From: MikeL [mailto:mike@advantagedsolutions.com]
Sent: Monday, August 30, 2010 2:53 PM
To: Billingsley, Ryan; Bean, Bobby R.
Cc: 'MikeL'
Subject: Alternatives to Credit Ratings

Ryan,

Thanks for taking my call this afternoon. As I mentioned the Kamakura Corporation has recently released a new paper discussing the use of default probability models as an alternative to credit ratings. I've attached that paper for you review. As I, also, mentioned, ASI and Kamakura have had an ongoing discussion with FDIC about Kamakura Corporation's Risk Management and Default Probability tools and services.

I've taken the liberty of copying Bobby Bean as well.

Sean Klein and David Boldon will be in the DC area the last week of September and the first week of October and could be made available for a detailed discussion if you have an interest. Please let me know.

Rgards,

Mikel
301-875-8493

Michael J. Leishman

Advantaged Solutions, Inc.
1300 Pennsylvania Avenue, NW
Suite 700
Washington, D.C. 20004

mike@advantagedsolutions.com

Credit ratings and credit risk

Jens Hilscher*
hilscher@brandeis.edu

Mungo Wilson†
mungo.wilson@sbs.ox.ac.uk

This version: August 2010

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

†Saïd Business School, Oxford University, Park End Street, Oxford OX1 1HP, UK. Phone +44-1865-288914. Wilson acknowledges the help of a Hong Kong RGC grant (project no. HKUST6478/06H).

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 Effi Benmelech, Max Bruche, John Campbell, Steve Cecchetti, Robert Jarrow, Blake LeBaron, Tarun Ramadorai, David Scharfstein, Andrei Shleifer, Monica Singhal, Jeremy Stein, Jan Szilagyi, Adrien Verdelhan, David Webb, Robert Whitelaw, Moto Yogo, and seminar participants at Royal Holloway, University of Zürich, LSE, Humboldt Universität zu Berlin, CEMFI, Brandeis University, the 2009 Venice C.R.E.D.I.T. conference, the Oxford-Man Institute for Quantitative Finance, the Federal Reserve Bank of Boston, Leicester University, and the 20th FDIC Derivatives Securities and Risk Management conference 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.

Abstract

This paper investigates the information in corporate credit ratings. We examine the extent to which firms' credit ratings measure raw probability of default as opposed to systematic risk of default, a firm's tendency to default in bad times. We find that credit ratings are dominated as predictors of corporate failure by a simple model based on publicly available financial information ('failure score'), indicating that ratings are poor measures of raw default probability. However, ratings are strongly related to a straightforward measure of systematic default risk: the sensitivity of firm default probability to its common component ('failure beta'). Furthermore, this systematic risk measure is strongly related to credit default swap risk premia. Our findings can explain otherwise puzzling qualities of ratings.

JEL Classification: G12, G24, G33

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

1 Introduction

Despite recent criticism, credit ratings remain the most common and widely used measure of corporate credit quality. Investors use credit ratings to make portfolio allocation decisions; in particular pension funds, banks, and insurance companies use credit ratings as investment screens and to allocate regulatory capital. Central banks use credit ratings as proxies for the quality of collateral. Corporate executives evaluate corporate policies partly on the basis of how their credit rating may be affected. Recent events and associated debate more than ever underline the importance of understanding if ratings are appropriate for these purposes. Increased regulatory pressure and discussion have focused on what the specific role of credit ratings and credit rating agencies should be and whether or not there are suitable alternatives to credit ratings.

Before we can assess the suitability of credit ratings or embark on a search for alternatives, it is important first to understand what credit ratings measure. Conventionally, credit ratings are thought to provide information about the likelihood of default and other forms of corporate failure.[1] However, a diversified investor cares not only about a firm's raw default probability but also its systematic risk, its tendency to default in bad times. This distinction can also be expressed in terms of the expected payoff on a corporate bond, which depends on its raw default probability, versus its discount rate or risk premium, which depends on its systematic default risk. In this paper we ask how good corporate credit ratings are at capturing these two different aspects of credit risk.

We approach this question in two steps. First, we examine the ability of credit ratings to forecast corporate failure – the extent to which ratings measure raw default probability. Second, we evaluate credit ratings' ability to capture systematic risk – the risk to a diversified investor of holding firms' bonds. We find that while credit ratings are dominated as a predictor of corporate failure by a simple model based on publicly available information, ratings are strongly related to a straightforward measure of systematic default risk, and this systematic risk measure is itself strongly related to credit default swap (CDS) risk premia. We now

[1] See, for example, West (1970), Blume, Lim, and MacKinlay (1998), Krahnen and Weber (2001), Löffler (2004b), Molina (2005), and Avramov, Chordia, Jostova, and Philippov (2009).

describe our procedure, and findings, for each of these two steps in more detail.

We begin by investigating the ability of credit ratings to forecast corporate default and failure. Following Campbell, Hilscher, and Szilagyi (2008) we define failure as the first of the following events: 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.

We evaluate the ability of credit ratings to predict failure and compare ratings' forecast accuracy to a benchmark measure that is constructed in a dynamic logit model. We build on recent models of default prediction (Shumway (2001), Chava and Jarrow (2004), and Campbell et al.)[2] by constructing a straightforward predictor of default based on accounting data and stock market prices.

We find that this measure, which we refer to as 'failure score,' is substantially more accurate than ratings at predicting failure over the next 1 to 5 years. The higher accuracy in predicting the cumulative failure probability is driven by a much higher ability of failure score at predicting marginal default probabilities at horizons of up to 2 years and the fact that credit rating adds little information to marginal default prediction at horizons up to 5 years. We show that the superior performance of failure score at predicting default is not driven by the fact that ratings update only infrequently, nor because ratings use a discrete, "broad-brush" ranking.

We next investigate in more depth how credit ratings relate to default probabilities and provide additional evidence that ratings are not primarily a measure of raw default probability. We begin by presenting further motivation for using fitted failure probability as a benchmark predictor of default: failure probability explains variation in CDS spreads within identically rated firms (i.e. the market views within-rating variation in failure probabilities as important);

[2] 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).

in addition, failure probability is a significant predictor of a deterioration in credit quality as measured by rating downgrades. Using fitted values as a measure of default probability, we then relate ratings directly to default probabilities. Contrary to the interpretation that credit rating reflects raw default probability there is considerable overlap of default probability distributions across investment grade ratings; many firms with investment grade ratings have the same or very similar default probabilities even though their ratings are quite different. This means that variation in rating explains only very little variation in raw default probability. Furthermore, there is important time-variation in failure probabilities not captured by ratings.

Our results in the first part of the paper suggest that if ratings are understood primarily as predictors of default, then they are puzzling for a number of reasons. First, they are easily improved upon using publicly available data. Second, they fail to differentiate between firms: firms with the same rating often have widely different default probabilities and firms with very different ratings often have very similar default probabilities. Third, they fail to capture variation in default probability over time.

In the second part of the paper, we investigate if instead credit ratings capture systematic default risk. We begin by identifying a measure of systematic risk. We assume a single factor structure for default probability and measure a firm's systematic risk as the sensitivity of its default probability to the common factor. We refer to this measure as 'failure beta.' To identify a measure of the common factor we extract the first principal component of default probability.[3] We find that median default probability is highly correlated with the first principal component and therefore use median default probability as our measure of the common factor.

For risk averse investors to be concerned about failure beta it must be the case that a bond's failure beta affects the non-diversifiable component of its risk. It is straightforward to show that failure betas are monotonically related to joint default probability for any pair of firms, so that higher failure beta is equivalent to higher non-diversifiable default risk. Furthermore, times of high default probabilities (high levels of the common factor) are bad times: the realized default rate varies countercyclically, being much higher during and immediately after recessions and

[3] The first principal component of default probability captures 70.3% of the variation in default probability across ratings and 41.7% across industries.

financial crises (e.g. Campbell et al. (2008), Duffie et al. (2009)).[4] A default during such times is of more concern to a risk-averse investor than a default during good times. Risk averse investors will therefore demand a higher risk premium as compensation for higher exposure to bad times.

We find that credit rating strongly reflects variation in systematic risk and that exposure to bad times is compensated by higher CDS risk premia. We estimate failure betas for each rating and find that failure beta is strongly related to rating; there is a virtually monotonic relationship between rating and failure beta and failure beta explains 95% of variation in rating. The default probability of a typical firm (the common component of default probability) increases strongly in recessions and financial crises and this increase is more pronounced for lower-rated (higher failure beta) firms. During and after recessions and financial crises ('bad times') failure probabilities are 40% (investment grade) and 98% (non-investment grade) higher than outside of these times. Investors demand compensation for the exposure to this risk – we find that variation in failure beta explains 91% of the variation in CDS risk premia across ratings.

The relationship between credit rating (and CDS risk premia) and systematic risk is robust to using more conventional measures of systematic risk such as CAPM beta, though the relationship is not as strong as it is with failure beta. Credit ratings are also closely related to down beta – the sensitivity of stock returns to negative market returns. It seems, therefore, that credit ratings are measuring exposure to bad times, something failure beta and down beta capture well and something corporate bond investors are particularly concerned about.

In summary 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. Interestingly, rating agencies themselves appear to be aware of this dual objective: Standard & Poor's website states that a AA rating means that

[4]The recent recession is no exception: An important consequence of the recent financial crisis and recession has been the ongoing wave of major corporate failures and near-failures. In the first eight months of 2009 216 corporate issuers defaulted affecting $523 billion of debt (September 2009 S&P report). High default rates in recessions may be the result of low fundamentals during these times (Campbell et al. (2008)), they may be driven by credit cycles (Sharpe (1994), Kiyotaki and Moore (1997), Geanakoplos (2009)), or by unobservable factors (Duffie et al. (2009)).

a bond is, in the agency's opinion, "less likely to default than the BBB bond."[5] On the same web-page, the agency states that a speculative-grade rating "factors in greater vulnerability to down business cycles."

Our results can explain a number of otherwise puzzling aspects of ratings: (1) why ratings are not as good as a simple alternative at forecasting default: to do so may not be their sole purpose; (2) why ratings do not distinguish well between firms with different default probabilities: default probability and systematic default risk are economically different attributes; (3) why agencies 'rate through the cycle': if systematic risk equals "vulnerability to down business cycles," (the measurement of which is a stated objective) it cannot vary over the business cycle, so neither can rating to the extent rating reflects systematic risk; (4) why investors are interested in ratings and why variation in borrowing cost is strongly related to rating: investors care both about expected payoff and about risk premia.

This paper adds to a large early literature that evaluates the ability of ratings to predict default, beginning with Hickman (1958). More recently, van Deventer, Li, and Wang (2005) evaluate Basel II implementations and compare accuracy ratios of S&P credit ratings to a reduced form measure of default probability. 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. We discuss this literature in more detail in the next section. Our paper adds to this line of work since we compare credit ratings and the most recent reduced form models of default prediction and measure differences in default probability within rating and over time.

Our findings are also related to several studies that investigate the determinants of corporate bond prices. The idea that both default probabilities and risk premia affect bond prices and CDS spreads is well understood (see e.g. Elton, Gruber, Agarwal, and Mann (2001)). Equivalently, studies have shown that prices depend on both objective and risk-neutral probabilities (Chen (2009), Bhamra, Kuehn, and Strebulaev (2010)). However, these papers do not relate their findings to credit ratings, other than using ratings as a control. In the context

[5] "... [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."

