The Credit Spread Puzzle - Myth or Reality? *

Peter Feldhütter†
London Business School

Stephen Schaefer‡
London Business School

September 30, 2014

Abstract

Many papers find that standard structural models predict corporate bond spreads that are too low compared to actual spreads, giving rise to the so-called credit spread puzzle. We show that the puzzle derives in large part from strong biases and low statistical power in commonly adopted approaches to testing the models. The biases are due to Jensen's inequality and arise when tests are carried out on a representative firm rather than on individual firms. Using data on individual firms during 2002-2012 we quantify the size of the bias in spread predictions and find it to be particularly severe for high-quality firms and short-maturity bonds. The problem of low statistical power arises because ex-post realized default frequency - often used to calibrate models - is a very poor estimate of ex-ante default probability. Finally, we test the Merton model via a bias-free approach using around 400,000 transactions in the period 2002-2012. We find that the Merton model captures both the average level and time series variation of the long-term BBB-AAA spread.

Keywords: Credit spread puzzle, Merton model, Structural models, Corporate bond spreads, Realized default frequencies;

JEL: C23; G01; G12

* We are grateful for valuable comments and suggestions received from Antje Berndt (discussant), Harjoat Bhamra, Hui Chen, Darrell Duffie, Stefano Giglio (discussant), Jean Helwege (discussant), Ralph Koijen, David Lando, Erik Loualiche (discussant), Gustavo Manso (discussant), Jens-Dick Nielsen, Lasse Heje Pedersen, Scott Richardson, and seminar participants at the WFA 2014, SFS Cavalcade 2014, ESSFM 2014 in Gerzensee, Fifth Risk Management Conference 2014, UNC's Roundtable 2013, AQR, London Business School, Vienna Graduate School of Finance, Stockholm School of Economics, Tilburg University, Duisenburg School of Finance Amsterdam, NHH Bergen, Rotterdam School of Management, University of Southern Denmark, Copenhagen Business School, Brunel University, Henley Business School, Bank of England, and University of Cambridge. We are particularly grateful for the research assistance provided by Ishita Sen.

†London Business School, Regent's Park, London, NW1 4SA, pfeldhutter@london.edu

‡London Business School, Regent's Park, London, NW1 4SA, sschaefer@london.edu

1. Introduction

Structural models were introduced by Merton (1974) and represent one of the most widely employed frameworks for the analysis of credit risk. However, many papers find that standard structural models predict spreads that are too low compared to actual spreads, giving rise to the so-called credit spread puzzle (see Table 1 for literature). We find that common approaches to testing structural models suffer from strong biases and low statistical power. When we test the Merton model in a bias-free approach we find that the Merton model captures both the average level and time series variation of the commonly studied long-term BBB-AAA spread.

One approach to testing structural models is to use *average* firm variables such as asset volatility and the leverage ratio as inputs to the model and then to compare the resulting model spread with the average actual spread (Leland (2006) and McQuade (2013)). This 'representative firm' calculation is typically carried out by averaging firm variables across firms within a rating category and over a long historical time period. However, a convexity bias arises in this case because the spread using average variables is typically lower than the average spread. Figure 1 illustrates this bias. The bias occurs both in the cross-section and in the time series dimension. David (2008) points out this bias but up to this point there is almost no empirical evidence on its size[1].

We examine the empirical importance of the convexity bias in our sample period 2002-2012 and find the bias to be strong. For example, the average model-implied spread on 3-year BBB-rated debt – averaging the model-implied spread across firms – is 234 basis points while the corresponding figure using a representative firm is 46 basis points. As a percentage of the spread, the bias increases as credit quality increases and as maturity shortens, precisely

[1] Bhamra, Kuehn, and Strebulaev (2010) examine the bias for 5- and 10-year spreads for BBB-rated firms in a structural model with macroeconomic risks. They do not examine the bias in a simple structural model or other rating categories.

in the directions where the credit spread puzzle is found to be most severe. This suggests that the conclusions in Leland (2006) and McQuade (2013) that standard structural models underpredict spreads may well be unwarranted. We show that a similar underprediction bias occurs when calculating default probabilities for a representative firm. This bias is also most severe at short maturities for high-quality issues, which may explain the finding in Leland (2004) and McQuade (2013) that standard structural models underpredict default probabilities at short horizons.

Another approach to structural model calibration is to compute the implied asset volatility that makes the default probability of a representative firm equal to the historical default frequency (Cremers, Driessen, and Maenhout (2008), Zhang, Zhou, and Zhu (2009), Chen, Collin-Dufresne, and Goldstein (2009), Chen (2010), and Huang and Huang (2012)). This typically results in an asset volatility that is too high because the default probability of a representative firm is lower than the average default probability. We show that the bias in asset volatility is often large; when calibrated to one-year default frequencies the implied asset volatility is around twice as high as average asset volatility in our sample. But since the yield spread of the representative firm is biased downwards relative to the average yield spread, it is not clear in which direction the model-implied yield spread will be biased relative to the average yield spread because two biases pull in opposite directions. We show that when the default probability and the computed spread have the same horizon, the Merton model-implied spread may be biased either upwards or downwards. For long-term BBB-rated bonds we find a downward bias of 15%. We also show that if the spread predictions are computed for shorter maturities than the horizon of default probabilities to which the model is calibrated (as in Cremers, Driessen, and Maenhout (2008) and Zhang, Zhou, and Zhu (2009)) the downward bias becomes much larger. We conclude that when fitting to historical default frequencies it is not meaningful to study the term structure of credit spreads because of large bias.

When calibrating to historical default frequencies, and no matter which structural model is being studied, the key assumption is that *ex-post* historical default frequencies provide a good proxy for *ex ante* average default probabilities. To test the statistical reliability of this assumption, we repeatedly simulate defaults over a period of 28 years for a large number of BBB-rated firms that have a realistic level of exposure to systematic risk. We find that the realized default frequency is often far from the average default probability. Furthermore, the distribution of default frequency is highly skewed to the right, so the observed default frequency - e.g., long-run data from rating agencies - most often observed is much lower than the mean. This implies that most of the time researchers using historical default frequencies will find a spurious credit spread puzzle even if it does not exist. We use the simulation results along with Bayes' rule and historical loss observations to calculate confidence bands for ex-ante default probabilities. Given the historical average 10-year default frequency of BBB-firms of 4.39% calculated over the period 1970-1998 - a number used in several studies - we estimate a 95% confidence band for the ex ante default probability ranging from 1.9% to 20.1%. We conclude that the ex-post realized default frequency cannot be used as a reliable measure of the ex-ante average default probability even when calculated over several decades.

Having documented that using a representative firm approach leads to biased spread predictions and fitting to historical default frequencies has low statistic power, we then test the Merton model using a bias-free approach (where we do not fit the model to historical default frequencies). The Merton model might not be the best structural model but it is certainly the simplest and we think it is useful to see how far we can go with a simple structural model. In other words, using the simplest model, we ask: "to what extent can the apparent failure of structural models be rectified by eliminating weaknesses in empirical implementation?"

In the bias-free approach we calculate a Merton spread for every transaction in the data set, compute an average, and compare with the average actual spread to the swap curve. Eom, Helwege, and Huang (2004) and Ericsson, Reneby, and Wang (2007) also use this approach. Eom, Helwege, and Huang (2004)'s data set consists of 182 trader quotes in the period 1986-1997 while Ericsson, Reneby, and Wang (2007)'s data set contains of 1387 transactions over the period 1994-2003. Our data set consists of 396,890 transactions for the period 2002-2012 from the TRACE database. This allows us to examine in much greater detail the ability of the Merton model to price bonds across maturity, across ratings, and over a time period that includes both a boom period and a major recession. In particular, we study the time series variation in detail.

The most common version of the credit spread puzzle is that the spread between long-term BBB yields and AAA yields is too high to be explained by the Merton model and other standard structural models. We find that over our sample period the median actual 10-year BBB-AAA spread is 88 basis points and the model-implied spread is 104 basis points, so on average the Merton model has no difficulty in predicting spreads at least as high as those in the data. There is a large amount of spread variation over the sample period, so we also look at the time series variation of spreads. We find that the model-implied BBB-AAA spread tracks the time series variation of the actual spread well. Based on the long-term BBB-AAA spread, we thus find no credit spread puzzle.

The Merton model predicts very low spreads for high-quality bonds with short maturity and "most researchers view this kind of result as a failure of diffusion-type structural models" according to Huang and Huang (2012). Even correcting for the convexity bias, we find that the Merton model mostly predicts spreads close to zero for investment grade bonds with a maturity less than one year. However, to our knowledge there is no evidence on the actual size of short-term spreads for bonds with a maturity below one year and we fill this gap in the literature[2]. Over the sample period the median actual spread to LIBOR for bonds with

[2]Covitz and Downing (2007) examine corporate spreads for maturities less than one year in the commercial paper market. The commercial paper market has its own set of institutional features compared to the corporate bond market. In addition, Ou, Hamilton, and Cantor (2004) find that "unlike bondholders, in most cases holders of defaulted Commercial Papers have ultimately been made whole" so recovery rates are

a maturity below one year is also close to zero for bonds rated AAA, AA, or A. In 2008-2009 short-term spreads were higher than zero, consistent with Dick-Nielsen, Feldhütter, and Lando (2012) finding an illiquidity premium during this period, but by 2010 spreads were back at a level close to zero. Overall, we conclude that the low short-maturity spreads in the Merton model align well with actual spreads for ratings A-AAA. However, the actual median short-term BBB spread is 55 basis points while the median Merton model spread is zero basis points. This suggests that a credit spread puzzle in short-term bonds appears when *moving down* in credit quality of investment grade firms.

For high-quality bonds the Merton model also predicts small spreads on long-term bonds. For our sample the median model-implied 10-year AAA spread is zero basis points while the median actual spread is 12 basis points. For A-rated bonds the actual spread is 67 basis points while the model-implied spread is 29 basis points. This shows that the Merton model cannot match the size of high quality long-term spreads, but the magnitude of underprediction is smaller than previously found. So even though we find a puzzle for high-quality long-term spreads, the puzzle is smaller compared to previous findings in terms of spread size and how far down in credit quality the puzzle extends.

