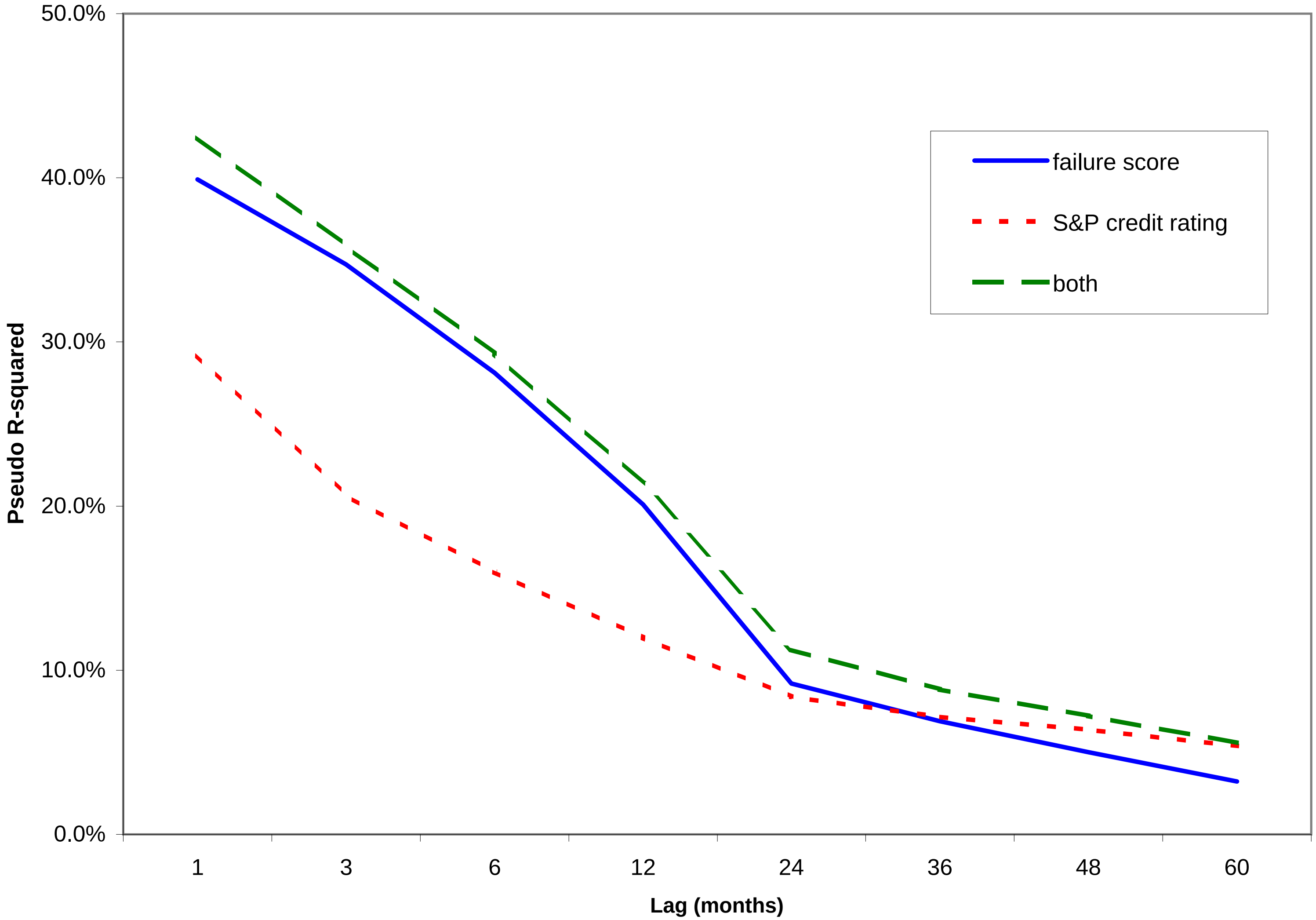
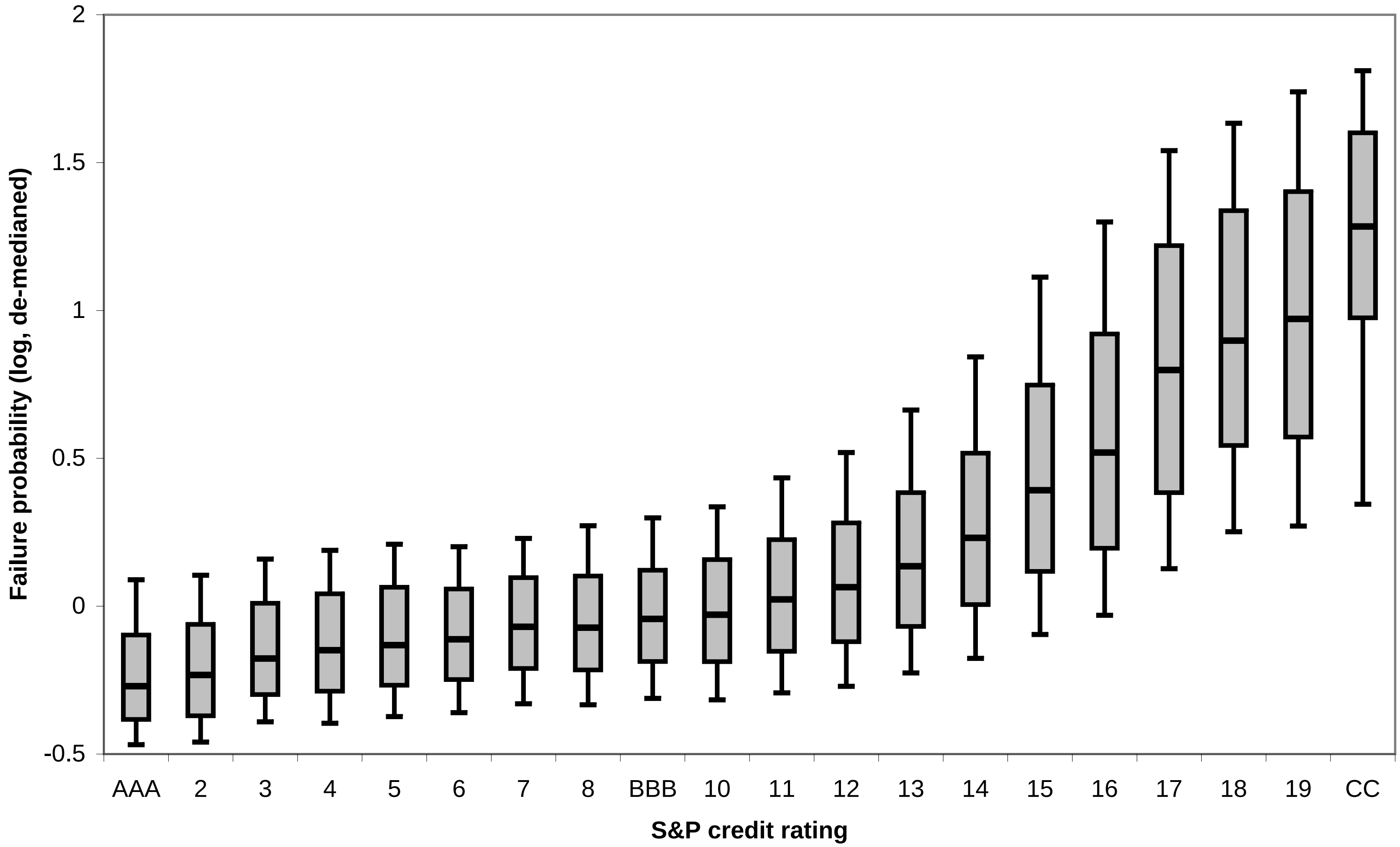


Figure 2: Failure probability distribution by S&P credit rating



This figure plots the percentiles of the de-medianed 12-month failure probability, annualized (10th, 25th, median, 75th, 90th of the distribution), by S&P credit rating (AAA=1, AA+=2, …, CCC-=19, CC=20). The failure probability is the annualized 12-month failure probability (from Table 2) and is de-medianed using the overall yearly median failure probability. The data included is from 1985 to 2008.