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 (2009) and Brennan, Hein, and Poon (2009).[6] 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. In our study we propose an explicit measure of systematic risk and find that credit ratings contain information not only about default probability but also about systematic risk.

The rest of the paper is organized 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 investigates further the information in credit ratings and failure score relevant to default; section 4 relates ratings to systematic default risk; the last section concludes.

2 Measuring corporate default probability

In the first part of the paper we explore the information about raw default probability in corporate credit ratings. To do this we perform two empirical exercises. We first propose a direct measure of raw default probability, an empirical measure based on publicly available accounting and market-based information. We then examine the ability both of our measure and of ratings to forecast default. We next analyze further the relationship between our measure of default probability and ratings.

We begin by introducing and discussing our measure of default probability. Our method for predicting default follows Campbell et al. (2008) and builds on the earlier work of Shumway (2001) and Chava and Jarrow (2004). Specifically, we use the same failure indicator and explanatory variables as Campbell et al. All of the variables, the specification, and the estimation procedure (described in more detail in section 2.2) are discussed in Campbell et al., who also show that this specification outperforms other standard methods of default prediction. The model is more accurate than Shumway and Chava and Jarrow, who use a smaller set of

[6] Our study does not examine credit ratings of complex securities. Instead it focusses in the accuracy of credit ratings in what is arguably the agencies' core competence: assessing corporate credit risk.

explanatory variables, and is also more accurate than using distance-to-default, a measure based on the Merton (1974) model (e.g. Vassalou and Xing (2004)).[7]

2.1 Corporate failures and explanatory variables

Our failure 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. The data was provided to us by Kamakura Risk Information Services (KRIS) and covers the period 1963 to 2008.

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 and the percentage of active firms that failed each year (failure rate) in columns 3 and 4. We repeat this information for those firms with an S&P credit rating in columns 5 through 7. Since our data on credit ratings begin in 1986 we mainly focus on reporting statistics for the period from 1986 to 2008. 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%.

From 1986 to 2008 the failure rate exhibits strong variation over time. This variation 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 1.4%. 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 failure rate for rated firms is almost three times higher during and immediately after recessions (2.4%) and crises (2.3%) than it is

[7] Bharath and Shumway (2008) also document that a simple hazard model performs better than distance-to-default.

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 daily and monthly CRSP files and 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.

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 estimate the vector $\hat{\delta}$ and refer to the linear combination $\hat{\delta}_s' x_{it}$ as the 'failure score' of firm i in month t. Failure score and failure probability are then (positively) related by equation (1).[8]

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. The market-to-book ratio enters with a positive sign. Firms with lower price per share are more likely to fail and size enters with a counterintuitive positive coefficient, which is most likely driven by the high correlation of size and price. At the 12-month horizon, the results are similar, except that size and price are insignificant.

As measures of fit we report McFadden's pseudo R^2 which is equal to 31.6% and 11.8% for the 1-month and 12-month models. For comparison, Campbell et al. report a pseudo R^2 of 31.2% for their 'best model,' 27% for Shumway's (2001) model, and 15.9% when using distance-to-default. We also report the accuracy ratio which 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 50% (random model) to 100% (perfect model). It is a commonly used measure when evaluating a binary response model. For the two models the accuracy ratios are equal to 95.5% and 86.2%.

[8] 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:

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

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

3 Information about default probability in credit rating

Having constructed our measure of raw default probability we can now compare our failure score with S&P long-term general corporate credit rating as a predictor of default.[9] Data on monthly S&P credit ratings are from Compustat.[10] 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:

$$P_t(Y_{i,t+s} = 1) = (1 + \exp(-\alpha_s - \gamma_s \left(\delta_s' x_{it}\right) - \phi_s Rating_{it}))^{-1} \tag{2}$$

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 rating 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.[11]

S&P credit ratings for firms that are not in default run from AAA to C. Ratings from AA to CCC are further divided into 3 subgroups each with a '+' or a '–' added to the rating (e.g. A+, A, A–). We assign a score of 1 to AAA and each reduction in rating receives an additional score of 1 so that BBB (the lowest investment grade rating) is assigned a score of 9 and C (one notch above default) receives a score of 21. Thus our ratings variable, like failure score, is positively related to default risk. The assumption of linearity does not affect our results of

[9] In addition to ratings provided by rating agencies, banks often develop internal ratings. Carey and Hrycay (2004) and Krahnen and Weber (2001) discuss that rating process.

[10] 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.

[11] If we instead estimate the unrestricted regression, failure score performs better, and outperforms rating by more at all horizons.

relative forecast accuracy; we discuss robustness checks in more detail in section 3.1.3.

3.1 Relative forecast accuracy of credit rating and failure score

3.1.1 Marginal forecast accuracy

Table 3 reports the results from our estimation of the baseline model in equation (2). 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. We are estimating the probability of default at these horizons conditional on no previous default. This means that we are intuitively estimating forecast accuracies of marginal default probabilities at different points in time. We consider cumulative forecast accuracies in section 3.1.2.

Failure score predicts default at horizons of one month with a pseudo R^2 of 40% versus 29.2% for rating alone, which means that failure score outperforms rating by 10.8 points. Adding rating to failure score increases the pseudo R^2 from 40% to 42.4%. Thus, rating appears to contain little additional information about the probability of failure in the immediate future, while failure score significantly outperforms rating.

Figure 1 plots the pseudo R^2 for all horizons from our baseline model for failure score only, rating only, and both together. Since we expect a large increase in uncertainty at longer horizons we expect marginal forecast accuracies to diminish with the forecast horizon. This is what we find; 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 42.4% at the 1-month horizon to 5.6% at the 60-month horizon. Failure score continues to outperform rating in the medium term, at horizons of 3, 6, 12, and 24 months. Failure score outperforms rating by 14.2 and 12.1 points at the 3 and 6 months horizons and by 7.8 and 0.7 points at the 12 and 24 months horizons. At 36 months both measures have close to the same forecast accuracy and for the 4 and 5-year horizons rating only is a slightly more accurate predictor than failure score only (we discuss shortly that this small advantage cannot make up for the lower accuracy of rating at short horizons). Nevertheless, using both measures is always better in terms of accuracy than using only one of the measures. Table 3 also reports accuracy ratios and we find that using them results in the same pattern across the different prediction horizons.

Table 3 panel B reports the coefficient estimates and their associated z-statistics for the specification including both failure score and rating. The significance levels of credit rating and failure score, when both are included, reflect the relative performance of the individual measures. The pattern in z-statistics reflects the pattern in pseudo R^2 – both are statistically significant at all horizons, but failure score is much more significant up to about 2 years, significance levels are similar at 3 years, and rating is more significant at 4 and 5 years. The significance levels of the coefficients also reflect the incremental information of the two measures. For example, if rating is not informative when added to failure score, the coefficient on rating when both are included would be indistinguishable from zero. If it is different from zero, rating adds information. This means that the additional information contained in failure score and rating is statistically significant at all horizons.

3.1.2 Cumulative forecast accuracy

We also consider the ability of ratings and failure score to predict cumulative failure probabilities at longer horizons. We expect that the slightly superior performance of rating at predicting the marginal probability of failure 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 at shorter horizons. This means that we expect failure score to also outperform when we compare it to rating as a predictor of default between date t and date $t + s$. The area under each line in figure 1 can be thought of as an estimate of the ability to forecast default over time (cumulative probability), rather than at some future point (marginal probability). 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.

To consider the relative accuracy more formally we construct for each January 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. Panel C of table 3 reports the pseudo R^2 measures which decline monotonically with the horizon but are always higher for failure score only than for rating only. At horizons of one year, failure score's forecast accuracy is 41.4% compared to 24.1% for rating, a difference of 17.3 points. Adding rating to failure score increases the pseudo R^2 from 41.4% to 42.7%. At 5-year horizons, failure score predicts correctly cumulative failure events 23.2% of the time versus 18.0% for rating only, a difference of 5.2 points, and failure score outperforms rating at all intervening horizons. Adding rating to failure score increases the pseudo R^2 from 23.2% to 26.7% at 5-year horizons.[12]

It may not be too surprising that failure score is a good forecast of default at short and medium horizons: most investors should presumably be aware of impending disaster at such short horizons and equity market data, such as past returns and volatility, will likely reflect this awareness. However, all the information available to the market is also available to the rating agencies.[13] We conclude that by ignoring or not responding to publicly available early warning signals of default at horizons of up to 3 years, ratings fail as an optimal forecast of default at horizons of up to 5 years. Whatever else ratings may be, they are not optimal forecasts of default.

3.1.3 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 estimate coefficients on failure score

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

[13] In fact, it may be that rating agencies have additional information, that is not available to the public (see, for example, Jorion, Liu, and Shi (2005)). If they do have such information, and if this information is reflected in the rating, it does not seem to make up for their seemingly slow response to market data.

from 1963 to 2003 (the same period as in Campbell et al.) and then test relative out-of-sample performance using data on failure events 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 the 2004-2008 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 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. We find that the discretized failure score predicts default at a similar level of accuracy as continuous failure score which means that it performs as well relative to ratings.

Third, one might be concerned that our results are 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 (2004a)). 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: (a) we include separate dummy variables for recessions and financial crises and compare relative performance. (b) 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. (c) 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. We find that none of these alternative specifications significantly affects the results in table 3.

Fourth, another concern could be that our results are 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. We group firms into 10 groups by rating and estimate a logit model allowing the coefficient on the dummy variables to vary freely.[14] Again, we find that failure score outperforms rating by a substantial margin in predicting default.

Fifth, it is possible that ratings do a poor job at predicting failure because a typical rating is stale, but that ratings changes or ratings that have recently changed are much better at predicting default.[15] We address this concern in two ways: (a) We add the interaction of rating change and rating to our standard specification. If ratings that recently changed contain more information this change should lead to an increase in forecast accuracy. (b) We include a downgrade indicator as an additional variable. Downgrades may contain important information about the possibility of future default and allowing for an additional effect may increase

[14] 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.

[15] Such an interpretation would be consistent with Hand, Holthausen, and Leftwich (1992) who document bond price effects in response to rating changes, implying that such changes are viewed as news.

accuracy. This check also addresses the concern that ratings changes are asymmetrically informative and that only downgrades really matter. Neither change to our main specification affects our results materially.

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

3.2 The relationship between default probability and rating

The fact that a simple model, combining accounting and market-based variables, dominates ratings as a default predictor provides evidence that ratings are not primarily an optimal estimate of raw default probability. We now explore this hypothesis further by analyzing the extent to which rating reflects information related to default probability. We first provide additional evidence that the fitted values of our model may be regarded as benchmark estimates of default probability and then present evidence on how ratings relate to these estimates.

3.2.1 Further motivation for failure score as a benchmark measure of default probability

If variation in default probability is viewed by market participants as being informative it should be reflected in market prices.[16] To check this we consider the ability of our estimates of default probability to explain variation in credit default swap (CDS) spreads. 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 a high default probability (expected loss) as well as the

[16] The idea that default probability is related to yield spreads on risky debt was suggested as early as Fisher (1959). Specifically, Fisher motivates his approach by an earlier quote that "No person of sound mind would lend on the personal security of an individual of doubtful character and solvency." More recently, Huang and Huang (2003) and Duffie et al. (2008) have explored this idea.

asset's risk premium. At this point we consider only if variation in spreads across issuers and over time can be attributed to raw default probability. We return to the effect of systematic risk in section 4.