In an Internet Appendix we show that are results are robust. In particular, we show that results are similar when using the Leland-Toft model (Leland and Toft (1996)) instead of the Merton model. Results are also similar when using quotes from Merrill Lynch for the period 1997-2012 instead of TRACE transactions data for the period 2012-2012.

The organization of the article is as follows: Section 2 explains the data and how the Merton model is implemented. Section 3 examines common approaches to testing structural models. Although this section uses data described in Section 2, one can easily skip section 2 and read section 3 directly. In Section 4 we examine the statistical properties of historical default frequencies. Section 5 tests the Merton model using transaction data for the period

likely different.

2002-2012, while Section 6 describes the robustness checks in an internet Appendix. Section 7 concludes.

2. The Merton model: basics and implementation

2.1 Data

This section gives a brief overview of the data; a detailed description is relegated to Appendix A.

Since July 1, 2002, members of the Financial Industry Regulatory Authority (FINRA) have been required to report their secondary market over-the-counter corporate bond transactions through the TRACE database. Our data comes from two sources, WRDS and FINRA, and covers almost all U.S. Corporate bond transactions for the period July 1, 2002 - June 30, 2012[3]. We limit the sample to senior unsecured fixed rate or zero coupon bonds and exclude bonds that are callable, convertible, putable, perpetual, foreign denominated, Yankee, have sinking fund provisions, or have covenants. Appendix A.1 describes details of the dataset and explains why it has more transactions than the typical TRACE dataset used in the literature.

To price a bond in the Merton model we need the issuing firm's asset volatility, leverage ratio, and payout ratio. *Leverage ratio* is calculated as the book value of debt divided by firm value (where firm value is calculated as book value of debt plus market value of equity). *Payout ratio* is calculated as the sum of interest payments to debt, dividend payments to equity, and net stock repurchases divided by firm value. *Asset volatility* is not directly observable and is estimated from equity volatility and leverage as we will explain in Section

[3]Rule 144A bond transactions are not covered. Rule 144A allows for private resale of certain restricted securities to qualified institutional buyers. According to TRACE Fact Book 2011, the percent of rule 144A transactions relative to all transactions is 2.0% in investment grade bonds and 8.4% in speculative grade bonds. Also, transactions reported on or through an exchange are not included in TRACE.

2.2. Equity volatility is an annualized estimate based on the previous three year's of daily equity returns. All firm variables are obtained from CRSP and Compustat and details are given in Appendix A.2.

2.2 Calibration of the Merton model

Asset value in the Merton model follows a Geometric Brownian Motion under the risk-neutral measure,

$$\frac{dV_t}{V_t} = (r - \delta)dt + \sigma_A dW_t \tag{1}$$

where r is the riskfree rate, δ is the payout rate, and σ_A is the volatility of asset value. The firm is financed by equity and a single zero-coupon bond with face value F and maturity T[4]. If the asset value falls below the face value at the bond's maturity, $V_T < F$, the firm cannot repay its bond holders and the firm defaults. In our calibration we try to be consistent with previous literature on the credit spread puzzle, and key calibrated parameters in previous literature are given in Table 1. In the original Merton model bondholders receive 100% of the firm's value in default, but to be consistent with empirical recovery rates, we follow the literature on structural models and assume that bondholders recover only a fraction of the firm's value in default. According to Moody's (2011) the average recovery rate for senior unsecured bonds for the period 1987-2010 was 49.2% and we follow Eom, Helwege, and

[4]To keep the model as simple as possible, we assume that the bond is a zero-coupon bond and implicitly account for coupons by estimating the payout rate as the total payout to debt and equity holders. An alternative would be to assume that payout is only dividends to equity holders and the firm refinances coupons and repays them at bond maturity. In this case the drift of the firm would be higher, but the amount of debt would also be higher and these two effects offset each other and the model is the same as the one we present. Finally, we could - at the expense of simplicity - allow for coupon payments as in Eom, Helwege, and Huang (2004). This generally leads to higher spreads because the firm can default not only at bond maturity but also before bond maturity.

Huang (2004) and set the payoff to bondholders to $\min(V_T, 0.492F)$.[5] The bond price at time 0 is calculated as:

$$P(0,T) = E^Q[e^{-rT}(1_{\{V_T \geq F\}} + \min(0.492, \frac{V_T}{F})1_{\{V_T < F\}})],$$

and the specific expression is given in Appendix B. The model implies a deadweight cost of bankruptcy; for a 10-year A-rated bond the expected deadweight cost of bankruptcy is 32.6%[6]. This is broadly consistent with the empirical estimate of 31.0% in Davydenko, Strebulaev, and Zhao (2012), 36.5% in Alderson and Betker (1995), 45.5% in Gilson (1997), 45% in Glover (2012), and the deadweight cost generally used in the literature as Table 1 shows.

Our assumption that the default boundary is the face value of debt is common (Eom, Helwege, and Huang (2004) and Cremers, Driessen, and Maenhout (2008)) although some assume a default boundary below (Chen, Collin-Dufresne, and Goldstein (2009) and Huang and Huang (2012)). Huang and Huang (2012) show that if one matches the model to historical default frequencies and recovery rates as in Cremers, Driessen, and Maenhout (2008), Huang and Huang (2012), and Chen, Collin-Dufresne, and Goldstein (2009), the value of the default boundary is not very important for implied spreads. However, in our implementation we do not use historical default frequencies as targets and therefore the value of the default boundary is important. If we chose a boundary below face value, say 60% of face value as in Huang and Huang (2012), default probabilities and spreads go down

[5] The average recovery rate in the model is slightly below 49.2% of face value because the recovery is either 49.2% or less. However, the firm value will rarely fall below 49.2% and therefore the expected recovery will be below but close to 49.2%. With an asset risk premium of 4% and using the mean values of leverage, asset volatility, and payout ratio for an A-rated bond in Table 3, the expected recovery on a 10-year bond is 48.3% (see Appendix B.2 for formulas).

[6] This assumes an asset risk premium of 4% and uses the mean values of leverage, asset volatility, and payout ratio for a A-rated bond in Table 3. The formula is given in Appendix B.2. An asset premium of 3% respectively 5% gives an expected deadweight cost of bankruptcy of 31.8% respectively 33.4%.

substantially. However, if one accepts the above mentioned recent evidence of a 31-45% deadweight cost of bankrupcy, one cannot match historical recovery values of close to 50% without having a default boundary close to the face value of debt.[7]

When we price a given bond we assume that all debt matures at bond maturity even if the firm has multiple bonds outstanding. This is an aggressive assumption for short maturity bonds and results in higher default probabilities compared to a structural model that incorporates the term structure of debt. Consistent with this view Moody's KMV use a default boundary equal to short-term liabilities plus one-half of long-term liabilities when calculating one-year default probabilities, see Sun, Munves, and Hamilton (2012). For longer maturity bonds this assumption is more reasonable because most of the existing debt will have matured when the bond matures. In fact, this assumption might lead to conservative estimates of default for long-maturity bonds because the firm cannot default before bond maturity in contrast to a model with a term structure of debt. In line with this reasoning Moody's KMV increase the default boundary with the horizon at which they calculate default probabilities (Sun, Munves, and Hamilton (2012)). We stress that we use the Merton model because it is the most simple and heavily tested structural model in the literature, not because it has the most realistic assumptions. The Leland-Toft model is a structural model where the debt matures at a constant rate, and in the Internet Appendix we show that results are similar when using the Leland-Toft model instead of the Merton model.

A crucial parameter in any structural model is the volatility of assets and we follow the approach of Schaefer and Strebulaev (2008) in calculating asset volatility. Since firm value is the sum of the debt and equity values, asset volatility is given by:

$$\sigma_A^2 = (1-L)^2\sigma_E^2 + L^2\sigma_D^2 + 2L(1-L)\sigma_{ED}, \tag{2}$$

where σ_A is the volatility of assets, σ_D volatility of debt, σ_{ED} the covariance between the

[7]See also Eom, Helwege, and Huang (2004) for a discussion of the importance of the deadweight cost of bankruptcy.

returns on debt and equity, and L is leverage ratio. If we assume that debt volatility is zero, asset volatility reduces to $\sigma_A = (1-L)\sigma_E$. This is a lower bound on asset volatility. Schaefer and Strebulaev (2008) (SS) compute this lower bound along with an estimate of asset volatility that implements equation (2) in full. They find that for investment grade companies the two estimates of asset volatility are very similar while for junk bonds there is a substantial difference. We compute the lower bound of asset volatility, $(1-L)\sigma_E$, and multiply this lower bound with SS's estimate of the ratio of asset volatility computed from equation (2) to the lower bound. Specifically, we estimate $(1-L)\sigma_E$ and multiply this by 1 if $L < 0.25$, 1.05 if $0.25 < L \leq 0.35$, 1.10 if $0.35 < L \leq 0.45$, 1.20 if $0.45 < L \leq 0.55$, 1.40 if $0.55 < L \leq 0.75$, and 1.80 if $L > 0.75$[8]. This method has the advantage of being transparent and easy to replicate.

Finally, the riskfree rate r is chosen to be the swap rate for the same maturity as the bond. For maturities shorter than one year we use LIBOR rates. Most of the previous literature uses Treasury yields as riskfree rates as Table 1 shows, but recent evidence shows that swap rates are more appropriate to use than Treasury yields. The reason is that Treasury bonds enjoy a convenience yield that pushes their yields below riskfree rates. Hull, Predescu, and White (2004) find that the riskfree rate is 63bps higher than Treasury yields and 7bps lower than swap rates, Feldhütter and Lando (2008) find riskfree rates to be approximately 53bps higher than Treasury yields and 8bps lower than swap rates, while Krishnamurthy and Vissing-Jorgensen (2012) find that riskfree rates are on average 73bps higher than Treasury yields.

[8]These fractions are based on Table 7 in SS apart from 1.80 which we have set somewhat ad hoc. Results are insensitive to other choices of values for $L > 0.75$. Note that we rely on estimates from SS instead of applying their procedure because the nature of our dataset is different from theirs. Their dataset consists of monthly quotes and therefore they have monthly return observations for every bond. Our dataset consist of actual transactions which are unevenly spaced in time and constructing a return series is considerably more difficult. See also Correia, Kang, and Richardson (2014) for an assesment of different approaches to calculating asset volatility.