Figure 3: Median failure probability by rating

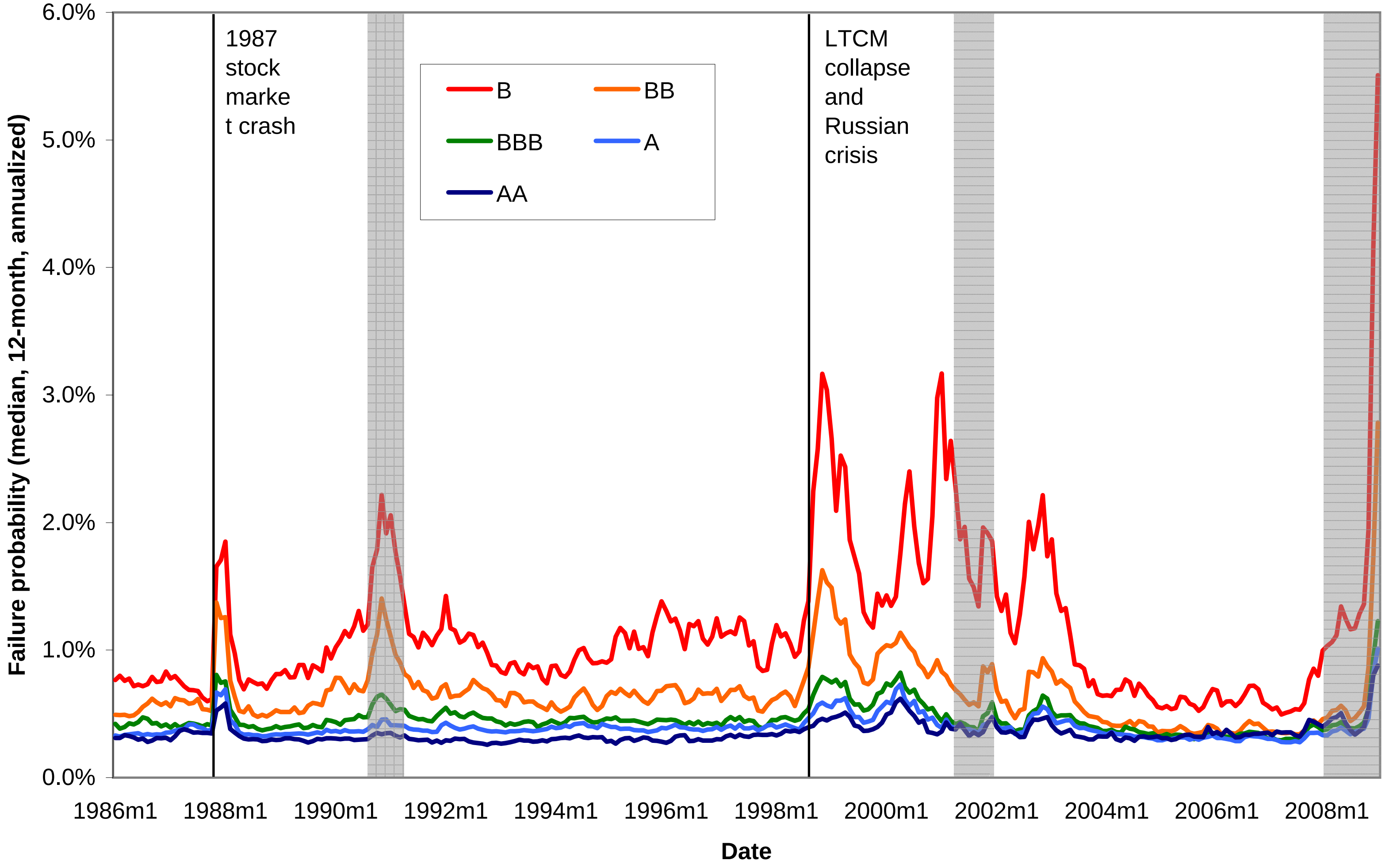


Figure 4: Share in different ratings

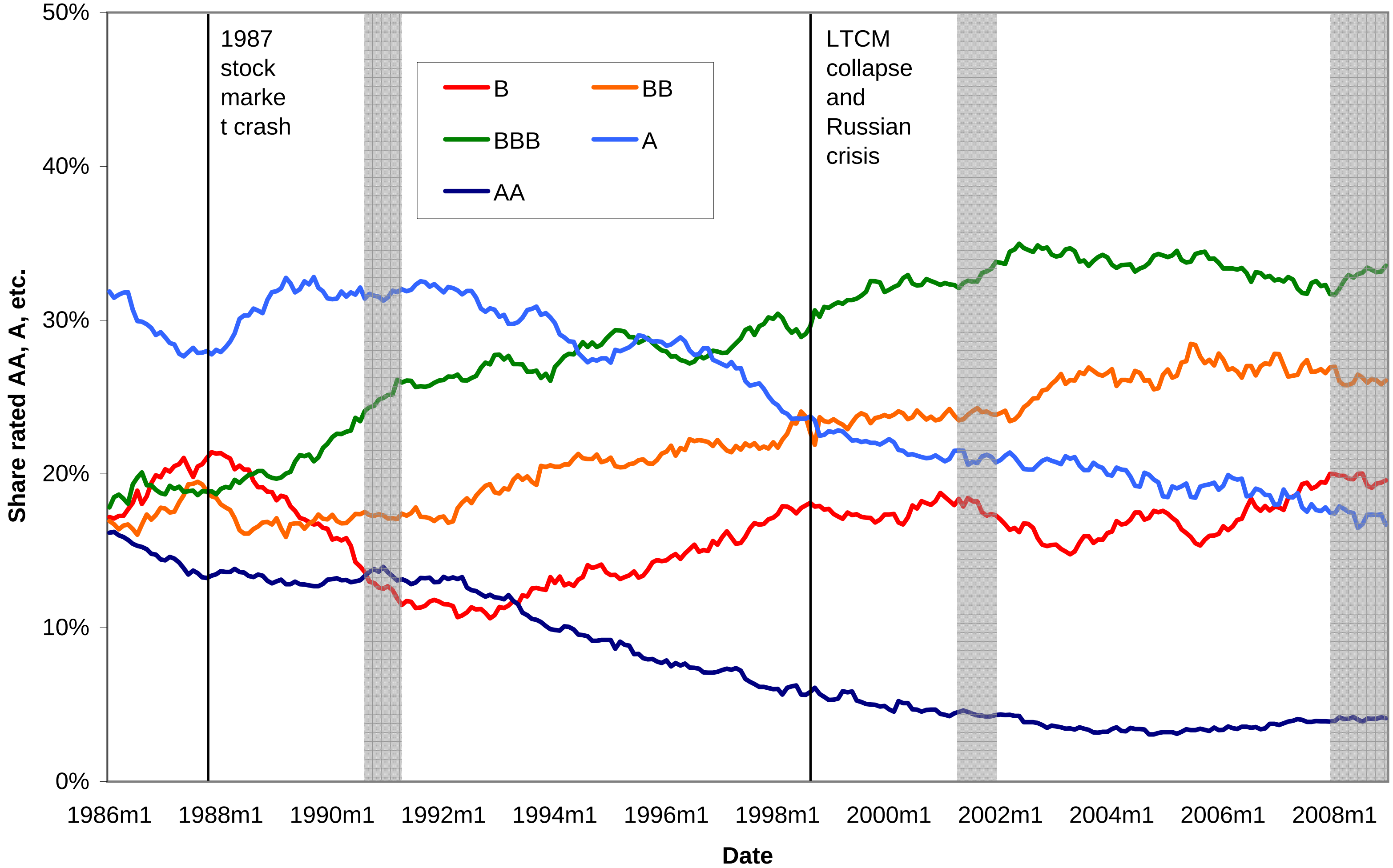
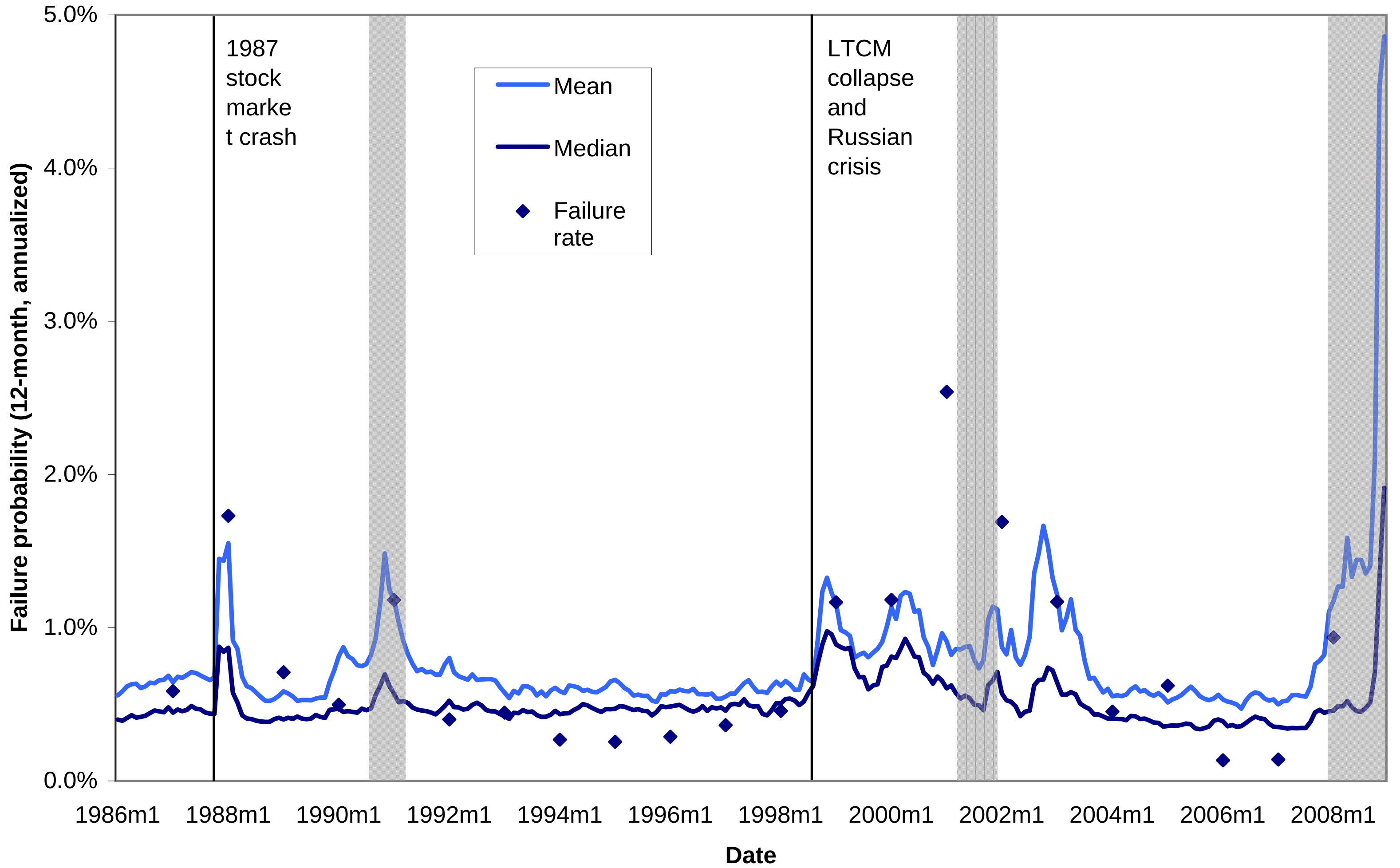
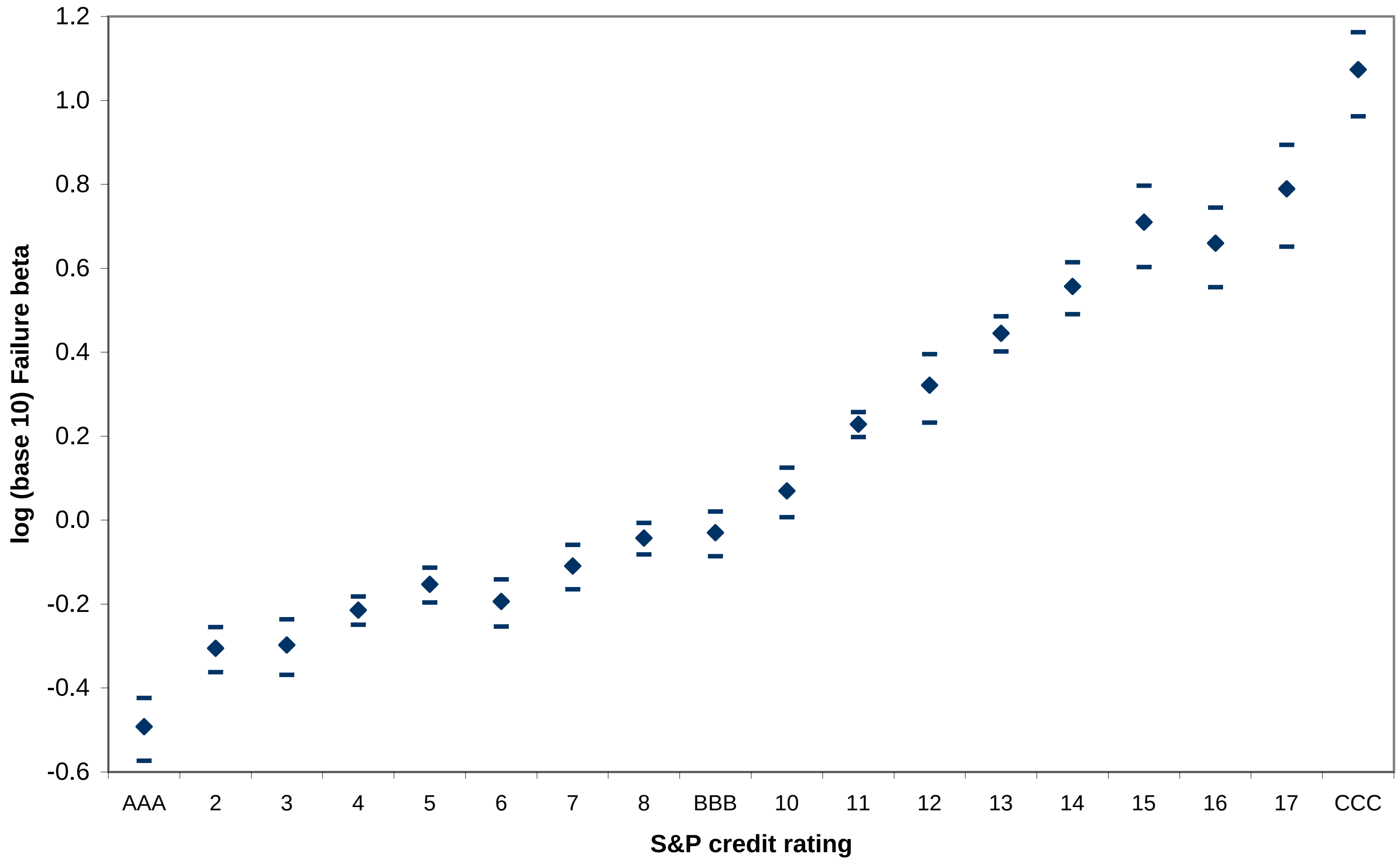


Figure 5: Average failure probability over time