Table 4 panel A presents results of regressions of log spreads on log 12-month failure probability.[17] We assume a linear relationship[18] and include rating fixed effects (columns (1) and (2)), rating and year fixed effects (columns (3) and (4)), and firm fixed effects (columns (5) and (6)). For each set of fixed effects we then run a regression with and without failure probability. In all specifications failure probability is a highly economically and statistically significant determinant of CDS spreads. A 1% increase in failure probability is associated with a 0.44% to 0.61% increase in spreads. Failure probability explains 30% of within rating variation and 30.5% of within firm variation in CDS spreads. The information in failure probability is also reflected in overall R^2: adding failure probability to a model containing only rating fixed effects results in an increase of overall R^2 from 64.5% to 75.2%; adding failure probability to rating and year fixed effects increases overall R^2 from 77.7% to 82.6%.[19] We conclude that our estimates of default probability contain important information reflected in market prices.

We also present evidence that failure probabilities predict downgrades. In panel B of table 4 we estimate logit regressions of an indicator variable that is equal to one if a firm is downgraded during the next month and use failure score as our explanatory variable. We control for rating effects (columns (1) and (2)) and rating and year effects (columns (3) and (4)). For each set of dummies we estimate models with and without failure score. Since the coefficient on failure score is very stable across rating groups we again pool the data. We find that the coefficient on failure score is highly statistically significant and that failure score adds substantial explanatory power. When including failure score together with rating dummies the pseudo-R^2 increases from 1.3% to 10.6%; when adding failure score to rating and year dummies

[17] Standard errors are clustered by year to take into account the possibility of cross-sectional correlation.

[18] The assumption of linearity is consistent with an earlier study by Berndt, Douglas, Duffie, Ferguson, and Schranz (2008). In addition, in unreported results, we find strong evidence for a linear relationship, with coefficients stable across ratings groups.

[19] These results are consistent with Ederington, Yawitz, and Roberts (1987), who find that accounting measures such as coverage and leverage contain information about spreads on industrial bonds that are not reflected by rating.

the pseudo-R^2 increases from 2.4% to 11.4%. The accuracy ratios reflect the same pattern: including failure score increases the accuracy ratios by 17.3 and 13.7 points respectively.

The evidence in table 4 indicates that variation in our estimates of default probability are reflected in market prices and contain information about ratings downgrades. In addition, tables 2 and 3 show that our estimates of default probability predict default well and better than ratings at horizons of up to five years. We conclude that failure score is an accurate measure of raw default probability.

3.2.2 How credit ratings relate to failure probabilities

We now treat our estimates of default probability as observations of actual raw default probability and continue to explore the information in rating relevant for predicting default.

Ratings do not clearly separate firms by default probability, though the ranking of average default probabilities is correct. If rating measures raw default probability then variation in rating should explain variation in default probability. To explore the information about default probability in rating we therefore compare fitted failure probabilities across credit ratings. In table 5 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 8.4%. In between, mean default probability increases monotonically from AAA to CCC. However, there is also substantial within-rating variation, which is even larger in magnitude than the mean default probability of the rating. For example, for BBB firms the mean default probability is equal to 0.61%, while the standard deviation is 0.68%. This large within-rating dispersion suggests that ratings do not clearly separate firms into categories by default probability. We note that, as we have previously shown, within-rating dispersion in default probability is reflected in CDS spreads and therefore does not represent only noise.

We next look at the distribution of fitted failure probabilities across credit ratings. Figure 2 presents box plots of failure 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 failure 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 subtract from every failure probability 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 shortly.

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 declines from AAA to CC. Second, as suggested by the results in table 5, there is considerable overlap across ratings. For example, the 75th percentile default probability for any rating is always higher than the median for the next lower rating, or even for that two notches lower. Third, the overlap in distributions is much more obvious for investment grade issuers: there appears to be almost total overlap for issuers rated between AA and BBB-. There is so much overlap 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 issuer rated BBB-, the last rating before reaching junk status. 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 raw default probability. In a regression of log default probability on rating, rating explains only 20% of the variation in default probability for non-investment grade issuers. For investment grade issuers, the relationship is even weaker: Credit rating only explains 3% of the variation in default probability.

For a given rating, a firm's default probability varies over time. We now turn our attention to examining variation in default probability by rating over time. We know that average default rates are higher during recessions and financial crises (table 1); however, ratings do not appear to capture this variation in default probability over the business cycle. In table 5, 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 0.10% in good times and 0.21% in bad times (table 5).[20]

To examine time variation in default probability further figure 3 plots median annualized 12-month failure probabilities over time for the 5 rating categories AA, A, BBB, BB, and B (the 5 letter ratings with the most available observations). Although the ranking of median failure probability by rating is largely 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 raw default probability, the lines in figure 3 would be roughly flat and parallel.[21]

The strong variation in default probabilities over time may be related to the rating agencies' stated practice to 'rate through the cycle' (Amato and Furfine (2004), Löffler (2004a)). This practice implies that ratings may do a poor job measuring time variation in default probability but leaves open the question of how large this underlying variation actually is. Figure 3 quantifies the inability of rating to reflect fluctuations in raw default probability and demonstrates that this variation is substantial.

We also confirm that, consistent with the practice of 'rating through the cycle,' the share of firms in a particular rating group does not vary directly with business conditions. Figure 4 plots the share of firms rated AA, A, BBB, BB, and B. 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

[20] We explore this pattern further in section 4 when we relate rating to measures of systematic risk.

[21] Our results relate to several previous studies that also find that default probabilities vary countercyclically. 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).

is no clear tendency of the share of lower-rated issuers to increase during and after recessions and financial crises.[22]

The results in this section present evidence that failure score, 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. Estimated default probabilities are strongly related to CDS spreads and also predict downgrades. Treating fitted failure probabilities as measures of actual raw default probabilities we find that although ratings rank firms correctly in terms of broad averages, they do not clearly separate firms by default probability. Furthermore, for a given rating, there is a strong tendency for default probabilities to vary over time, especially increasing in bad times. All of these results indicate that ratings, contrary to what is often assumed, are not primarily or exclusively a measure of firm-specific raw default probability.

4 Systematic risk and credit rating

We now ask if ratings instead measure systematic risk. 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's 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 having "[a]dequate capacity to meet financial commitments, but more subject to adverse economic conditions."

Figure 3 shows a strong tendency for median default probabilities of different ratings to spread out in recessions and crises so that the default probability of lower-rated firms increases by more during bad times. This suggests that rating reflects sensitivity of credit risk to bad times and, therefore, that rating may at least partly capture systematic risk of default. We now consider this hypothesis directly. In the next subsection, we introduce our measure of a firm's systematic default risk: failure beta. We then present evidence that ratings separate firms by

[22] We note that the lack of a strong association of rating share and business conditions is also consistent with the inability of year dummies to explain variation in the downgrade rate in table 4.

failure beta and that failure beta is priced in the cross-section of CDS risk premia. Finally, we check that our claim that ratings capture systematic risk is robust to using alternative measures of systematic risk.

4.1 Measuring systematic default risk: failure beta

We begin by identifying a measure of systematic default risk. In order to measure the extent to which a firm's default risk is exposed to common and therefore undiversifiable variation in default probability we must first construct a measure of such common variation. To do this we assume that default probabilities have a single common factor, and estimate this common factor using principal component analysis. Extracting principal components in the standard way from the full panel of rated firms is problematic because the cross-section is much larger than the time series. We therefore first shrink the size of the cross-section by assigning each firm-month to a given rating-month and calculating equal-weighted average 12-month cumulative default probabilities (as used in figure 3). We perform the same exercise grouping the firms by industry. This leaves us with two panels: the ratings panel consists of 18 ratings groups with 276 months of data; the industry panel consists of 29 Fama-French industries (30 industries excluding coal, for which we have insufficient data) again with 276 months. For each panel we extract principal components in the standard way.

We find clear evidence of common variation in default probabilities. For the ratings panel, we find that the first principal component explains 70.3% of variation in default probability, while the second principal component explains 9.5% and the third 5.8%. For the industry panel, the corresponding figures are 41.7%, 10.8% and 7.5%. In addition, both first principal components are capturing very similar variation: the correlation between the two is 0.954. Our assumption of a single factor is therefore a good approximation of the factor structure of default probabilities, however grouped. We also find that the first principal component is a measure of economy-wide variation in default probability: Both first principal components are close to equally-weighted across ratings and industry groups.

Principal component analysis is a statistical tool used to shrink the dimensionality of the data. For the presence of a common factor to be relevant for asset prices it must be related to the stochastic discount factor. Since default probabilities are high in bad times (see table 5, figure 3) a diversified, risk-averse investor will care about exposure to variation in the common component. In order to gain more insight about the common component figure 5 plots the first principal component of the rating panel over time, together with the median default probability for the full panel of rated firms. Figure 5 also plots the corresponding mean, weighted by book value of liabilities, and the realized failure rate over the following 12 months for each January. Thin vertical lines show when financial crises occur, and grey bars show NBER recessions.

The first principal component, median, and value-weighted mean default probabilities all move together and are noticeably higher during and immediately after recessions and financial crises, when economic theory suggests the stochastic discount factor is high. The first principal component and the median default probability move closely together and have a correlation of 0.945.[23] We therefore interpret the single common factor of default probabilities as the median default probability. Median failure probability is visibly higher in and shortly after bad times and is clearly related to the realized failure rate: The graph reflects the correlation between the median failure probability and the failure rate of 0.64.[24] This means that median failure probability is also a good measure of bad times.

Having identified the common factor and having interpreted it as correlated with the stochastic discount factor, we can estimate factor exposures: the sensitivity of a firm's default probability to the common factor. Specifically, for firm i, with cumulative failure probability P_{it}, and with credit rating CR we estimate:

$$P_{it} = \alpha_{CR} + \beta_{CR} P_t^{median} + \varepsilon_{it}. \tag{3}$$

[23]The correlation of the first principal component and the value-weighted mean default probability is 0.85. For the industry panel the correlation with the median is 0.913 and 0.811 for the mean. The first differences are also highly correlated for both measures.

[24]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 overall median issuer and did not account for a large proportion of total rated corporate issuance.

P_{it} is the 12-month annualized default probability and P_t^{median} is its median across firms.[25] 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 an important determinant of within-rating spread variation in the CDS market (table 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.[26] In order to ensure a sufficient number of observations for each rating, we combine all observations rated CCC and below together.[27]

The specification of (3) does not of itself constrain the dependent variable to lie between zero and one. However, investigation of the volatility of the residuals reveals a strong linear relationship with the square root of default probability.[28] Since almost all of our estimated default probabilities are small (so that $P(1-P) \approx P$), this is consistent, for example, with a continuous time model of default probability and its common component in which the innovations are proportional to $(P_t(1-P_t))^{\frac{1}{2}}$. Such a formulation does constrain default probability, under suitable initial conditions, to lie between zero and one. In addition, under this specification, OLS estimates of β_{CR} will be unbiased and consistent, since the first principal component is by construction orthogonal to the remaining residual variation. This means that we need only be concerned about heteroskedastic errors. We therefore use White standard errors to allow for heteroskedasticity of unknown form. We also cluster by date, to correct for cross-sectional correlation of the residuals.