2.3 Summary statistics

Table 2 shows statistics for the whole TRACE data set, for the final sample, and how the sample reduces as we apply more criteria. We see that the final sample of straight industrial bonds with no option-like features or covenants is a small subset of the universe of TRACE bonds. The final sample consists of 389 bonds[9]. Summary statistics for the firms in the final investment grade sample are shown Table 3 and those for the final speculative grade sample in Table 4. A firm can be in more than one rating category if the rating changes over time. We combine AAA and AA into one rating group because there are only four AAA-rated firms in our sample. When we refer to AAA-rated bonds later in the paper, we are referring to the combined group of bonds rated AAA or AA.

Focusing on investment grade firms, we see in Table 3 that the average leverage ratio in the sample period is 0.11 for AAA/AA, 0.24 for A, and 0.41 for BBB. The leverage ratios are similar to those found in other papers[10]. The payout ratio is similar across ratings, with average payout ratio ranging from 4.8% to 6.0%. Average equity volatilities for A-AAA rated firms are very similar to the estimates in Schaefer and Strebulaev (2008), while average equity volatility for BBB firms of 0.45 is higher than 0.33 estimated in Schaefer and Strebulaev (2008)[11]. Very interestingly, asset volatility is much the same in the second

[9]Papers reporting number of bonds in their sample are Eom, Helwege, and Huang (2004) with 182 bonds, Cremers, Driessen, and Maenhout (2008) with 524 bonds, and Schaefer and Strebulaev (2008) with 1360 bonds. If we allowed for covenants (but excluded callable bonds) as in Schaefer and Strebulaev (2008) the number of bonds in the sample would be 1297. If we allowed for covenants and call features as in Cremers, Driessen, and Maenhout (2008) the number of bonds in the sample would be 5349. The Internet Appendix examines bonds with covenants and call options in detail.

[10]Huang and Huang (2012) use a leverage ratio of 0.13 for AAA, 0.21 for AA, 0.32 for A, and 0.43 for BBB while Schaefer and Strebulaev (2008) find an average leverage of 0.10 for AAA, 0.21 for AA, 0.32 for A, and 0.37 for BBB.

[11]The sample period is 1996-2003 in Schaefer and Strebulaev (2008) while our sample period is 2002-2012, so the two sample periods only have a small overlap. In an Internet Appendix we use the Merrill Lynch data

half of our sample (2007Q3 to 2012Q4), which includes the recent crisis, as the first half (2002Q3 to 2007Q2). This is the case even for BBB that sees a strong increase in average equity volatility from 40% to 58%. This suggests that the high equity volatility during the subprime crisis was caused primarily through increased leverage and not because of increased asset volatility.

In Table 5 we see that the number of speculative grade bonds in our sample is small relative to the number of investment grade bonds. The reason is that speculative grade bonds frequently contain call and covenant features which leads to their exclusion from our sample. Since the number of speculative grade firm and bond observations in the sample is small relative to the number of observations in the investment grade segment, we group all speculative grade bonds in one group and generally focus on investment grade bonds. This does not substantially limit our examination of the credit spread puzzle since the puzzle is mostly confined to investment grade bonds. For example, HH report that between 16% and 29% of the 10-year spread on investment grade bonds can be explained by credit risk while the corresponding range for speculative grade bonds is 60-83%.

3 Why most existing tests of structural models are biased

A number of papers find that standard structural models - often the Merton model - cannot match the level of credit spreads, particularly for short maturities and high credit quality issuers (see Table 1). This finding has been coined the credit spread puzzle and the standard reference for the puzzle is Huang and Huang (2012). A recent review of the inability of the Merton model to capture the level of credit spreads and extensions of the model is

used in Schaefer and Strebulaev (2008) and for the period 1997-2007 we find an average equity volatility of 0.36, close to 0.33.

Sundaresan (2013). While credit spreads might be influenced by bond illiquidity and other factors, default probabilities are less so[12]. Leland (2004) and McQuade (2013) find that the Merton model underpredicts default probabilities at short horizons consistent with the underprediction evidence for spreads.

In this Section we discuss the methods used in these papers to test structural models and show that common tests are biased and suffer from low statistical power.

3.1 Notation

We begin by defining notation. For a given corporate bond transaction at time t in a bond with maturity T issued by firm i we call the actual spread $s_T^A(i,t)$. To calculate the theoretical credit spread in the Merton model we need firm leverage L_{it}, asset volatility σ_{Ait}, payout ratio δ_{it}, and the riskfree rate for maturity T r_{Tt}. We denote the vector of parameters for firm i at time t by $\theta_{it} = (L_{it}, \sigma_{Ait}, \delta_{it}, r_{Tt})$[13]. We define the parameter vector without asset volatility as $\theta_{it}^{\setminus\sigma_A} = (L_{it}, \delta_{it}, r_{Tt})$. We denote the model-implied Merton credit spread as $s_T^M(\theta_{it})$ and the explicit formulas are given in equations (9) and (12) in the Appendix. Given an asset risk premium π_A Appendix B.2 reviews the calculation of the cumulative default probability over the next T years and we call this $PD_T(\theta_{it}, \pi_A)$. Often we suppress the dependence on π_A. The asset risk premium can be firm specific, but consistent with previous literature we assume that it is the same for all firms.

[12]See He and Milbradt (2013) for a structural model where there is an interaction between default and liquidity risks.

[13]Payout ratio, asset volatility, and the risk free rate are assumed to be constant in the Merton model, but in the implementation we estimate them day by day and therefore they might vary over time for a given firm. This approach is standard in the literature (see for example Eom, Helwege, and Huang (2004), Schaefer and Strebulaev (2008), and Bao and Pan (2013)) and since we explicitly study previous results in the literature we follow this tradition.

3.2 Convexity bias in spreads

Leland (2006) and McQuade (2013) use values of the leverage ratio, payout rate, the riskfree rate, and asset volatility (obtained from Schaefer and Strebulaev (2008)) averaged over time and firms to calculate model-implied credit spreads from a standard structural model and then compare these with historical averages of actual credit spreads. We call this approach the representative firm approach. Following the tradition of the literature they do this for individual rating classes and in the rest of the paper we assume that all such comparisons are done within a rating class without explicitly mentioning this.

In the representative firm approach

$$s_T^M(\frac{1}{T_{i,t}}\sum_{i,t}\theta_{it})$$

is compared with

$$\frac{1}{T_{i,t}}\sum_{i,t}s_T^A(i,t)$$

where $T_{i,t}$ is the number of spread observations for bonds with maturity T over the corresponding time period. The approach suffers from a Jensen's inequality bias because spreads are typically convex in firm variables. Figure 1 illustrates the bias. David (2008) discusses this bias but there are no empirical results on the severity of the bias in a standard structural model. We examine the bias by computing both the average Merton spread and the Merton spread for the average firm in our sample.

The average Merton spread is calculated by computing the Merton spread for each transaction in the sample, computing the volume-weighted average monthly spread, and taking a simple average over the monthly spreads. Large bond issues trade more often than small bond issues and by volume-weighting we obtain a spread that is plausibly close to the spread of a portfolio holding the size-weighted bond market, consistent with how bond indices are calculated. The result is $E[s(\theta_{it})]$ in Table 6. To calculate the Merton spread using the representative firm approach, we calculate average firm variables by averaging monthly average volume-weighted firm variables and calculate the spread using average firm variables. The result is $s(E[\theta_{it}])$ in Table 6 where we also compute the ratio $\frac{s(E[\theta_{it}])}{E[s(\theta_{it})]}$. If the ratio is below one the representative firm approach underestimates spreads.

We see in the table that the bias is large and increases with credit quality and as maturity shortens. Averaging over the whole sample period 2002-2012, spreads based on average firm variables are only 5% of the average model-implied spread for 10-year AAA rated bonds while for BBB-rated bonds it is 85%[14]. For bonds with a maturity of less than one year spreads based on both the representative firm approach are essentially zero for investment grade bonds while spreads using the individual firm approach are as high as 100 basis points. We also see that the bias as a ratio is similar in the first (2002-2007) and second half (2007-2012) of the sample period despite the large differences in spreads between the two periods.

Overall, the results show that the representative firm approach causes a downward bias in spreads for investment grade bonds and this bias becomes larger where the credit spread puzzle has been found to be most severe, for short maturities and high-quality firms. This calls into question the findings in Leland (2006) and McQuade (2013) who, using this approach, conclude that the Merton model underpredict spreads[15].

The analysis in Table 6 takes into account correlation between model parameters. For example, the correlation between leverage ratio and asset volatility is -0.23 (not reported), so firms with high leverage tend to have low asset volatility. Nevertheless, it is useful to examine the contribution of each model parameter one at a time. We do so for each parameter by calculating the spreads using the 10% and 90% quantile of that parameter and comparing this average spread with the spread calculated using the average value of the 10% and 90%

[14]We use 'long-term' and '10-year' interchangeably since the average maturities in the groups of long-term bonds are fairly close to 10 year.

[15]One might worry that since there is variation in recovery rates, using a constant recovery rate is another source of a convexity bias. However, spreads are almost linear in recovery rates and therefore assuming a constant recovery rate is unlikely to have a significant effect.

quantile. In the calculation we set the remaining parameters to their average. Table 7 shows the results for bonds of different maturities. We see that payout rate has a modest contribution to the convexity effect and interestingly it is almost independent of bond maturity. The contribution of leverage ratio and asset volatility increases as bond maturity decreases and for long maturities the contribution of asset volatility is strongest while for short maturities the contribution of leverage ratio is strongest.

3.3 Convexity bias in default probabilities

Leland (2004), Zhang, Zhou, and Zhu (2009), and McQuade (2013) examine default probabilities implied by structural models. The advantage of looking at default probabilities instead of spreads is that default probabilities are arguably not contaminated by bond illiquidity, recovery rates, and other potential factors influencing spreads, but, unlike spreads, we don't know what they are (as we point out later). They use the representative firm approach and compute model-implied default probabilities, using the leverage ratio, payout rate, and the riskfree rate based on historical averages, and an asset risk premium of 4%. All three studies find that model-implied default probabilities are below historical default frequencies at horizons shorter than five years.

However, Table 8 shows that a similar bias for default probabilities as for spreads occurs when using the representative firm approach; average default probabilities are typically substantially higher than the default probability of the average firm and the bias increases in credit quality and as maturity shortens. Therefore, default probabilities calculated using a representative firm cannot be used to assess structural models' ability to match default frequencies.