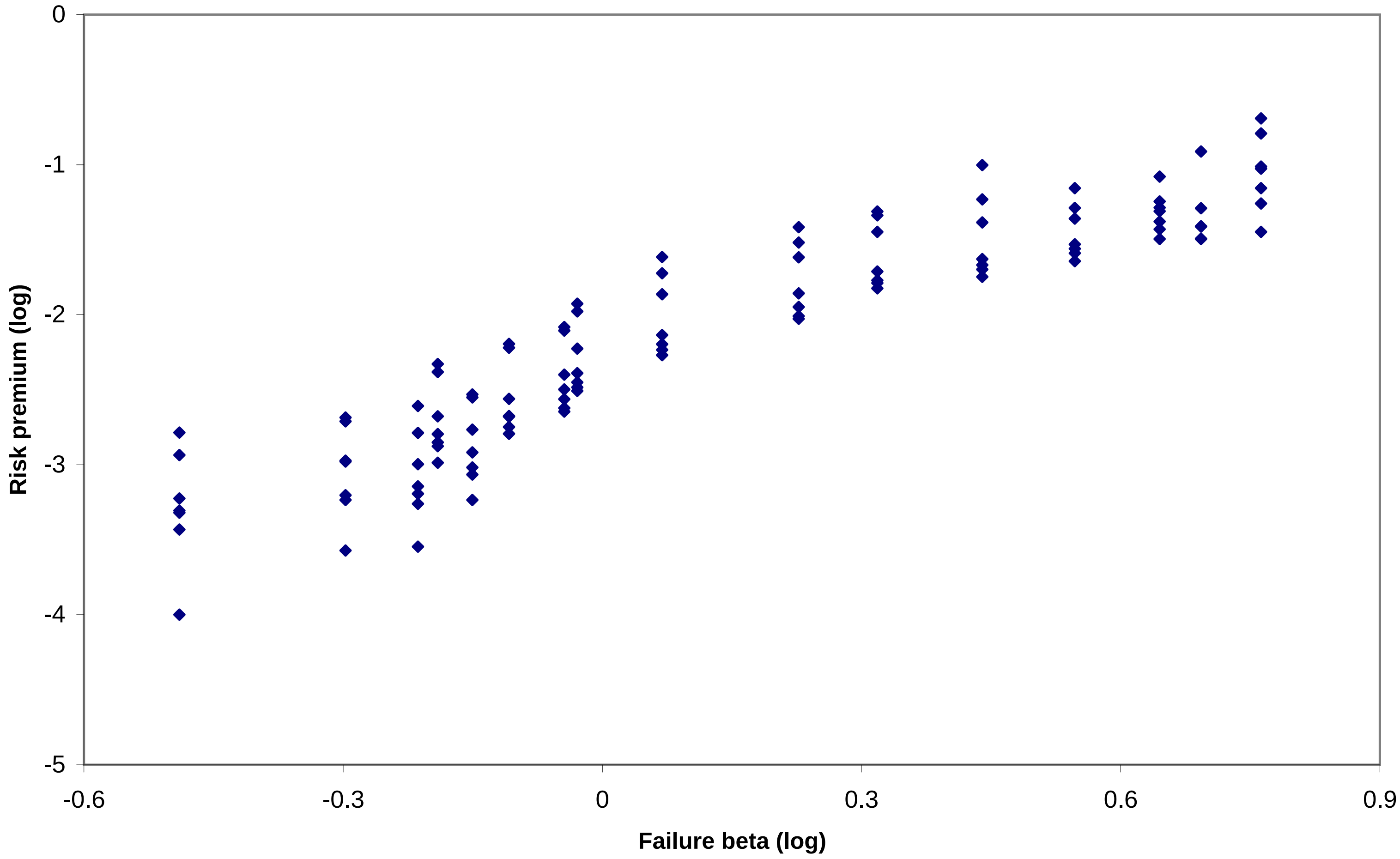
This figure plots the median and total liability-weighted mean 12-month failure probability from 1985 to 2008. It also plots the annual empirical failure rate, which is normalized to the mean predicted failure probability for ease of comparison. The annual failure rate is reported for the year is reported in January.

Figure 6: Failure beta by S&P credit rating



This figure plots the failure beta (12-month failure sensitivities) by S&P credit rating. The failure beta is measured as the coefficient from a regression of failure probability on median failure probability, including firm fixed effects. 95% confidence intervals of the coefficient estimates are included, standard errors are robust and clustered by month.

Figure 7: Risk premium and failure beta



This figure plots the failure beta (12-month failure sensitivities), calculated by S&P credit rating, and median risk CDS premium for each year from 2001 to 2007. The risk premium is calculated using the 1-year cumulative default probability, the 1-year USD swap rates as the risk free rate, and Markit recovery values.