[25] Or equivalently its first principal component. Our results are very similar if either principal component is used instead.

[26] To control for outliers for each rating group we winsorize default probabilities at the 0.5% and 99.5% levels; to control for firm-specific variation in default probability we include firm fixed effects.

[27] 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.

[28] Results are available from the authors on request.

4.2 Failure beta and credit rating

We estimate failure beta across 18 ratings ranging from AAA to CCC and below. Consistent with the previous suggestive evidence (figure 3, table 5) we find a strong relationship between failure beta and credit rating. Figure 6 plots estimates of failure beta by rating, together with 95% confidence intervals. There is substantial variation in sensitivities to bad times; failure beta is equal to 0.34 (AAA), 0.92 (BBB), 2.07 (BB), and 12.6 (CCC and below). The variation in failure beta is closely related to 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 and the correlation between rating score and log failure beta is equal to 0.972. 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 establishes that ratings are informative about the tendency of an issuer's propensity to default to worsen in bad times. Ratings, therefore, are measuring at least two distinct forms of credit risk: default probability and systematic risk.

We note that our results are not at odds with the existing literature on the determinants of corporate bond spreads: The results in Huang and Huang (2003), Chen (2009), Coval at al., (2009), and Bhamra et al. (2010) indicate that a higher share of higher-rated credit spreads is due to systematic risk (e.g. that the fraction of the spread due to systematic risk is higher for AAA than it is for junk). Our results imply that even though the *share* of the spread due to systematic risk may be higher for higher-rated credit, the absolute *level* of the systematic risk is lower. For example, even though AAA bonds are unlikely to default in bad times (they have low systematic risk), if they ever default the default will almost surely occur in bad times and so the share of the spread due to systematic risk is high. However, they are very unlikely to default at any time. By contrast, BB firms are much more likely not to survive bad times than are AAA firms and so they have higher systematic risk (and higher failure beta). But BB bonds are also much more likely to default overall so their share of systematic risk may be lower. Elkamhi and Ericsson (2008) show that such a pattern is consistent with structural models of corporate bond pricing.

The high sensitivity of low-grade issuers to bad times (as well as the presence of an important common factor in default probability) also establishes that the risk of failure is systematic. This means that our results contribute to the ongoing discussion of whether or not bankruptcy and default risk represent systematic risk. Dichev (1998) finds that high bankruptcy and distress risk is not associated with high equity returns and concludes that high bankruptcy risk is not a systematic risk.[29] In contrast, our results suggest that there is such a thing as systematic bankruptcy risk. In the next section we demonstrate that systematic default risk is reflected in asset prices. Specifically, we show that high failure beta is associated with high CDS risk premia.[30]

4.3 Failure beta and CDS risk premia

In this section we show that variation in failure beta can explain variation in CDS risk premia, providing supporting evidence that failure beta is related to systematic risk and that investors demand compensation for exposure to it. Economic theory suggests that investors will demand a higher expected return (and that they will use a higher discount rate) for those firms' credit risks that have higher levels of systematic risk. This means that if failure beta is a good measure of systematic risk, and if variation in systematic risk is priced in the credit market, we should expect failure beta to be related to CDS risk premia.

We use our measure of default probability and expected recovery rate data from Markit to calculate a simple measure of CDS risk premia. We consider the price of a discount bond that defaults with probability P_t. In the case of no default it has a payoff of 1 and in the case of default its expected payoff equals the expected recovery rate, which we denote as rec. The expected payoff is discounted at a rate that includes a risk premium RP and the risk-free rate

[29] Campbell et al. point out that low equity returns for financially distressed firms may be the result of information or arbitrage-related frictions. Consistent with this interpretation, Dichev and Piotroski (2001) find that after downgrades low returns occur mainly around earnings announcements which means that they are probably associated with under-reaction to news instead of low systematic risk.

[30] Consistent with our finding a close link between CDS risk premia and failure beta (discussed in the next section), Anginer and Yildizhan (2010) find that when using credit spreads to proxy for distress risk such stocks no longer deliver anomalously low returns.

R_f.[31] Today's price then satisfies:

$$price = \frac{1 - P_t(1 - rec)}{1 + R_f + RP}. \tag{4}$$

Holding default probability P_t constant, a higher failure beta, which is 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 by extracting risk premia from CDS spreads and relating them to failure beta.

For a pure discount bond spreads are defined as

$$price = \frac{1}{1 + R_f + spread}. \tag{5}$$

Therefore

$$RP = (1 + R_f + spread)(1 + P_t(1 - rec)) - (1 + R_f). \tag{6}$$

Using CDS spreads and recovery rates from Markit,[32] together with our estimated failure probabilities, we use this relationship to calculate implied risk premia for all issuers in our CDS sample. We then calculate the median risk premium for each rating-year group for which we have data.

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 vertical group of points corresponds to a rating, for each of which we estimate a separate failure beta. As expected, there is a strong positive relationship between failure beta and CDS risk premia.

Table 6 quantifies the relationship between risk premia and failure beta. We find a strongly statistically significant relationship that is steeper for investment grade issuers. Column (1)

[31] Similar calculations have been used by Elton, Gruber, Agarwal, and Mann (2001), De Jong and Driessen (2007), Campbello, Chen, and Zhang (2008), and Anginer and Yildizhan (2010).

[32] We have checked that our results are robust to using alternative measures of expected recovery rates: following Elton, Gruber, Agarwal, and Mann (2001) we use recovery rates from Altman and Kishore (1998). We have also used zero recovery rates. Our results are robust to either of these two alternatives.

regresses median log risk premia on log failure beta. Log failure beta is statistically significantly related to CDS risk premia and explains 81% of the variation in risk premia across the 108 rating-years. Column (2) shows that this result is robust to the inclusion of year fixed effects. Column (3) adds an interaction of log failure beta with an investment grade dummy and shows that the relationship is significantly stronger for investment grade ratings.

The next two columns include log default probability as an additional control. Our calculation of CDS risk premia (equation (6)) already controls for the level of default probability and so we do not expect default probability to contain explanatory power.[33] Consistent with our intuition we find that default probability is not important in explaining variation in risk premia over and above failure beta and that the relationship between risk premia and failure beta is robust to this control. We find a marginally significant effect of default probability (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.

Our conclusion is that our straightforward measure of systematic default risk is strongly related to credit rating and CDS risk premia. This relationship is robust to controlling for time effects and default probabilities. 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.

4.4 Alternative measures of systematic risk

Failure beta is an explicit measure of the tendency of default probability to increase in bad times. We now consider the robustness of the relationship between credit ratings and systematic risk to obvious alternative measures of issuer systematic risk. Our objective is first, to investigate the robustness of our finding of the relationship between rating and systematic risk and second, to increase our understanding of the reason why failure beta and CDS risk premia are related.

[33] This is in contrast to the earlier regression in table 4 where we relate CDS spreads (as opposed to risk premia) to default probabilities.

The first measure of systematic risk we consider is CAPM beta. Table 7 presents evidence on how CAPM beta is related to rating. In panel A we include all firms, while in panel B we focus on investment grade firms only. As before, our estimates of beta are pooled by rating group. The left half of the table presents evidence on how credit rating is related to measures of systematic risk and the right half of the table relates our estimates of CDS risk premia to these same measures. We find that CAPM beta is strongly related to rating: variation in CAPM beta explains 87% of the variation in rating. Even though rating and beta are widely studied measures of risk, to our knowledge this relationship is not widely appreciated. Yet it was pointed out in an early study by Schwendiman and Pinches (1975), who document that CAPM beta is higher for lower-rated issuers.[34] However, although CAPM beta is strongly related to rating overall, we find that the relationship is weaker than it is for failure beta, which explains 95% of variation in rating. Furthermore, CAPM beta explains a much smaller share of the variation in rating for investment grade issuers (68%) than does failure beta (92%). The pattern for CDS risk premia is similar: CAPM beta explains 50% of the variation in risk premia across investment grade firms compared to 82% of variation explained by failure beta. We conjecture that this pattern is due to the greater ability of failure beta to identify off bad times, such as severe crises and recessions.

To investigate this hypothesis we also consider up and down beta, which are simply estimates of CAPM beta that condition respectively on positive and negative equity index excess returns. If stock returns are non-linearly related to 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)). If failure beta has a greater ability to identify sensitivity to bad times than CAPM beta, we would expect down beta to be related more closely to rating than up beta. This is indeed what we find: down beta is more strongly statistically and economically related to both rating and CDS risk premia (table 7).

The relationship between beta and rating and between beta and risk premia is strongest for failure beta, followed by down beta, then followed by CAPM beta. When considering

[34] Consistent with this finding, Kaplan and Urwitz (1979) find that both accounting ratios and CAPM beta contain information about rating.

investment grade only (panel B) the relationships are similar, but more pronounced. Up beta is not significant and enters with a counterintuitive negative coefficient estimate both when explaining variation in rating as well as risk premia. These results suggest that failure beta captures more of the variation in rating and risk premia than CAPM beta because failure beta is better at measuring the risk of negative events. The fact that failure beta captures more variation than down beta suggests 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.

We conclude that the relationship between rating and systematic risk is robust to alternative measures of systematic risk, though the relationship with rating and risk premia is strongest with failure beta. We argue that this is because failure beta directly measures the tendency of issuers' credit quality to deteriorate in bad times. These are the risks that diversified corporate bond investors care about and ratings seem to reflect this. Moreover, market participants seem to set prices in accordance with this idea given that failure beta is most closely related to CDS risk premia.

5 Conclusion

In this paper we investigate the information in corporate credit ratings relevant to investors concerned about credit risk. We stress, in particular, that credit risk can mean both individual firm probability of default as well as systematic default risk: the tendency to default in bad times.

In the first part of the paper we analyze the information about corporate default contained in corporate credit ratings. Based on the following evidence we argue that ratings are not an optimal measure of default probability. (1) 'Failure score,' a linear combination of accounting and market-based measures of financial distress (based on Campbell, Hilscher, and Szilagyi

(2008)) is better than ratings at forecasting corporate default and failure. Forecasting default over the next year from 1986 to 2008, failure score explains 41% of the variation as compared to 24% for credit rating. Ratings do provide incremental information for predicting default at all horizons, though including them increases forecast accuracy only marginally. (2) Credit ratings explain little of the variation in default probability. There is substantial variation in default probability within ratings, especially for investment grade firms, and there is an extremely high degree of overlap in the distribution of default probabilities across ratings; a regression of default probability on rating for investment grade bonds has an R^2 of 3%. (3) Ratings do not reflect the considerable variation in default probabilities and empirical failure rate over the business cycle: Although the empirical default frequency during recessions and crises is much higher (2.4%) than during normal times (0.9%) the shares of firms with different ratings changes very little.

In contrast, failure score, a simple measure based on publicly available information, dominates rating as a predictor of default and contains important information relevant for market prices and the risk of deteriorating credit quality. Fitted failure probabilities are better at predicting marginal default probabilities at horizons up to 2 years and better at forecasting the cumulative failure probability at horizons up to 5 years; they explain 30% of the within rating variation in CDS spreads, and they predict downgrades. If the primary objective of ratings were to measure default probability, rating agencies could adopt our simple methodology. We conclude, therefore, that credit ratings are not optimal measures of default probability. This means that either credit ratings are simply not at the frontier of default prediction or that delivering optimal default probability forecasts is not the sole objective of rating agencies.