3.4 Calibrating to historical default rates

The most common approach when calibrating structural models is to leave asset volatility as a free parameter, set the leverage ratio and payout ratio to their historical averages, and imply out asset volatility by matching the default probability of an average firm to historical default frequency. Table 1 gives an overview of papers using this approach. Specifically, for a given bond maturity T (and rating), the asset volatility $\hat{\sigma}_A$ is implied out by the equation

$$PD_T(\frac{1}{T_{i,t}}\sum_{i,t}\theta_{it}^{\setminus\sigma_A},\hat{\sigma}_A) = RD_T \tag{3}$$

where RD_T is the historical realized average default frequency over a long period (in Huang and Huang (2012)'s case, 28 years). The backed-out asset volatility $\hat{\sigma}_A$ is then used to calculate the model-implied spread

$$s_T^M(\frac{1}{T_{i,t}}\sum_{i,t}\theta_{it}^{\setminus\sigma_A},\hat{\sigma}_A) \tag{4}$$

and this is compared with average actual spreads $\frac{1}{T_{i,t}}\sum_{i,t}s_T^A(i,t)$.

This approach leads to a biased calibration. As shown in Table 8 default probability using average firm variables is lower than average default probability. Therefore, $\hat{\sigma}_A$ implied out in equation (3) will be higher than average σ_A to compensate for this. However, spreads using average firm variables are lower than average spreads as Table 6 shows. Since the two biases have opposite effects on spreads, the combined effect on the predicted spread in equation (4) is not clear.

We examine the extent of the bias in Table 9. We set the asset risk premium to 4% and calculate for each transaction the default probability over the same horizon as the bond. For different bond maturities and ratings we calculate the monthly volume-weighted average default probability and then average across months. This is E(Def Prob). We use E(Def Prob) in place of historical default frequency RD_T in equation (3). This means that here we assume that average default probabilities equal historical default frequencies, an assumption that we investigate in Section 4. Likewise we calculate average asset volatility which is $E(\sigma_A)$ in Panel A. We also calculate average leverage ratio, average payout ratio, average riskfree rate and use these for a representative firm. We then imply out asset volatility - $\hat{\sigma}_A$ - by making the default probability of the representative firm match E(Def Prob) according to equation (3). The calibrated asset volatility is shown in Panel A. We see that for investment grade bonds asset volatility is biased upwards and the bias becomes more severe as bond maturity shortens. In Panel B we see that predicted spreads using $\hat{\sigma}_A$ in equation (4) can be higher or lower than the average spread. AAA spreads are 27 % higher, while for example long-term BBB spreads are 16% lower. Note that in this Panel we predict spreads at the same horizon as the horizon to which we fit default probabilities.

Since asset volatility is biased upwards other model predictions are conceivably biased. And even though the biases partially cancel out in spread predictions when default rates and spreads have the same horizon as in Panel B, they do not cancel out at all when default rates and spreads have different time horizons. For example, Cremers, Driessen, and Maenhout (2008) fit to historical 10-year historical default frequencies and look at implied spreads and default probabilities for maturities between one and 10 years while Zhang, Zhou, and Zhu (2009) fit to historical 5-year historical default frequencies and look at implied spreads and default probabilities for maturities between one and five years. To illustrate the bias when using different horizons for spreads and default probabilities, we repeat the previous procedure of implying out $\hat{\sigma}_A$ by calibrating to average default probabilities and predicting spreads except that we now use the implied asset volatility from fitting to 10-year default probabilities when calculating spreads for bonds of shorter maturity. Panel C shows that the bias is much more severe than in Panel B, and for short and medium term bonds model-implied spreads are hugely underestimated relative to average spreads. Put differently, it is not meaningful to calculate a term structure of credit spreads for a representative firm and compare it with the average actual term structure of spreads.

4 Why default probabilities are hard to estimate

Most papers testing structural models use ex post realized default frequencies to proxy for ex ante default probabilities.[16] We show in this section that this is highly problematic because the variance of the estimator is large. Note that what we show in this section is unrelated to the convexity bias documented in the previous section.

There is a tradition in the literature to use realized default frequencies published by Moody's.[17] To explain how Moody's calculate default frequencies, let us consider the 10-year BBB cumulative default frequency of 4.39% used in Cremers, Driessen, and Maenhout (2008) and Huang and Huang (2012). This number is published in Keenan, Shtogrin, and Sobehart (1999) and is based on default data in the period 1970-1998. For year 1970, Moody's define a cohort of BBB-rated firms and track how many of those firms default over the next 10 years. The 10-year BBB default frequency for 1970 is the number of defaulted firms divided by the number of firms in the cohort.[18] The average default rate of 4.39% is calculated as an average of 10-year default rates of cohorts formed at yearly intervals over the period 1970-1988[19].

We show in a simulation that in an economy where the ex ante 10-year default probability is 4.39% for all firms, the ex post realized 10-year default frequency averaged over 28 years can be dramatically different. We do so by populating the economy with identical firms, thus eliminating, or abstracting from, any convexity bias due to cross-sectional variation in the firms. What is crucial in the simulation is that firms are exposed to systematic risk which

[16]See for example Zhang, Zhou, and Zhu (2009) and the articles mentioned in Table 1.

[17]All articles mentioned in Table 1 use Moody's estimates of default frequencies.

[18]Some firms have their rating withdrawn and Moody's have incomplete knowledge of subsequent defaults once firms are no longer rated. Moody's adjust for this by assuming that firms with withdrawn ratings would have faced the same risk of default as other similarly rated issuers if they had stayed in the data sample. Evidence in Hamilton and Cantor (2006) suggests that this is a reasonable assumption.

[19]In recent years Moody's calculate average default frequencies based on monthly cohorts instead of yearly cohorts; the difference between default frequencies using monthly and yearly cohorts is small.

leads to correlation in defaults.

The simulation is done as follows. We assume that in year 1 we have 1,000 identical firms, where firm i's value under the historical measure follows the process

$$\frac{dV_t^i}{V_t^i} = (\pi_A + r - \delta)dt + \sigma_A dW_{it}^P \tag{5}$$

and π_A is the asset risk premium while W_i^P is a standard Brownian motion. We choose 1,000 firms each month because the average number of firms in Moody's BBB cohorts during the last decade is close to 1,000[20]. We assume every firm has one T-year bond outstanding, and a firm defaults if firm value is below bond face value at bond maturity, $V_T^i \leq F$. Using the properties of a Geometric Brownian Motion, the default probability is

$$p = P(W_{iT}^P - W_{i0}^P \leq c) \tag{6}$$

where $c = \frac{\log(F/V_0) - (\pi_A + r - \delta - \frac{1}{2}\sigma_A^2)T)}{\sigma_A}$. This implies that the unconditional default probability is $N(\frac{c}{\sqrt{T}})$ where N is the cumulative normal distribution. Note that for a given default probability p we can always find c such that equation (6) holds, so in the following we use p instead of the underlying Merton parameters that give rise to p.

We introduce systematic risk by assuming that

$$W_{iT}^P = \sqrt{\rho} W_{sT} + \sqrt{1-\rho} W_{iT} \tag{7}$$

where W_i is a Wiener process specific for firm i, W_s is a Wiener process common to all firms, and ρ is the pairwise correlation between percentage firm value changes. All Wiener processes are independent. The realized 10-year default frequency in the year 1-cohort is found by simulating one systematic and 1,000 idiosyncratic processes in equation (7).

In year 2 we form a cohort of 1,000 new fims. The firms in year 2 are identical to the previous firms as they entered the index in year 1. We calculate the realized 10-year default

[20]The average number of BBB cohort firms during 1970-2012 is 606. There has been an increasing trend from 372 in 1970 to 1,245 in 2012. The results are very similar if we use 600 or 1,250 firms instead of 1,000.

frequency of the year 2-cohort as we did for the year 1-cohort. Note that the common shock in year 1-9 for the year 2-cohort is the same as the common shock in year 2-10 for firms in the year 1-cohort. We do this for 18 years and calculate the overall realized cumulative 10-year default frequency in the economy by taking an average of the default frequencies across the 18 cohorts. We repeat this simulation 100,000 times.

There are two parameters in our simulation; the defalt probability p and the default correlation ρ. We focus on BBB firms and set $p = 4.39\%$, the realized 10-year cumulative BBB default frequency used in Huang and Huang (2012). We assume that half of firm volatility is systematic ($\rho = 0.25$) consistent with evidence in Choi and Richardson (2012)[21]. In each of the 100,000 simulations, we simulate 18 years of firms (in total 18,000 firms) and since we look at 10-year default frequencies, we simulate 28 years of data.

Figure 2 shows the distribution of realized default frequencies in the simulation study. A 95% confidence interval is [0.56%; 13.50%]. The black vertical line shows the ex ante default probability of 4.39%. Given that we simulate 18,000 firms over a period of 28 years, it might be surprising that realized default frequency can be so far from ex ante default probability. The reason is that there is systematic risk in the economy and this induces correlation in defaults among firms. If there is no systematic risk in the economy Table 10 shows that a 95% confidence interval for the realized default frequency is $[4.11\%; 4.68\%]$.

We also see from Figure 2 that the default frequency is significantly skewed to the right, i.e., the modal value (1.77%) is significantly below the mean (4.39%). This means that the default frequency *most often* observed – e.g., long-run data from the rating agencies – is below the mean and, if spreads reflect the true expected default rate, they will appear too high relative to the most frequently observed historical loss rate[22].

[21]See Table 5 in Choi and Richardson (2012). $\rho = 0.25$ is also broadly consistent with the fact that during our sample period the annualized volatility of daily returns of the S&P500 index was 21.5% while the median equity volatility for A-rated forms respectively BBB-rated firms in our sample is 31% respectively 47%.

[22]This is a point that Moody's KMV are aware of, see for example Kealhofer, Kwok, and Weng (1998)

If we denote the realized cumulative 10-year default frequency, averaged over 28 years, by X, then the above simulation gives the distribution of $X|p$. We can use the simulation to estimate the ex ante default probability that is consistent with a realized default rate of 4.39%. Bayes' rules gives

$$f(p|X) \propto f(X|p)f(p)$$

where $f(p)$ is the prior distribution of p and $f(X|p)$ is the distribution of default losses conditional on p. If we assume a flat prior on p, the distribution of p is given by $f(p|X) \propto f(X|p)$. In Appendix C we derive an approximate closed-form solution for the distribution of $X|p$ which we use in the following.