In the second part of the paper we argue that credit ratings capture variation in systematic default risk well and that such variation is important to investors. Systematic default risk is the tendency to default in bad times, times when share prices are low, unemployment is high, and other bonds are more likely to default. We propose failure beta, the sensitivity of a firm's default probability to the common component of default probability, as our measure of systematic risk and find that it is strongly related to credit rating. We document that investors care about differences in systematic risk: CDS risk premia are closely related to failure beta, even after controlling for default probability. Our findings are robust to using more conventional measures of systematic risk, such as CAPM beta, though failure beta is better at explaining the variation. We argue that this is the case since failure beta in particular identifies sensitivity to adverse conditions. We conclude that ratings relate to two somewhat different aspects of credit risk: raw default probability (expected payoff) and systematic default risk (discount rate).

If rating agencies are in fact deliberately targeting systematic risk (as is supported by their explanations of what different ratings mean), we can rationalize a number of potentially puzzling aspects of corporate credit ratings: First, why ratings are relatively inaccurate at predicting default when compared to using simple default predictors based on publicly available information – doing so is not their sole objective. Second, why agencies 'rate through the cycle' and respond sluggishly to new information – sensitivity to bad times may not change over the business cycle. Third, why investors pay so much attention to ratings and why ratings are strongly related to bond risk premia.

Another possibility is that rating agencies, perhaps for institutional reasons, simply focus on providing smooth and slow-moving indicators that capture credit risk over long horizons and change little over the business cycle. Such a product may be designed to cater to the demands of buy and hold investors, for example pension funds who may have a desire to rebalance portfolios only infrequently and are concerned mainly with their portfolios' long run risk, or corporate executives with an eye on their cost of capital. In such a setting ratings may possibly end up capturing systematic default risk even though measuring it may not be their primary focus. Intuitively, if a firm's default probability is more sensitive to bad times then its overall likelihood of failure is higher in the long run. However, intentional or not, if ratings are good measures of systematic default risk rational investors will pay attention to them and will demand a premium for a lower rating. This is true even if ratings are not optimal measures of default probability.

What are the implications for policy? There has been recent regulatory pressure on rating agencies and a call for potential alternatives to credit ratings. Our findings suggest that the ability to capture systematic risk may mean that, despite the availability of superior default prediction methods, ratings will continue to be part of investors' and executives' toolkits. However, there is no restriction that states that credit risk has to be summarized by only one measure. Instead, it may be fruitful to break out short-run default prediction from the measurement of long-run systematic risk. The former measure could update frequently and rapidly respond to news while the latter could be similar to current credit ratings.

References

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

[2] Altman, Edward I., and Vellore M. Kishmore, 1998, "Defaults and returns on high-yield bonds: analysis through 1997," NYU Salamon center working paper.

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

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

[5] Anginer, Deniz, and Celim Yildizhan, 2010, "Is there a distress risk anomaly? Corporate bond spread as a proxy for default risk," unpublished paper, Stephen M. Ross School of Business, University of Michigan at Ann Arbor.

[6] Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2009, "Dispersion in analysts' earnings forecasts and credit rating," *Journal of Financial Economics* 91, 83-101.

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

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

[9] 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.

[10] Bhamra, Harjoat S., Lars-Alexander Kuehn, and Ilya A. Strebulaev, 2010, "The levered equity risk premium and credit spreads: a unified framework," *Review of Financial Studies* 23, 2, 645-703.

[11] Bharath, Sreedhar T. and Tyler Shumway, 2008, "Forecasting default with the Merton Distance to Default model," *Review of Financial Studies* 21, 3, 1339-1369.

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

[13] 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* 53, 4, 1389-1413.

[14] Brennan, Michael J., Julia Hein and Ser-Huang Poon, 2009, "Tranching and rating," *European Financial Management* 15, 5, 891-922.

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

[16] Campello, Murillo, Long Chen, and Lu Zhang, 2008. "Expected returns, yield spreads, and asset pricing tests," *Review of Financial Studies* 21, 1297-1338.

[17] Carey, Mark, and Mark Hrycay, 2001, "Parameterizing Credit Risk Models with Rating Data," *Journal of Banking and Finance* 25, 1, 197-270.

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

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

[20] Chen, Hui, 2009, "Macroeconomic conditions and the puzzles of credit spreads and capital structure," *Journal of Finance*, forthcoming.

[21] Coval, Joshua, Jakub Jurek, and Erik Stafford, 2009, "Economic catastrophe bonds," *American Economic Review* 99, 3, 628-666.

[22] De Jong, Frank, and Joost Driessen, 2007, "Liquidity risk premia in corporate bond markets" unpublished paper, Tilburg University and University of Amsterdam.

[23] Deventer, Donald R. van, 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.

[24] Dichev, Ilia D., 1998, "Is the risk of bankruptcy a systematic risk?" *Journal of Finance* 53, 3, 1141–1148.

[25] Dichev, Ilia D., and Joseph D. Piotroski, 2001, "The long-run stock returns following bond ratings changes" *Journal of Finance* 56, 1, 173-203.

[26] Duffie, Darrell, Andreas Eckner, Guillaume Horel, and Leandro Saita, 2009, "Frailty correlated default," *Journal of Finance* 64, 5, 2089-2123.

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

[28] Ederington, Louis H., Jess B. Yawitz, and Brian Roberts, 1987, "The Informational Content of Bond Ratings," *Journal of Financial Research* 10, 3, 211-226

[29] Elkamhi, Redouane, and Jan Ericsson, 2008, "Time varying risk premia in corporate bond markets," unpublished paper, Henry B. Tippie College of Business, University of Iowa and Faculty of Management, McGill University.

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

[31] Fisher, Lawrence, 1959, "Determinants of risk premiums on corporate bonds" *Journal of Political Economy* 67, 3, 217-237.

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

[33] Geanakoplos, John, 2009, "The Leverage Cycle," Cowels foundation discussion paper No. 1715, Yale University.

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

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

[36] Hand, John R. M., Robert W. Holthausen, and Richard W. Leftwich, 1992, "The effect of bond rating agency announcements on bond and stock Prices," *Journal of Finance* 47, 2, 733-752.

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

[38] Hillegeist, Stephen A., Elizabeth Keating, Donald P. Cram and Kyle G. Lunstedt, 2004, "Assessing the probability of bankruptcy," *Review of Accounting Studies* 9, 5-34.

[39] Huang, Jing-Zhi and Ming Huang, 2003, "How much of the corporate-treasury yield spread is due to credit risk?" unpublished paper, Smeal College of Business, Penn State University and Graduate School of Business, Stanford University.

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

[41] Jorion, Philippe, Zhu Liu, and Charles Shi, 2005, "Informational effects of regulation FD: evidence from rating agencies," *Journal of Financial Economics* 76, 309-330.

[42] Kaplan, Robert S., and Gabriel Urwitz, 1979, "Statistical models of bond ratings: A methodological inquiry," *Journal of Business* 52, 231-261.

[43] Kiyotaki, Nobuhiro, and John Moore, 1997 "Credit cycles," *Journal of Political Economy* 105, 2, 211-248.

[44] Krahnen, Jan Pieter, and Martin Weber, 2001, "Generally accepted rating principles: A primer," *Journal of Banking and Finance* 25, 1, 3-23.

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

[46] Löffler, Gunter, 2004b, "Ratings versus market-based measures of default risk in portfolio governance," *Journal of Banking and Finance* 28, 11, 2715-720.

[47] 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.

[48] Merton, Robert C., 1974, “On the pricing of corporate debt: the risk structure of interest rates,” *Journal of Finance* 29, 449-70.

[49] Molina, Carlos A., 2005, “Are firms underleveraged? An examination of the effect of leverage on default probabilities,” *Journal of Finance* 60, 1427-1459.

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

[51] Schwendiman, Carl J., George E. Pinches, 1975, “An Analysis of Alternative Measures of Investment Risk,” *Journal of Finance* 30, 1, 193-200.

[52] Sharpe, Steven A., 1994, “Financial market imperfections, firm leverage, and the cyclicality of employment,” *American Economic Review* 84, 1060-1074.

[53] Shumway, Tyler, 2001, “Forecasting Bankruptcy more Accurately: A Simple Hazard Model,” *Journal of Business* 74, 101-124.

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

[55] West, Richard R., 1970, “An Alternative Approach to Predicting Corporate Bond Ratings,” *Journal of Accounting Research* 8, 1, 118-125.

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

Panel A lists the number of firms and failures for all active firms (1963 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. Panel B reports failure rates during the 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 (%)**
1963-1975	2218	50	0.14	.	.	.
1976-1985	4645	339	0.68	.	.	.
1986	5896	101	1.71	907	20	2.20
1987	6331	58	0.92	999	9	0.90
1988	6445	94	1.46	974	30	3.08
1989	6346	77	1.21	932	11	1.18
1990	6269	80	1.28	889	8	0.90
1991	6291	72	1.14	873	15	1.72
1992	6622	49	0.74	913	5	0.55
1993	7149	40	0.56	988	7	0.71
1994	7793	33	0.42	1082	4	0.37
1995	8099	43	0.53	1156	5	0.43
1996	8474	34	0.40	1288	5	0.39
1997	9273	61	0.66	1482	9	0.61
1998	9572	148	1.55	1647	10	0.61
1999	9270	207	2.23	1741	39	2.24
2000	9018	167	1.85	1719	34	1.98
2001	8379	333	3.97	1691	67	3.96
2002	7757	229	2.95	1646	50	3.04
2003	7334	165	2.25	1597	32	2.00
2004	6777	38	0.56	1633	14	0.86
2005	6781	36	0.53	1631	16	0.98
2006	6786	18	0.27	1613	6	0.37
2007	6919	24	0.35	1544	6	0.39
2008	6896	50	0.73	1454	27	1.86

Panel B: Failures during and after recessions and crises

	all firms			rated firms		
	months	**failures**	**rate (%)**	**months**	**failures**	**rate (%)**
normal	359	1167	0.75	183	175	0.89
recession	167	1039	1.42	67	187	2.41
crisis	26	340	1.98	26	67	2.27

Table 2: Failure prediction in a logit model

This table reports 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 (the model is the same as in Campbell, Hilscher, and Szilagyi (2008)): 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 excess return relative to value-weighted S&P 500 return over the previous 12 months (EXRETAVG), log of firm's market equity over the total valuation of S&P 500 (RSIZE), square root of the sum of squared firm stock returns over the previous three-month period, annualized (SIGMA), stock of cash and short term investments over the market value of total assets (CASHMTA), market-to-book ratio 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 total liabilities. Z-statistics (reported in parentheses) are calculated using standard errors that are robust and clustered by year; ** denotes significant at 1%.