Figure 3 shows the distribution of p given an average cumulative 10-year default experience of 4.39% over 28 years. We see that the distribution of p is wide with a 95% confidence band of $[1.9\%; 20.1\%]$. Intuitively, as high values of p as 20.1% are consistent with an ex post default rate of 4.39% because the distribution of default rates is highly skewed. The modal value of 5.35% is higher than the realized default frequency of 4.39%. Since the distribution of p corresponds to the likelihood function, this implies that the maximum likelihood estimator of p is 5.35% and so the realized default frequency is a downward biased estimator of p.

Moody's started to record default rates in 1920 and we can reduce statistical uncertainty by extracting an estimate of the ex ante default probability from the average default frequency for the period 1920-2012 instead of for the period 1970-1998. The average 10-year BBB cumulative default frequency for 1920-2012 is 7.112% according to Ou, Chiu, Wen, and Metz (2013) and Figure 3 shows the density for p using this default frequency and time period as well. Although the width of the confidence band is reduced to $[4.2\%; 14.1\%]$, the statistical uncertainty is still large. It is well-known that a very long time series is necessary to estimate the equity premium accurately, see for example Merton (1980), but our results

and Bohn, Arora, and Korablev (2005).

document that a very long time series is necessary to estimate the ex ante default probability as well.

How large does the ex ante default probability have to be to make the actual spread consistent with the spread in the Merton model? Using the results in Chen, Collin-Dufresne, and Goldstein (2009) (CCG) we estimate the ex ante default probability that makes the credit spread puzzle disappear to be 8.16% and plot this probability in Figure 3[23]. The figure clearly shows that $p = 8.16\%$ cannot be rejected on the basis of the distribution of p conditional on the realized default frequency in either the short 1970-1998 or long 1920-2012 sample period.

In Table 11, we estimate for different rating classes and bond maturities confidence bands for model-implied spreads when using the realized default rate as a proxy for expected default probability. Note that we are ignoring any convexity bias and therefore spreads might be biased and confidence bands therefore too tight. For actual spreads we use estimates from Duffee (1998) commonly used in the literature as shown in Table 1. We see in Table 11 that using default rates from 1920-2012 gives considerably higher investment grade spreads compared to using default rates from 1970-1998, while this is not the case for speculative grade bonds. Partially, this can be explained by the fact that the distribution of realized

[23]CCG derive the credit spread for a T-year zero coupon bond in the Merton model as

$$(y - r) = -\frac{1}{T} \log \left(1 - LN[N^{-1}(p) + \theta\sqrt{T}]\right) \tag{8}$$

where p is the T-year default probability under the historical measure, L is the loss given default, and θ is the asset Sharpe ratio. CCG assume that the recovery rate is $1 - L$ while we assume it is $\min(1 - L, \frac{V_T}{F})$, but this is unlikely to lead to any significant differences. CCG calculate θ to be 0.22 based on equity data from 1927-2005. We focus on the 10-year BBB-AAA spread since this spread is most widely studied in the literature and there is no issue about which riskfree rate to use. Table 1 shows that estimates in the literature of the actual 10-year BBB-AAA spread for non-callable bonds are fairly consistent around 105bps. If we assume that the ratio between AAA and BBB default rates (from the period 1920-2012) is constant, a BBB default rate of 8.16% (and AAA default rate of 0.92%) gives a 10-year BBB-AAA spread of 105bps.

defaults becomes less skewed as default probability increases and for high default probabilities there is a negative skewness. So using default rates from a "calm period" results in a credit spread puzzle that becomes more severe as credit quality increases. The table also shows that none of the actual spreads are statistically significantly different from Merton model spreads, so we cannot reject that both spreads and default frequencies are consistent with predictions from the Merton model.

In the Merton model the expected equity return is related to the expected bond return, and one can be inferred from the other. Campello, Chen, and Zhang (2008) use this insight to measure expected equity returns from bond yields. The motivation is that the use of noisy realized equity returns to proxy for expected equity returns is avoided. In their calculation of the expected bond return, they compute the expected default loss using historical default frequencies. However, our finding that default frequencies are poor proxies for default probabilities raises concerns regarding this methodology, since it replaces noise in realized equity returns with noise in realized default rates. Other papers measuring expected bond (or CDS) returns by using historical default frequencies includes among others, Elton, Gruber, Agrawal, and Mann (2001b) and Bongaerts, Driessen, and de Jong (2011).

In the simulations we assume that the cohorts are reformed each year and so a given firm only appears in a single cohort. This is a simplification because when Moody's forms cohorts of BBB-rated firms from year to year, there is a substantial overlap of firms from one cohort to the next. This overlap adds additional correlation in defaults on top of the correlation caused by systematic risk. If we allowed firms to stay in more than one cohort, the variation in default frequency would be even larger. Also, the amount of systematic risk is assumed to be $\rho = 0.25$, but as Table 10 shows, even with a very low levels of systematic risk, confidence bands are still wide. Finally, we assume in the simulation that firms are identical, ignoring cross-sectional variation in firm characteristics. As we showed in the previous section firm heterogeneity is important, but precisely for BBB-rated firms at a horizon of 10 years the impact of firm heterogeneity on default probabilities is small relative to the effect of systematic risk. We see in Table 8 that the bias is modest for long-term BBB-rated bonds and we have confirmed this in simulations.

5. New evidence on the credit spread puzzle

We showed in the previous two sections that tests of structural models using the representative firm approach are biased and, when calibrated to historical default rates, suffer from low statistical power. An approach to testing structural models that avoids these issues is to compare model-implied and actual spreads on a transaction-by-transaction basis. To our knowledge only Eom, Helwege, and Huang (2004) and Ericsson, Reneby, and Wang (2007) take this approach and it is noteworthy that neither of those papers finds that structural models systematically underpredict spreads. Eom, Helwege, and Huang (2004)'s data consist of 182 trader quotes in the period 1986-1997 while Ericsson, Reneby, and Wang (2007)'s consist of 1387 transactions over the period 1994-2003. With the availability of TRACE we can conduct a large-scale examination of the Merton model using 396,890 transactions for the period 2002-2012. This allows us to examine in detail the ability of the Merton model to price bonds across maturity, across ratings, and over a time period that includes both a boom period and a recession.

There are broadly three versions of the puzzle[24]:

- *Long-term yield spreads between BBB- and AAA-rated bonds are too high to be ex-*

[24] We searched the top-3 finance journals (Journal of Finance, Review of Financial Studies, and Journal of Financial Economics) for articles proposing extensions of standard structural models that raise yield spreads relative to standard models and found 10 articles. All 10 articles examine the long-term BBB-AAA spread (Puzzle 1), 5 examine short-term spreads (Puzzle 2), and 6 examine long-term AAA spreads (Puzzle III). Eom, Helwege, and Huang (2004), Ericsson and Renault (2006), Hackbarth, Miao, and Morellec (2006), Zhang, Zhou, and Zhu (2009), and He and Xiong (2012) examine all three puzzles, Cremers, Driessen, and Maenhout (2008) examine Puzzle I and III, while David (2008), Chen, Collin-Dufresne, and Goldstein (2009), Chen (2010), and Bhamra, Kuehn, and Strebulaev (2010) examine Puzzle I.

plained by standard structural models of credit risk. The yield spread analysed is typically for bonds with a maturity close to 10 years. If potential non-default components of yield spreads such as taxes or liquidity are the same for AAA- and BBB-rated bonds, this version of the credit spread puzzle offers a "clean" spread uncontaminated by non-credit effects. Table 1 shows that the actual spread is often estimated to be around 105bps while the model spread is 47-77bps.

- *Yield spreads on high-quality bonds with short maturity are too high to be explained by standard structural models of credit risk.* Short maturity typically means one year or less and high-quality refers to investment grade bonds. Since standard structural models typically predict yield spreads close to zero for high-quality bonds at short maturities, this version of the puzzle is not very sensitive to model specification; if there is a significantly positive short-term spread there is a puzzle. Note that there is no empirical evidence on corporate bond spreads on bonds with a maturity below one year.

- *Yield spreads on high-quality bonds with long maturity are too high to be explained by standard structural models of credit risk.* Long maturity is typically 10 years and high-quality refers to bonds with a rating of AAA. Table 1 shows that previous estimates of the actual 10-year AAA spread **to Treasury yields** is 47-63bps while the model-implied spread is 2-12bps.

The long-term spread between BBB and AAA-rated bonds has received most attention in the literature and we will focus on this puzzle first before examining the other two versions. When we calculate spreads we take the median across all bond transactions and weight by the volume of each transaction.[25] The use of medians is robust to the presence of

[25]To calculate the volume-weighted median, we sort spreads in increasing size $s_1 < s_2 < ... < s_T$ with corresponding normalized volumes $\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_T$, where $\tilde{v}_i = \frac{v_i}{\sum_{j=1}^{T} v_j}$ and then find t such that $\sum_{i=1}^{t} \tilde{v}_i \geq 0.5$ and $\sum_{i=1}^{t-1} \tilde{v}_i < 0.5$. The volume-weighted median spread is then s_t.

potential outliers and in most cases we also report volume-weighted 10 and 90pct quantiles which are informative about the distribution of spreads. We weight by volume following the recommendation by Bessembinder, Kahle, Maxwell, and Xu (2009).

5.1 Long-term BBB-AAA yield spreads

Table 12 shows actual and model-implied bond spreads for our sample. The actual median long-term BBB-AAA spread is 100bps-12bps=88bps while the model-implied spread is 104bps-0bps=104bps. Thus, in contrast to what most of the previous literature has found, on average the Merton model does not underpredict long-term BBB-AAA spreads. Since spreads are volume-weighted over all transactions, the results in the table have to be interpreted with caution. For example, trading in AAA/AA bonds went up strongly in 2009 while trading in BBB went down, so the weight of 2009 in AAA/AA spreads is higher than in BBB spreads. To get a more accurate picture of the ability of the Merton model to match spreads, we examine the time series variation in long-term BBB-AAA spreads in Figure 4[26]. The graph shows that the Merton model cannot quite match the level of the actual spread during 2005-2007, but apart from this period the model-implied spread tracks the actual spread surprisingly well.