	Logit model (1963 to 2008)	
	(1)	(2)
Lag (months)	0	12
NIMTAAVG	-29.00	-20.12
	(16.65)**	(14.11)**
TLMTA	3.509	1.60
	(12.77)**	(8.34)**
EXRETAVG	-8.017	-7.88
	(12.40)**	(9.16)**
SIGMA	1.69	1.55
	(9.31)**	(5.44)**
RSIZE	0.138	-0.005
	(2.84)**	(0.15)
CASHMTA	-2.49	-2.27
	(7.07)**	(7.39)**
MB	0.05	0.07
	(3.23)**	(5.57)**
PRICE	-0.974	-0.09
	(10.26)**	(0.84)
Constant	-8.63	-8.87
	(14.17)**	(17.44)**
Observations	2,022,562	1,870,481
Failures	1,756	2,159
Pseudo R-squared	0.316	0.118
Accuracy ratio	0.955	0.862

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 corresponding 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 (reported in parentheses) are calculated using standard errors that are robust and clustered by year; ** denotes significant at 1%, * significant at 5%. 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	40.0%	34.8%	28.1%	19.8%	9.1%	6.8%	4.9%	3.1%
credit rating only	29.2%	20.6%	16.0%	12.0%	8.4%	7.2%	6.4%	5.4%
both	42.4%	35.9%	29.2%	21.1%	11.2%	8.8%	7.1%	5.6%
Accuracy Ratio								
failure score only	97.0%	96.2%	94.5%	91.8%	83.3%	79.5%	76.8%	72.0%
credit rating only	93.2%	90.3%	88.0%	84.7%	80.8%	79.1%	78.2%	76.2%
both	97.4%	96.5%	95.1%	92.2%	84.9%	82.0%	80.0%	76.7%

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

prediction month	1	3	6	12	24	36	48	60
failure score	0.802	0.929	0.968	1.033	0.831	0.806	0.653	0.388
	(14.10)**	(28.04)**	(29.10)**	(16.72)**	(15.49)**	(9.32)**	(5.58)**	(2.16)*
credit rating	0.307	0.180	0.166	0.170	0.207	0.201	0.213	0.222
	(4.01)**	(4.52)**	(5.00)**	(4.41)**	(8.74)**	(6.71)**	(8.26)**	(8.60)**
# 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	41.4%	33.3%	28.1%	25.5%	23.2%
credit rating only	24.1%	21.4%	19.6%	18.7%	18.0%
both	42.7%	35.4%	30.8%	28.6%	26.7%

Table 4: Information in default probabilities and failure score

Panel A reports results from regressions of log(CDS spread) on log(12-month failure probability) and sets of dummy variables: S&P credit rating, year, and firm. CDS spread data are from Markit and are end-of-month spreads for the 5-year contract. The sample period is 2001 to 2007. All firms with CDS spread, ratings, and available default probability data in the CRSP-COMPUSTAT universe are included. t-statistics (reported in parentheses) are calculated using standard errors that are robust and clustered by year; ** denotes significant at 1%. Panel B reports results from logit regressions of downgrades (a decrease in S&P credit rating, e.g. from BB to BB-) on 12-month failure score estimated using the model in Table 2. Z-statistics (reported in parentheses) are calculated using standard errors that are robust and clustered by year.

Panel A: Explaining variation in CDS spreads using default probability

	(1)	(2)	(3)	(4)	(5)	(6)
log(failure probability)		0.612		0.442		0.713
		(7.52)**		(10.63)**		(15.69)**
R-squared (overall)	64.5%	75.2%	77.7%	82.6%	76.5%	83.7%
R-squared (within)		30.0%				30.5%
Rating fixed effects	X	X	X	X		
Year fixed effects			X	X		
Firm fixed effects					X	X

Number of observations: 38,569

Panel B: Predicting downgrades using failure score

	(1)	(2)	(3)	(4)
Failure score		1.068		1.080
		(26.55)**		(26.37)**
Pseudo R-squared	1.3%	10.6%	2.4%	11.4%
Accuracy Ratio	58.0%	75.3%	62.7%	76.4%
Rating fixed effect	X	X	X	X
Year fixed effect			X	X

Number of observations: 340,496
Number of downgrades: 5,277

Table 5: Fitted failure probability by credit rating

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. Share (%) is the percentage share of the sample rated AAA etc.

rating	# obs	share (%)	PD	PD (good)	PD (bad)	stdev(PD)
AAA	5,011	1.5	0.40	0.36	0.47	0.45
AA	23,100	6.8	0.46	0.42	0.56	0.59
A	79,454	23.3	0.52	0.47	0.62	0.51
BBB	97,925	28.7	0.61	0.52	0.77	0.68
BB	75,481	22.1	1.02	0.80	1.45	1.76
B	54,810	16.1	2.35	1.63	3.61	4.17
CCC, below	5,592	1.6	8.42	6.53	10.89	10.53

Table 6: Regression of risk premia on failure 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 default probability. t-statistics (reported in parentheses) are calculated using standard errors that are robust and clustered by year; ** denotes significant at 1%, * significant at 5%. We include median annual observations for 16 rating groups.

	(1)	(2)	(3)	(4)	(5)
log(failure beta)	1.78	1.84	1.45	1.46	2.24
	(25.02)**	(24.20)**	(28.14)**	(6.90)**	(18.38)**
log(failure beta)*inv. grade			1.47		
			(9.23)**		
log(default probability)				0.151	-0.188
				(1.99)	(4.31)**
Year fixed effects		X	X		X
R-squared (overall)	80.8%	91.3%	94.4%	82.0%	92.5%
R-squared (within)		90.8%	94.1%		92.0%

Number of observations: 108

Table 7: Failure beta and alternative measures of systematic risk

This table reports results from regressions of rating and risk premia on different measures of systematic risk (failure beta, CAPM equity beta, up beta, down beta). Panel A reports results when including 18 rating groups, Panel B reports results for investment grade issuers (rated AAA to BBB-). For risk premia calculations we include median annual risk premia by rating (as in Table 8), and include year fixed effects in the regression; we report within-R^2. CAPM equity beta is estimated using equally-weighted equity portfolio returns; up beta and down beta are 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.

	Regression of rating on beta		**Regression of risk premia on beta**	
Panel A: Investment grade and non-investment grade				
	Coefficient	R-squared	Coefficient	R-squared
failure beta (log)	11.35	94.6%	1.84	90.8%
	(16.66)**		(31.37)**	
CAPM beta	16.24	87.4%	2.34	81.0%
	(10.53)**		(20.62)**	
Up beta	23.81	50.0%	3.40	50.6%
	(4.00)**		(10.12)**	
Down beta	12.09	93.0%	1.78	86.8%
	(14.63)**		(25.59)**	
Panel B: Investment grade only				
failure beta (log)	19.27	91.8%	2.71	81.9%
	(9.44)**		(15.80)**	
CAPM beta	29.01	68.0%	3.76	50.0%
	(4.12)**		(7.42)**	
Up beta	-19.65	16.6%	-2.52	8.6%
	(1.26)		(2.27)*	
Down beta	18.78	86.4%	2.78	69.6%
	(7.12)**		(11.23)**	

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

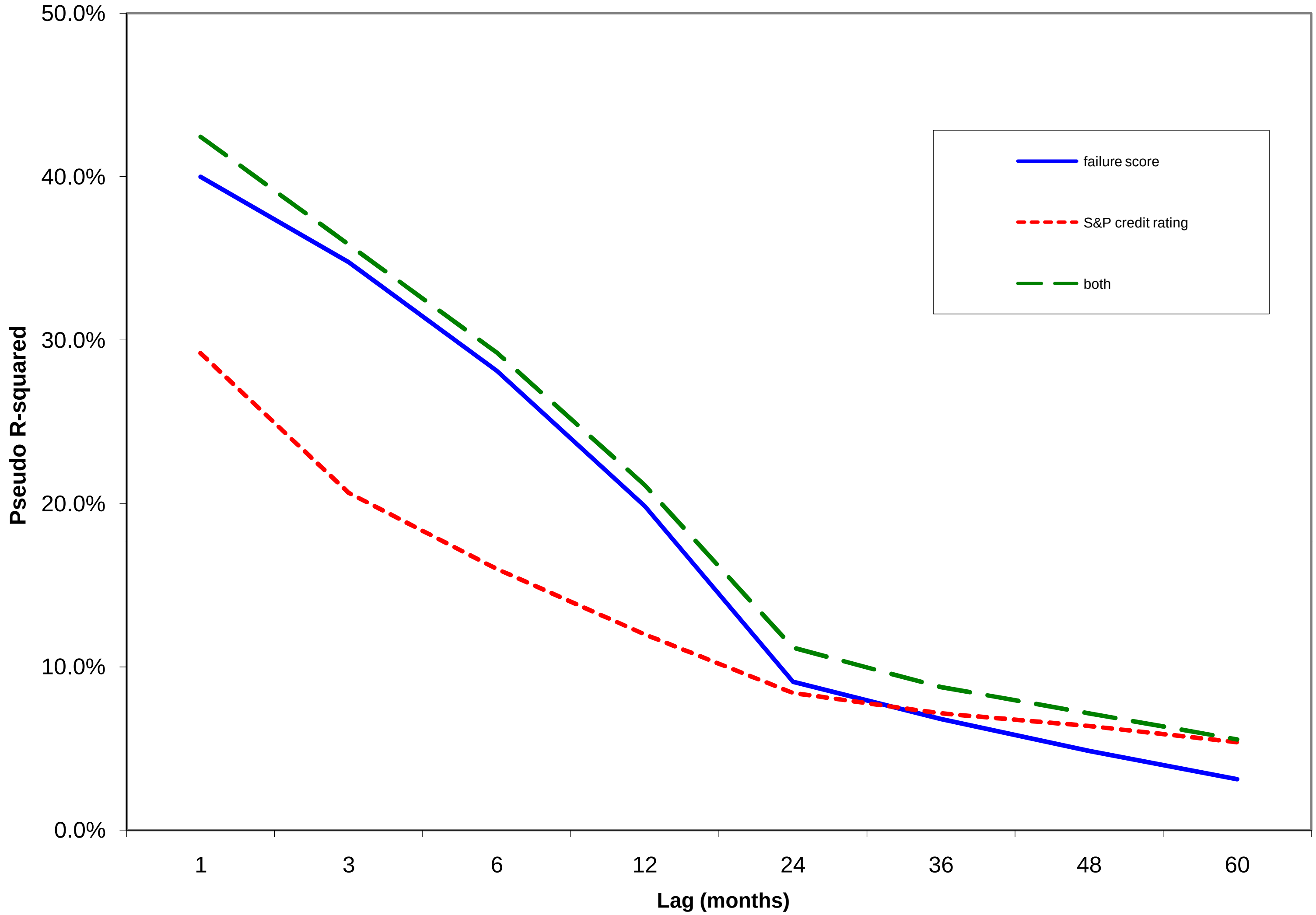
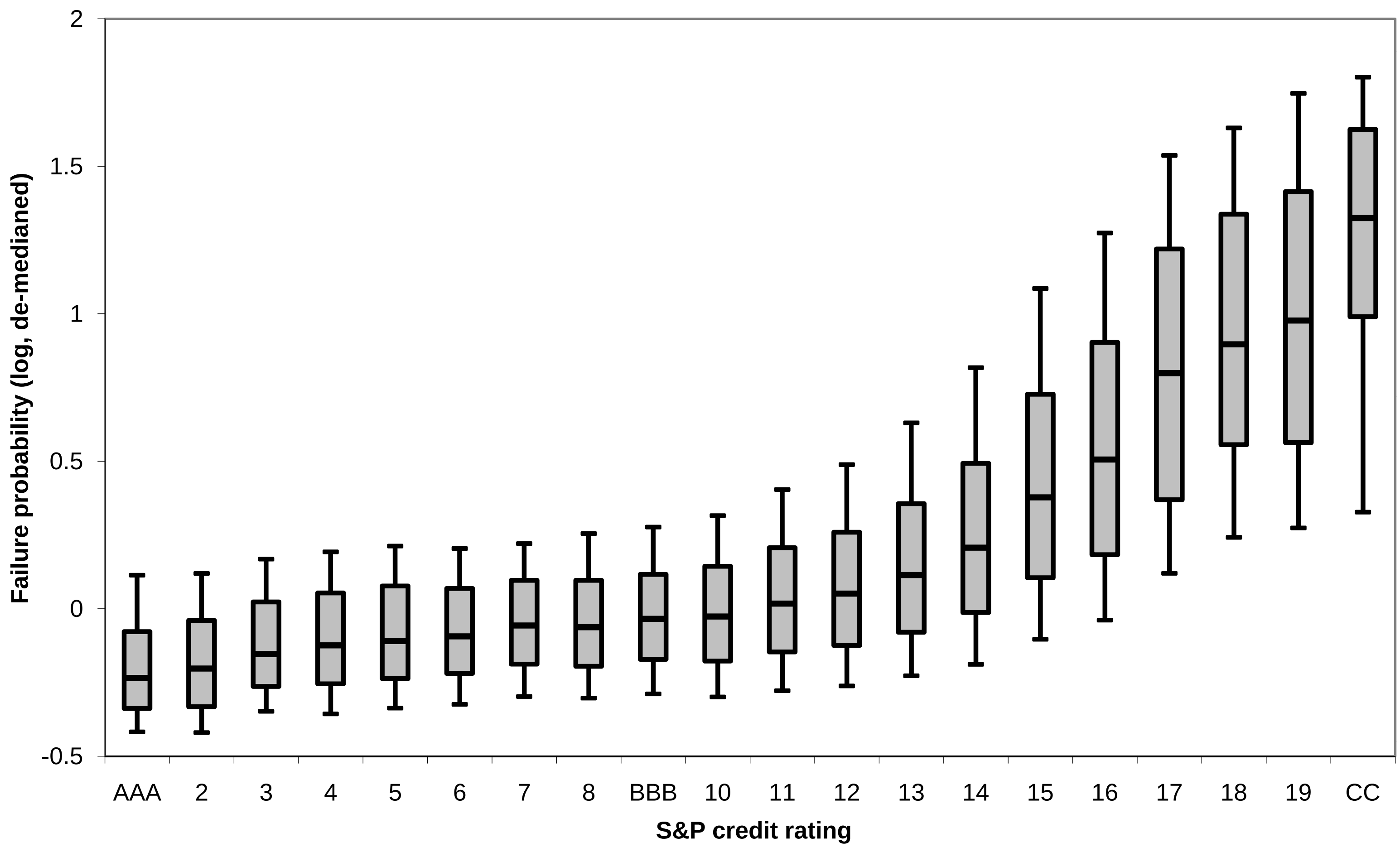