Overall, we find no evidence that actual long-term BBB-AAA spreads are consistently higher than model-implied long-term BBB-AAA spreads.

[26]We calculate the 10pct and 90pct quantiles in the Figure by simulation on a quarterly basis: we draw a transaction in a BBB bond from the pool of actual BBB transactions where the probability of drawing a particular transaction is proportional to transaction volume. In the same way we draw a AAA/AA transaction. We calculate from the two transactions a BBB-AAA spread. We repeat this procedure 5,000 times and calculate the 10pct and 90pct quantile in the 5,000 simulated BBB-AAA spreads.

5.2 Short-term yield spreads on high quality bonds

Predicted spreads in standard models of credit risk for short-maturity investment grade bonds are very low and this has been viewed as a failure of structural models. However, to the best of our knowledge there is no empirical evidence on the actual size of corporate bond credit spreads for maturities shorter than one year. The reason is that most previous research had to rely on quotes and typically quotes were only available for bonds that were included in an index, such as the Lehman or Merrill Lynch indices. Bonds usually drop out of indices when the maturity falls below one year and, when this happens, these bonds typically stop being quoted. Since we use transactions data, we observe transactions on bonds with any maturity, and our results on short-term bonds provide new evidence on the actual size of short-term corporate bond spreads.

In Table 12 we see that the median yield spread for AAA/AA-rated bonds with a maturity less than one year is -4bps and 7bps for A-rated bonds. This is surprisingly close to zero[27]. In contrast, the median actual yield spread for short maturity BBB bonds is 55bps while the model-implied spread is zero. Figure 5 shows the time variation of short-term spreads for investment grade bonds. Actual spreads for AAA and AA bonds are close to zero except during the volatile period in 2008 where we see a jump in short-term spreads of around 50bps. For A-rated bonds we see a similar pattern except that there is a larger jump in 2009 to around 200bps. Actual short-term spreads for BBB-rated bonds in first half of the sample period are in the range of 20-100bps and confidence bands do not contain zero. For the first half of the sample period model-implied spreads are zero. Proposed explanations in the literature for positive short-term spreads are incomplete accounting information (Duffie and Lando (2001)) or jumps in firm value (Zhou (2001)). However, any explanation raising short-term BBB credit spreads must have only a small effect on AAA/AA/A short-term

[27] Feldhütter and Lando (2008) find that there is a 4-5bps credit risk premium in LIBOR/swap rates, so correcting for this premium the yield spread for AAA bonds is 0-1bps.

spreads since these are close to zero.

In the second half of the sample period we see that actual BBB short-term credit spreads are high but model-implied spreads are sometimes even higher. When calculating the credit spread, we assume that all of the firm's debt is due when the bond matures as discussed in Section 2.2 and this can raise short-term spreads significantly. In the Internet Appendix we relax the assumption that all debt matures at bond maturity and implement the Leland-Toft model. This lowers short-term spreads substantially relative to the Merton model while long-term spreads are similar.

5.3 Long-term yield spreads on high-quality bonds

Table 12 shows that the median actual long-term AAA spread over the sample period is 12bps while the model-implied is, once again, zero. This finding is consistent with the previous literature that finds that structural models cannot reproduce spreads of long-term high quality yield spreads. The difference of 12bps is smaller than previous literature has found (see Table 1), but this is simply due to the use of swap rates as the riskfree rate instead Treasury yields. Figure 6 shows the time variation of the long-term AAA spread. In 2002-2003 the model-implied spread tracks the actual spread fairly well, while from 2003 and onwards the actual spread is higher than the model spread.

The underprediction of long-term high-quality spreads could be related to the fact that some firms that have no or little leverage despite the tax advantage of debt; the "under-leverage puzzle" (see Strebulaev and Yang (2013)). However, if we restrict our sample to firms that have at least 5, 10, or even 15% leverage the results for AAA are similar. Therefore our results are not driven by firms with no or little leverage.

Overall, our findings imply that the Merton model captures the size of average spreads better than previously documented. Consistent with the previous literature on the credit spread puzzle, the results above apply to average spreads over a period of a quarter or longer.

However, our results do not necessarily imply that the Merton model accurately prices bonds in the cross section - at least within rating categories. At the level of individual bonds there are significant deviations. This is consistent with the evidence in Collin-Dufresne, Goldstein, and Martin (2001).[28]

6 Robustness checks

To assess how specific our results are to the use of the Merton model, we implement the Leland-Toft model in an Internet Appendix. The Leland-Toft model has a maturity structure of debt with the default time endogenously chosen by the equity holders and the assumption in the Merton model that all debt matures at bond maturity is relaxed. We find the convexity bias in the Leland-Toft model to be very similar to that in the Merton model. Furthermore, the level and time series variation of the long-term BBB-AAA spread in the Leland-Toft model are similar to in the Merton model. Thus, the conclusions we draw in the previous three sections are not specific to the Merton model.

In the Internet Appendix we also use daily quotes from Merrill Lynch (ML) for the period 1997-2012 instead of transactions data from TRACE for the period 2002-2012. Corporate bond quotes are typically dealer bid quotes and this is also the case for the ML data. A disadvantage when using the ML data is that quoted yields are higher than midyields and in the Internet Appendix we document how large the difference is over time and across rating. The advantage of using the ML data is that it allows us to examine the period 1997-2002. Since there is no transaction volume and one quote per day per bond, using the ML data also allows us to see the effect of equally-weighting bonds. The main result when using the ML data is that the level and time series variation of the long-term BBB-AAA spread is still captured by the Merton model.

While we only examine bonds issued by industrial firms in the precious sections, we

[28] See also Correia, Richardson, and Tuna (2012) for further evidence.

look at the pricing of bonds issued by financial firms in the Internet Appendix. We find that Merton model spreads are higher than actual spreads for all maturities and investment grade rating categories, so the Merton model does not price bonds issued by financial firms well.

Several previous studies of the credit spread puzzle use callable bonds in their analysis as Table 1 shows. In the Internet Appendix we therefore repeat the analysis for callable bonds. We find that the Merton model matches the level of the long-term BBB-AAA spred, but actual long-term AAA and BBB spreads are approximately 50bps higher than implied spreads. Using a sample of firms that have issued both callable and noncallable bonds, Jarrow, Li, Liu, and Wu (2010) estimate that the call option on average raises spreads by 48 basis points. If their finding holds for our sample period, the Merton model matches the option-adjusted levels of corporate spreads for both AAA and BBB. For short-maturity bonds the value of the call option is likely close to zero and we find that actual spreads for bonds rated A or higher are close to zero while for BBB they are 33 basis points, consistent with the evidence for non-callable bonds. When we allow for bonds that may contain covenants as well as being callable the results are similar.

Finally, the Internet Appendix gives results where we use the Treasury yield as riskfree rate instead of the swap rate to compute spreads. Not surprisingly, actual spreads are 17-45 basis points higher, but there is no significant change on the BBB-AAA spread.

7. Conclusion

The credit spread puzzle – the underprediction of credit spreads by structural models – arises in studies that typically use one or both of two methods. In the first - "representative firm" - approach, spreads are computed using average firm variables as model inputs. In the second, model parameters are obtained by calibrating the model to historical default frequencies. We find that there are problems with both these approaches. The first is that spreads are typically convex in the input parameters such as asset volatility and the leverage ratio and this means that average spreads are higher than spreads for a representative firm. A very similar bias occurs when using these models to compute default probabilities. We examine these biases empirically and find them to be economically highly significant. The second problem is that when fitting to historical default frequencies, an implicit assumption is made that the ex-post historical default frequency is a good proxy for the ex-ante default probability. In a simulation study we find that even when almost 30 years of data is used to compute the historical default frequency, it can differ dramatically from the ex-ante average default probability. Thus the statistical power of fitting to historical default frequencies is low.

We then test the Merton model using a bias-free approach and find that the Merton model captures the level and time series variation of the long-term BBB-AAA US corporate bond spread. We find that, consistent with predictions from the Merton model, actual yield spreads on short-term AAA-A-rated bonds are close to zero in normal times while short-term BBB spreads are higher than those predicted by the model. We further find that the Merton model undershoots long-term AAA-A rated spreads by 12-38 basis points. Overall, we find that the evidence for a credit spread puzzle is much weaker than previously found.

Our results show that when testing structural models it is critically important to take into account the cross-sectional and time series variation of leverage and other firm variables. They also show that the statistical uncertainty associated with historical default frequencies is, for most purposes, too large make them useful proxies for expected default frequency.

We show in an Internet Appendix that the empirical results using the Leland-Toft model are similar to those using the Merton model, so our conclusions are not particular to the Merton model. This implies that extensions of standard structural models face harder cross-sectional restrictions than previously thought. An extension raising long-term A-AAA rates spreads by 12-38 basis points on average would make model spreads consistent with actual spreads for high credit quality firms. However, since the model already explains the long-term BBB-AAA spread, BBB spreads should not be raised significantly more. At the very short end of the yield curve, an extension raising BBB spreads by 55 basis points to a level consistent with actual spreads, should only have a small effect on AAA, AA, and A spreads.

Our results suggest that the failure of structural models to explain credit spreads may have less to do with deficiencies of the model than with the way in which it has been implemented. As we have stated earlier, our claim is certainly not that the Merton model is the "best" model of corporate debt; rather it is that even a model as simple as the Merton model does quite a good job if implemented in a manner that avoids both bias and problems of statistical power.

A Data

This Appendix gives details on the corporate bond transactions dataset and how firm variables, leverage, payout rate, and equity volatility are calculated using CRSP/Compustat.

A.1 Bond data

Since July 1, 2002, members of the Financial Industry Regulatory Authority have been required to report their secondary over-the-counter corporate bond transactions through the Trade Reporting and Compliance Engine (TRACE) and the transactions are disseminated to the public within 15 minutes.[29]

Initially, the collected trade information was publicly disseminated only for investment grade bonds with issue sizes greater than $1 billion. Gradually, the set of bonds subject to transaction dissemination increased and since January 9, 2006 transactions in all non-144A bonds transactions have been immediately disseminated.[30] Goldstein and Hotchkiss (2008) provide a detailed account of the dissemination stages. In the publicly disseminated data the trade size is capped at $5 million in investment grade transactions and $1 million in speculative grade transactions. Since November 3, 2008, the publicly available TRACE data indicate whether a transaction is an interdealer transaction or a transaction with a customer and, if a customer transaction, whether the broker-dealer is on the buy or the sell side. This publicly disseminated data is available through Wharton Research Data Services (WRDS) and is used in for example Dick-Nielsen, Feldhütter, and Lando (2012) and Bao, Pan, and Wang (2011). We use this data for the period September 15, 2011- June 30, 2012.