Figure 2: Failure probability distribution by S&P credit rating



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

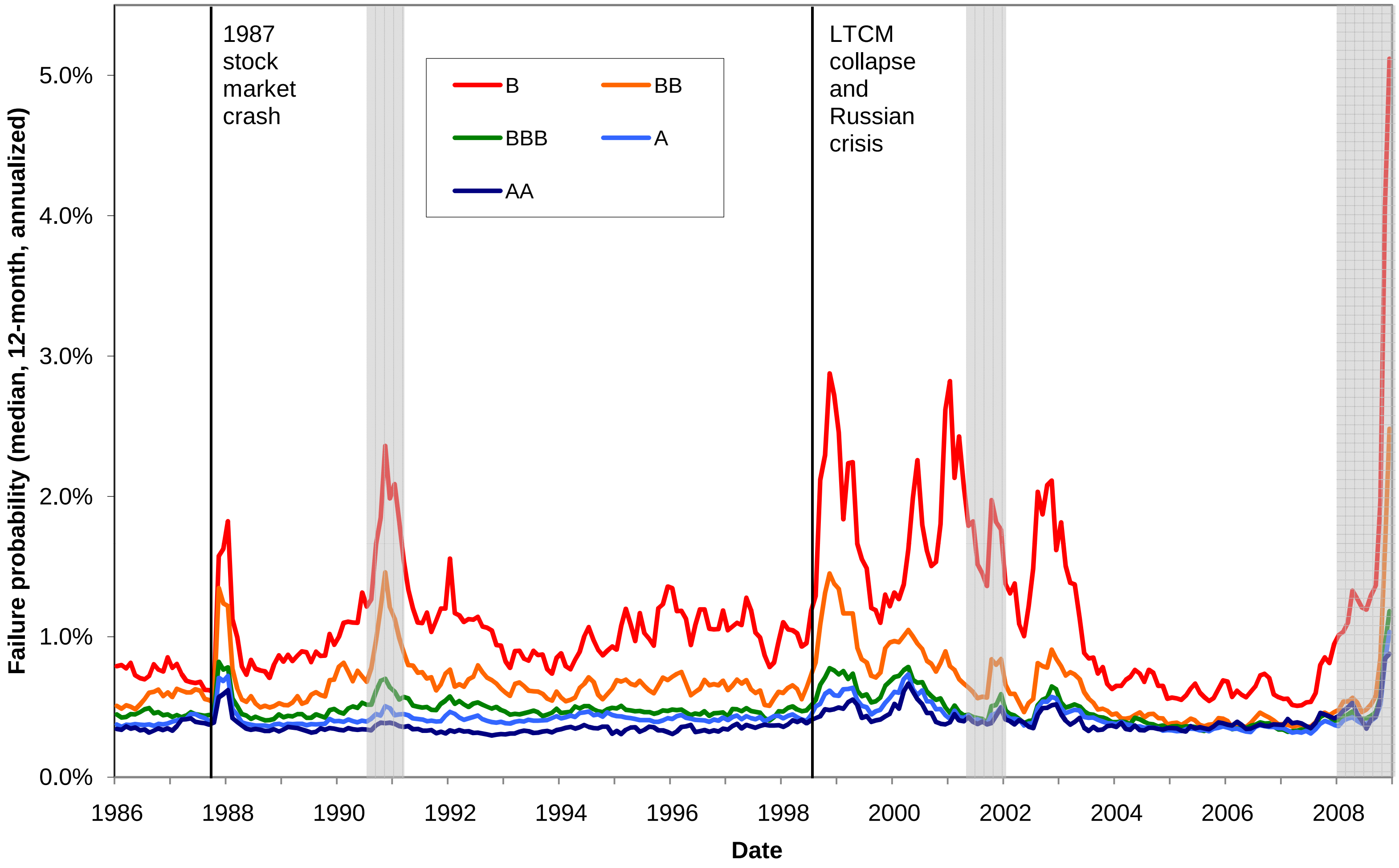


Figure 4: Share in different ratings

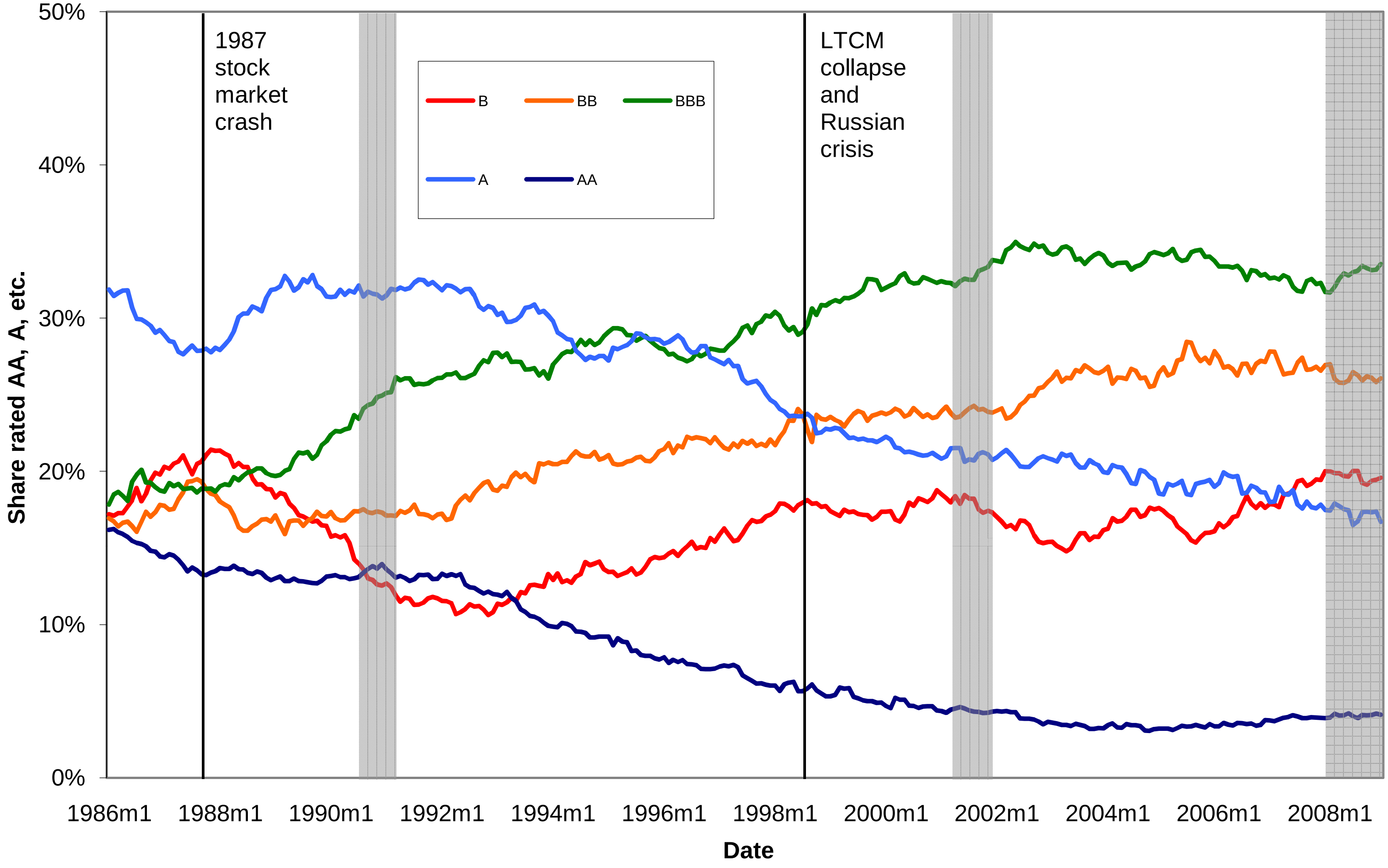
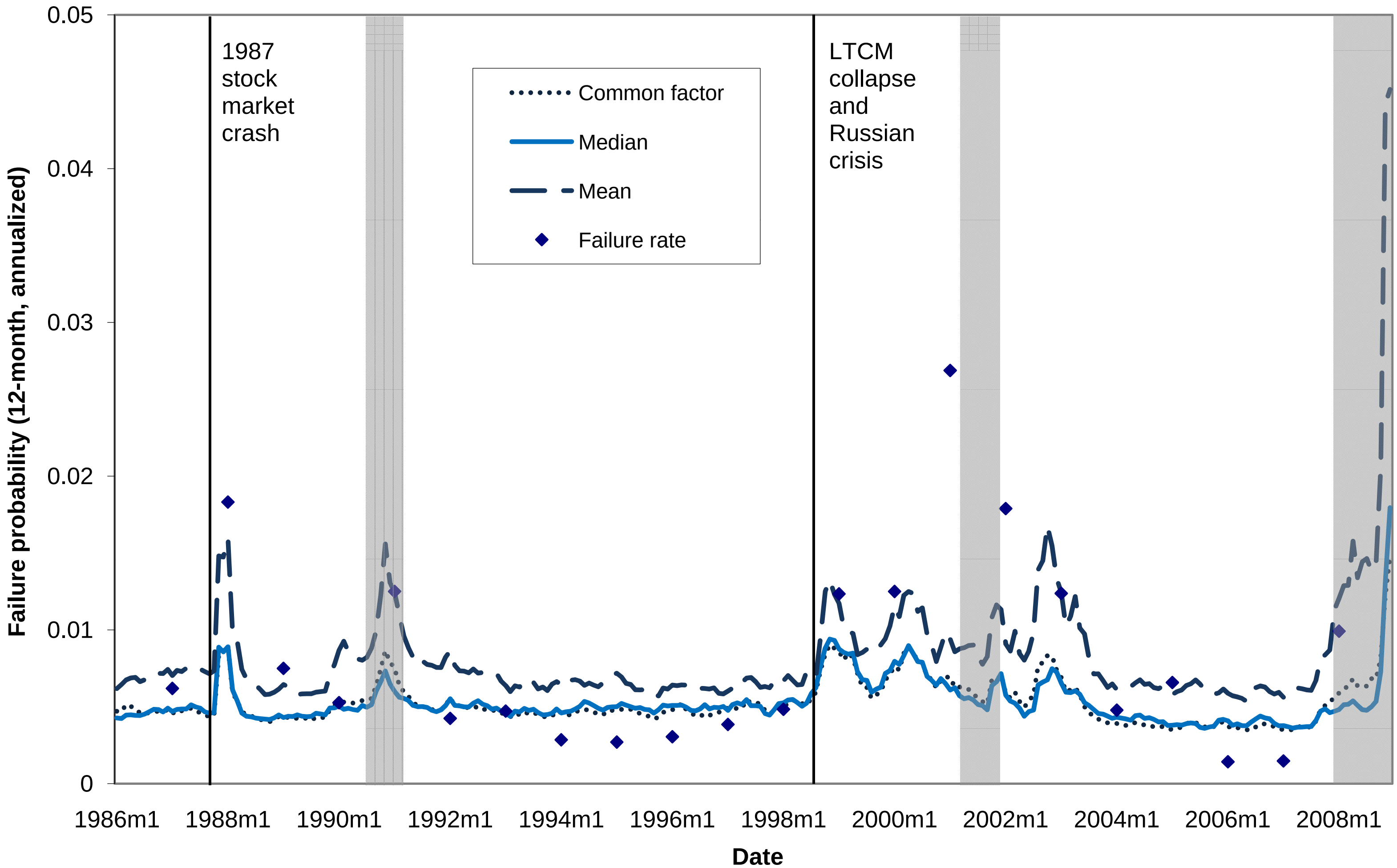
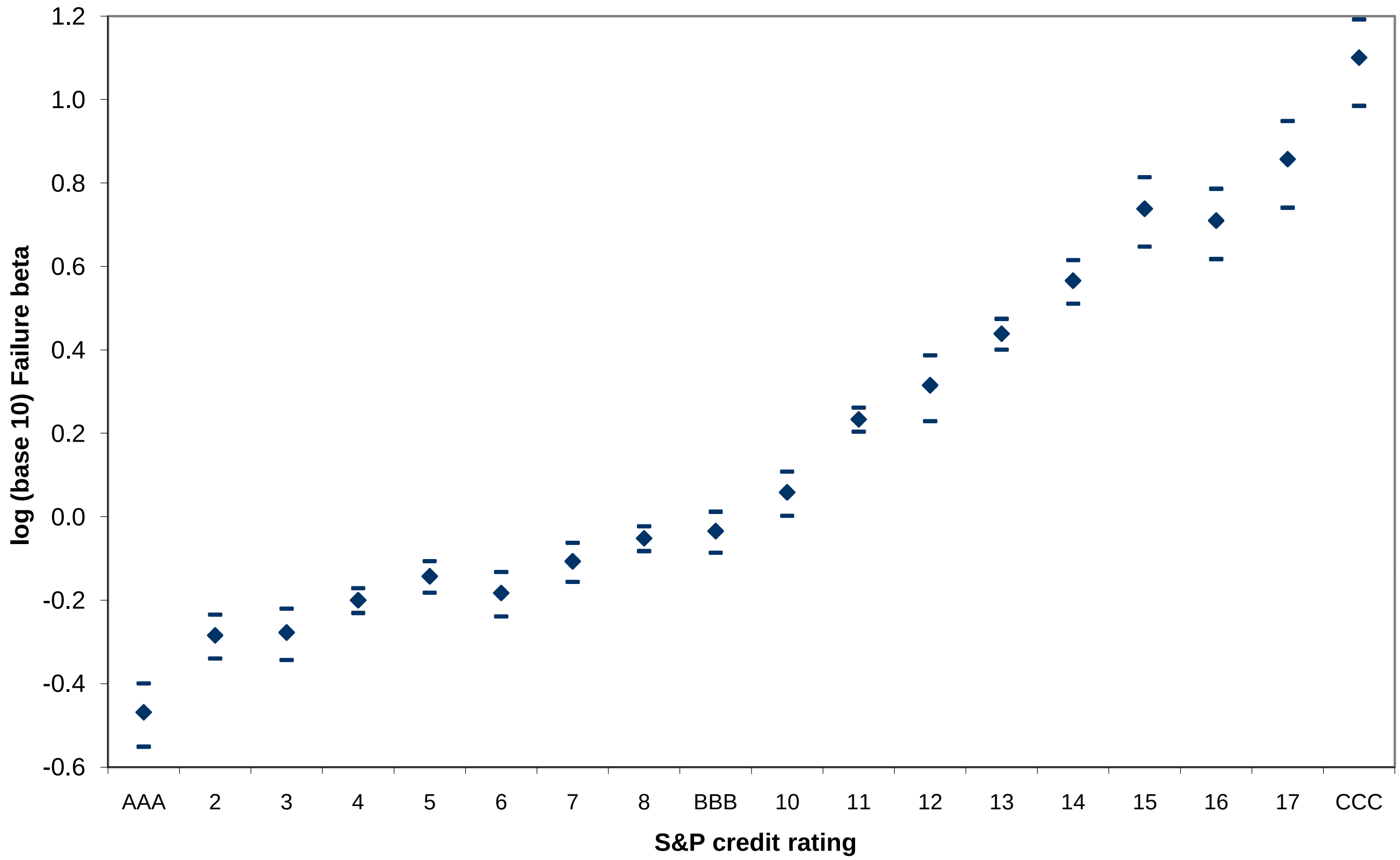


Figure 5: Average failure probability over time