[29]In the initial phase of TRACE the disseminating times were longer than 15 minutes. Since July 1, 2005 the reporting and dissemination is required to occur within 15 minutes after the trade.

[30]Rule 144a allows for private resale of certain restricted securities to qualified institutional buyers. According to TRACE Fact Book 2011, the percent of rule 144A transactions relative to all transactions was 2.0% in investment grade bonds and 8.4% in speculative grade bonds. Also, transactions reported on or through an exchange are not included in TRACE.

Through FINRA we have access to historical transactions information not previously disseminated. The historical data is richer than the WRDS data in three aspects. First, the data contains all transactions in non-144A bonds since July 2002, so the data set for the first years of TRACE is significantly larger than the WRDS data set. Second, the data has buy/sell indicators for all transactions, not just after October 2008 as in the WRDS data set. Third, trade volumes are not capped. FINRA provide access to the enhanced historical data with a lag of 18 months. We use this data for the period July 1, 2002-September 14, 2011.

We obtain bond information from the Mergent Fixed Income Securities Database (FISD) and limit the sample to senior unsecured fixed rate or zero coupon bonds. We exclude bonds that are callable, convertible, putable, perpetual, foreign denominated, Yankee, have sinking fund provisions, or have covenants. For bond rating, we use the lower of Moody's rating and S&P's rating and discard any transactions where the bond does not have a Moody's or S&P rating on the transaction day. We track rating changes on a bond, so the same bond can appear in several rating categories over time. Bonds for which FISD do not provide information are dropped from the sample. Erroneous trades are filtering out as described in Dick-Nielsen (2009). We exclude transactions with a yield of 99999.9999% or 99999.99%.

A.2 Firm data

To compute bond prices in the Merton model we need the issuing firm's leverage ratio, payout ratio, and asset volatility. Firm variables are collected in CRSP and Compustat. To do so we match a bond's CUSIP with CRSP's CUSIP. In theory the first 6 digits of the bond cusip plus the digits '10' corresponds to CRSP's CUSIP, but in practice only a small fraction of firms is matched this way. Even if there is a match we check if the issuing firm has experienced M&A activity during the life of the bond. If there is no match, we hand-match a bond cusip with firm variables in CRSP/Compustat.

Leverage ratio: Equity value is calculated on a daily basis by multiplying the number of shares outstanding with the price of shares. Debt value is calculated in Compustat as the latest quarter observation of long-term debt (DLTTQ) plus debt in current liabilities (DLCQ). Leverage ratio is calculated as $\frac{\text{Debt value}}{\text{Debt value+Equity value}}$.

Payout ratio: The total outflow to stake holders in the firm is interest payments to debt holders, dividend payments to equity holders, and net stock repurchases. Interest payments to debt holders is calculated as the previous year's total interest payments (previous fourth quarter's INTPNY). Dividend payments to equity holders is the indicated annual dividend (DVI) multiplied by the number of shares. The indicated annual dividend is updated on a daily basis and is adjusted for stock splits etc. Net stock repurchase is the previous year's total repurchase of common and preferred stock (previous fourth quarter's PRSTKCY). The payout ratio is the total outflow to stake holders divided by firm value, where firm value is equity value plus debt value.

Equity volatility: We calculate the standard deviation of daily returns (RET in CRSP) in the past three years to estimate daily volatility. We multiply the daily standard deviation with $\sqrt{255}$ to calculate annualized equity volatility. If there are no return observations on more than half the days in the three year historical window, we do not calculate equity volatility and discard any bond transactions on that day.

B Pricing formulas

For completeness we include in this Appendix formulas for bond prices, default probabilities, and deadweight losses.

B.1 Bond price

Equation (1) states firm value as a Geometric Brownian motion under the pricing measure,

$$\frac{dV_t}{V_t} = (r - \delta)dt + \sigma_A dW_t,$$

and the firm defaults if firm value is below face value of the bond at bond maturity, $V_T < F$. The bond price at time 0 is calculated as

$$P(0,T) = E^Q[e^{-rT}(1_{\{V_T \geq F\}} + \min(R, \frac{V_T}{F})1_{\{V_T < F\}})] \tag{9}$$

where R is the recovery rate. From Eom, Helwege, and Huang (2004) Appendix A.1 we have that

$$E^Q[1_{\{V_T \geq F\}}] = N(d_2(F,T)) \tag{10}$$

and

$$E^Q[1_{\{V_T < F\}} \min(\psi, V_T)] = \frac{V_0}{D(0,T)} e^{-\delta T} N(-d_1(\psi,T)) + \psi[N(d_2(\psi,T)) - N(d_2(F,T))], \tag{11}$$

where $\psi \in [0, K]$, N represents the cumulative standard normal function,

$$\begin{aligned} d_1(x,T) &= \frac{\log(\frac{V_0}{xD(0,T)}) + (-\delta + \sigma_A^2/2)T}{\sigma_A \sqrt{T}}, \\ d_2(x,T) &= d_1(x,T) - \sigma_A \sqrt{T}, \end{aligned}$$

and $D(0,T) = \exp(-rT)$. Using equation (10) and (11) in (9) gives the solution to the bond price. The yield spread is calculated as

$$s(0,T) = -\frac{\log(P(0,T))}{T} - r. \tag{12}$$

B.2 Default probabilities and deadweight loss

Under the physical measure firm value is given as the Geometric Brownian motion

$$\frac{dV_t}{V_t} = (\pi_A + r - \delta)dt + \sigma_A dW_t,$$

where π_A is the asset risk premium.

Default probabilities are given by (10) where the expectation is taken under the physical measure.

The expected deadweight loss in bankruptcy as a fraction of asset value is given as

$$1 - E^P[\frac{\min(V_T, RF)}{V_T}|V_T < F]$$

and this expression can be solved numerically using that V_T is log-normally distributed.

C Analytic formula for the distribution of realized default frequencies

Section 4 explains how Moody's calculate realized default frequencies over longer periods of time, and the realized default frequencies are used in a number of papers as moments to match when calibrating structural models. In this Appendix we derive an approximate analytical solution for the distribution of realized default frequencies.

We assume as in Section 4 that firms are identical, exposed to systematic risk, and firm default occurs according to the Merton model. The cumulative T-year default probability for firm i in the cohort born at time t is given in equation (6) as

$$p = P(c \leq W_{i,t+T}^P - W_{i,t}^P)$$

where c depends on the underlying firm parameters as described in the main text. Furthermore, equation (7) gives the Brownian motion as

$$W_{i,t}^P = \sqrt{\rho} W_{s,t} + \sqrt{1-\rho} W_{i,t}$$

where W_s is common to all firms and W_i is specific to firm i.

Assume that Moody's calculate the average T-year cumulative default frequency over M cohorts (where there is one cohort each year).

First let $M = 1$; for example Moody's calculate the 10-year cumulative default frequency based on one cohort born in 1970. Vasiček (1991) derives an approximate analytical solution for the distribution for the realized default frequency[31]: the cumulative distribution function is given as

$$W(s|p,\rho) = N\Big(\frac{1}{\sqrt{\rho}}\big(\sqrt{1-\rho}N^{-1}(s) - N^{-1}(p)\big)\Big)$$

where N is the cumulative normal distribution function and s is the default frequency. Its density is

$$f(s|p,\rho) = \frac{\sqrt{1-\rho}}{\rho}\exp\Big(-\frac{1}{2\rho}\big(\sqrt{1-\rho}N^{-1}(s) - N^{-1}(p)\big)^2 + \frac{1}{2}\big(N^{-1}(s)\big)^2\Big). \tag{13}$$

Assume now $M > 1$. If firms are from the same cohort the correlation of their Brownian motions is 0.25. If firm i is from cohort t and firm j is from cohort $t+1$ the covariance of their Brownian motions is

$$\begin{aligned} cov(W^P_{i,t+T} - W^P_{i,t}, W^P_{j,t+1+T} - W^P_{j,t+1}) &= cov(\sqrt{\rho}(W_{s,t+T} - W_{s,t}), \sqrt{\rho}(W_{s,t+1+T} - W_{s,t+1}) \\ &= \rho cov(W_{s,t+T} - W_{s,t+1}, W_{s,t+T} - W_{s,t+1}) \\ &= \rho(T-1) \end{aligned}$$

and the correlation is

$$\frac{\rho(T-1)}{\sqrt{T}\sqrt{T}} = \frac{T-1}{T}\rho.$$

It is easy to see that for two firms in cohorts k years a part where $k < M$ the correlation is $\frac{T-k}{T}\rho$ and 0 if $k \geq T$, so the correlation is $\max(0, T-k)\rho$. We can now define the average correlation as

$$\rho\frac{1}{M}\sum_{i=1}^{M}\frac{1}{M}\sum_{j=1}^{M}\frac{\max(0, T-|i-j|)}{T} \tag{14}$$

[31]The approximation lies in that Vasicek assumes there is an infinite number of firms in the cohort instead of a finite number.

and use this average correlation in the Vasiček (1991) formula.

Table A1 shows quantiles from both the simulation and approximation where the default probability is $p = 4.39\%$ and correlation is $\rho = 0.25$. We see that the approximation underestimates the widths of confidence bands slightly. For a sample period of 30 years for example, the width of the 95% confidence band in the simulated distribution is $13.15\% - 0.61\% = 12.54\%$ while it is $12.61\% - 0.71\% = 11.90\%$ in the approximating distribution. We therefore underestimate the uncertainty of estimated default rates slightly by using the approximating distribution.

References

Alderson, M. and B. Betker (1995). Liquidation costs and capital structure. *Journal of Financial Economics 39*, 45–69.

Bao, J. and J. Pan (2013). Bond Illiquidity and Excess Volatility. *Review of Financial Studies, forthcoming*.

Bao, J., J. Pan, and J. Wang (2011). The illiquidity of Corporate Bonds. *Journal of Finance 66*, 911–946.