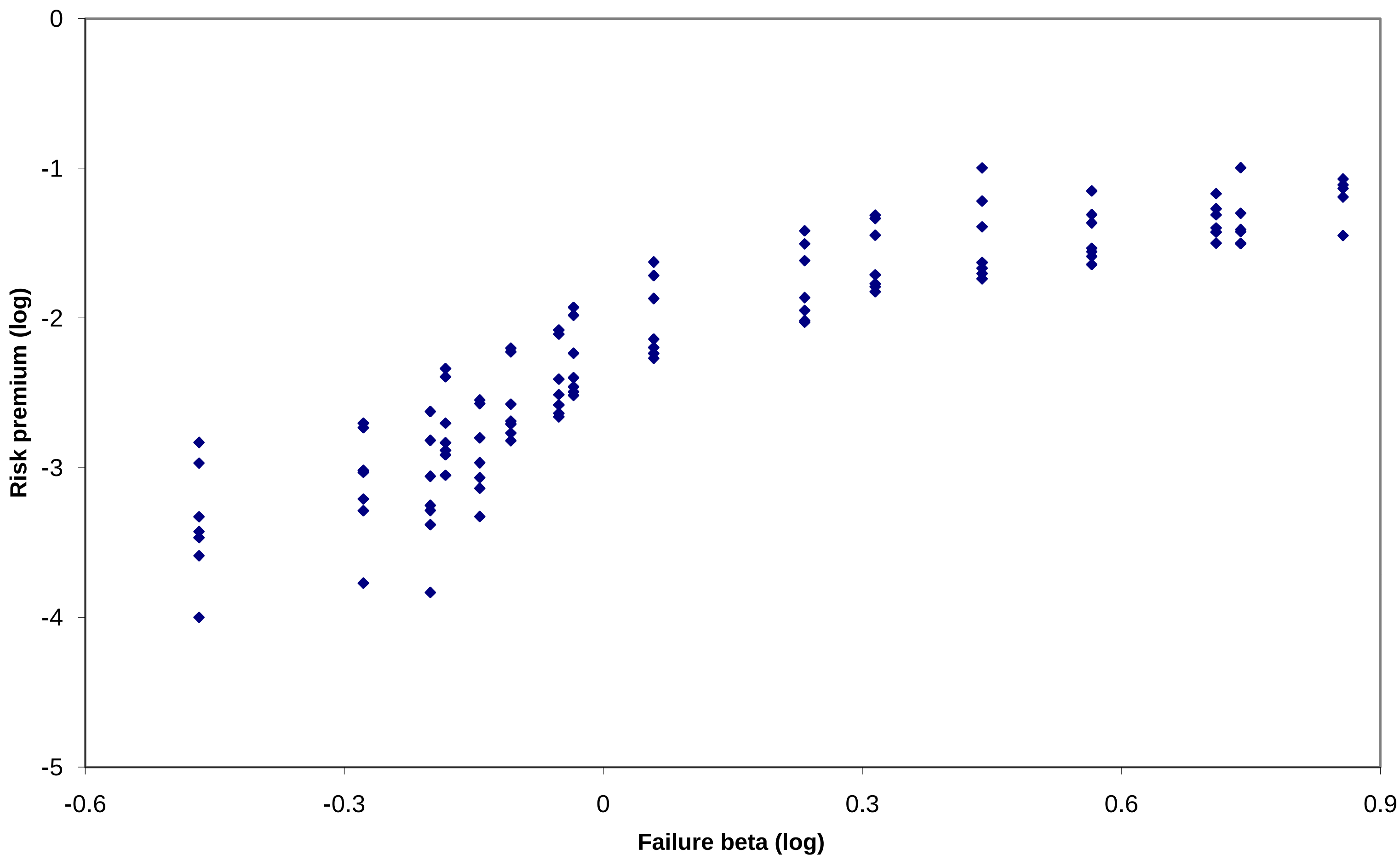
This figure plots the median,and total liability-weighted mean 12-month failure probability, as well as the common factor from 1986 to 2008. It also plots the annual empirical failure rate, which is normalized to the mean predicted failure probability for ease of comparison. The annual failure rate is reported 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.