Bessembinder, H., K. M. Kahle, W. F. Maxwell, and D. Xu (2009). Measuring abnormal bond performance. *Review of Financial Studies 22*(10), 4219–4258.

Bhamra, H. S., L.-A. Kuehn, and I. A. Strebulaev (2010). The Levered Equity Risk Premium and Credit Spreads: A Unified Framework. *Review of Financial Studies 23*(2), 645–703.

Bohn, J., N. Arora, and I. Korablev (2005). Power and level validation of the EDF credit measure in the U.S. market. *Moody's KMV Company*.

Bongaerts, D., J. Driessen, and F. de Jong (2011). Derivative Pricing with Liquidity Risk: Theory and Evidence from the Credit Default Swap Market. *Journal of Finance 66*(1), 203–240.

Campello, M., L. Chen, and L. Zhang (2008). Expected returns, yield spreads, and asset pricing tests. *Review of Financial Studies 21*(3), 1297–1338.

Caouette, J., E. Altman, and P. Narayanan (1998). *Managing credit risk: The next great financial challenge*. New York: John Wiley and Sons, Inc.

Chen, H. (2010). Macroeconomic conditions and the puzzles of credit spreads and capital structure. *Journal of Finance 65*(6), 2171–2212.

Chen, L., P. Collin-Dufresne, and R. S. Goldstein (2009). On the relation between the credit spread puzzle and the equity premium puzzle. *Review of Financial Studies 22*, 3367–3409.

Choi, J. and M. Richardson (2012). The volatility of firm's assets and the leverage effect. *Working Paper*.

Collin-Dufresne, P., R. Goldstein, and S. Martin (2001). The Determinants of Credit Spread Changes. *Journal of Finance 56*, 2177–2207.

Correia, M., J. Kang, and S. Richardson (2014). Asset volatility. *Working paper*.

Correia, M., S. Richardson, and I. Tuna (2012). Value investing in credit markets. *Review of Accounting Studies 17*, 572–609.

Covitz, D. and C. Downing (2007). Liquidity or credit risk? The determinants of very short-term corporate yield spreads. *Journal of Finance 62*(5), 2303–2328.

Cremers, M., J. Driessen, and P. Maenhout (2008). Explaining the level of credit spreads: Option-implied jump risk premia in a firm value model. *Review of Financial Studies 21*, 2209–2242.

David, A. (2008). Inflation Uncertainty, Asset Valuations, and the Credit Spread Puzzle. *Review of Financial Studies 21*(6), 2487–2534.

Davydenko, S. A., I. A. Strebulaev, and X. Zhao (2012). A Market-Based Study of the Cost of Default. *Review of Financial Studies 25*(10), 2959–2999.

Dick-Nielsen, J. (2009). Liquidity biases in TRACE. *Journal of Fixed Income 19*(2), 43–55.

Dick-Nielsen, J., P. Feldhütter, and D. Lando (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics 103*, 471–492.

Duffee, G. R. (1998). The Relation Between Treasury Yields and Corporate Bond Yield Spreads. *Journal of Finance 53*(6), 2225–2241.

Duffie, D. and D. Lando (2001, May). Term Structures of Credit Spreads with Incomplete Accounting Information. *Econometrica*, 633–664.

Elton, E., M. Gruber, D. Agrawal, and C. Mann (2001a). Explaining the rate spread on corporate bonds. *Journal of Finance 56*, 247–277.

Elton, E. J., M. Gruber, D. Agrawal, and C. Mann (2001b). Explaing the rate spread on corporate bonds. *Journal of Finance 56*, 247–277.

Eom, Y. H., J. Helwege, and J.-Z. Huang (2004). Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies 17*(2), 499–544.

Ericsson, J. and O. Renault (2006). Liquidity and Credit Risk. *Journal of Finance 61*(5), 2219–2250.

Ericsson, J., J. Reneby, and H. Wang (2007). Can structural models price default risk? Evidence from bond and credit derivative markets. *Working Paper*.

Feldhütter, P. and D. Lando (2008). Decomposing Swap Spreads. *Journal of Financial Economics 88*, 375–405.

Gilson, S. (1997). Transactions costs and capital structure choice: Evidence from financially distressed firms. *Journal of Finance 52*(1), 161–196.

Glover, B. (2012). The Expected Cost of Default. *Working Paper*.

Goldstein, M. A. and E. Hotchkiss (2008). Dealer Behavior and the Trading of Newly Issued Corporate Bonds. *Working Paper*.

Hackbarth, D., J. Miao, and E. Morellec (2006). Capital structure, credit risk, and macroeconomic conditions. *Journal of Financial Economics 82*, 519–550.

Hamilton, D. T. and R. Cantor (2006). Measuring Corporate Default Rates. *Special Comment. New York: Moody's Investors Services*, 1–16.

He, Z. and K. Milbradt (2013). Endogenous Liquidity and Defaultable Debt. *Forthcoming, Econometrica*.

He, Z. and W. Xiong (2012). Rollover Risk and Credit Risk. *Journal of Finance 67*, 391–429.

Huang, J. and M. Huang (2012). How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk? *Review of Asset Pricing Studies 2*(2), 153–202.

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

Jarrow, R., H. Li, S. Liu, and C. Wu (2010). Reduced-Form Valuation of Callable Corporate Bonds: Theory and Evidence. *Journal of Financial Economics 95*, 227–248.

Kealhofer, S., S. Kwok, and W. Weng (1998). Uses and Abuses of Bond Default Rates. *KMV Corporation Document*.

Keenan, S., I. Shtogrin, and J. Sobehart (1999). Historical Default Rates of Corporate Bond Issuers, 1920-1998. *Special Comment. New York: Moody's Investors Services*, 1–84.

Krishnamurthy, A. and A. Vissing-Jorgensen (2012). The Aggregate Demand for Treasury Debt. *Journal of Political Economy 120*(2), 233–267.

Leland, H. (2004). Predictions of Default Probabilities in Structural Models of Debt. *Journal of Investment Management 2*.

Leland, H. (2006). Structural Models of Corporate Financial Choice. *Princeton Lectures in Finance, Lecture 1*.

Leland, H. and K. Toft (1996). Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance 51*(3), 987–1019.

McQuade, T. J. (2013). Stochastic Volatility and Asset Pricing Puzzles. *Working Paper, Harvard University.*

Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance 29*, 449–470.

Merton, R. C. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics 8*, 323–361.

Moody's (2011). Corporate Default and Recovery Rates, 1920-2010. *Moody's Investors Service*, 1–66.

Ou, S., D. Chiu, B. Wen, and A. Metz (2013). Annual Default Study: Corporate Default and Recovery Rates, 1920-2012. *Special Comment. New York: Moody's Investors Services*, 1–64.

Ou, S., D. Hamilton, and R. Cantor (2004). Short-Term Rating Performance and Corporate Commercial Paper Defaults, 1972-2004. *Special Comment. New York: Moody's Investors Services*, 1–18.

Schaefer, S. and I. Strebulaev (2008). Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds. *Journal of Financial Economics 90*, 1–19.

Strebulaev, I. and B. Yang (2013). The mystery of zero-leverage firms. *Journal of Financial Economics 109*, 1–23.

Sun, Z., D. Munves, and D. Hamilton (2012). Public Firm Expected Default Frequency (EDF) Credit Measures: Methodology, Performance, and Model Extensions. *Moody's Analytics*, 1–30.

Sundaresan, S. (2013). A Review of Merton's Model of the Firm's Capital Structure with its Wide Applications. *Annual Review of Financial Economics 5.*

Vasiček, O. (1991). Limiting Loan Loss Probability Distribution. *Technical Report, KMV Corporation.*

Zhang, B. Y., H. Zhou, and H. Zhu (2009). Explaining Credit Default Swap Spreads with the Equity Volatility and Jump Risks of Individual Firms. *Review of Financial Studies 22*(12), 5101–5131.

Zhou, C. (2001). The term structure of credit spreads with jump risk. *Journal of Banking and Finance 25*, 2015–2040.

Article	rep. firm	default freq.		benchmark model		
		fit	period	Ψ	α	δ
This paper	No	No		49%	30-35%	4.1-6%(est)
Huang and Huang(2012)	Yes	Yes	70-98	51%	15%	6%
Chen, Collin-Dufresne, and Goldstein(2009)	Yes	Yes	70-01	45%	35-40%	5.89%
Chen(2010)	Yes	Yes	70-00s	41%	0-10%(end)	-
Cremers, Driessen, and Maenhout(2008)	Yes	Yes	70-98	51%	49%	6%
Leland(2006)	Yes	Yes	70-00	50%	30%	6%
McQuade(2013)	Yes	No		51%	30%	6%
Ericsson, Reneby, and Wang(2007)	No	No		40%	15%	2.65%(est)
Eom, Helwege, and Huang(2004)	No	No		51%	30-35%	4.8%(est)

Article	benchmark model		data				
	spread	bps	bps	source	period	callable	riskfree
This paper				TRACE	02-12	No	Swap
Huang and Huang(2012)	10y AAA	10	63	Lehman	73-93	Yes	Treasury
	10y BBB-AAA	47	131	Lehman	73-93	Yes	
	4y AAA	1	55	Duffee(1998) + call spread	85-95	Yes	Treasury
	4y BBB-AAA	31	103	Duffee(1998) + call spread	85-95	Yes	
Chen, Collin-Dufresne, and Goldstein(2009)	10y BBB-AAA	77	109	Moody's	70-01	Yes	
	4y BBB-AAA	57	94-102	Lehman	74-98	No	
Chen(2010)	10y BBB-AAA	68	105	Duffee(1998) + 4bps	85-95	No	
Cremers, Driessen, and Maenhout(2008)	10y AAA	12	66	Lehman	83-02	Yes	Treasury
	10y BBB-AAA	57	105	Lehman	83-02	Yes	
Leland(2006)	10y BBB	150	50-60	Duffee(1998)+others	85-95	No	Treasury
McQuade(2013)	10y AAA	2	47	Duffe(1998)	85-95	No	Treasury
	10y BBB-AAA	43	103	Duffee(1998)	85-95	No	
	4y AAA	1	46	Duffee(1998)	85-95	No	Treasury
	4y BBB-AAA	35	103	Duffee(1998)	85-95	No	
Ericsson, Reneby, and Wang(2007)	10y	73-109	111	NAIC	97-03	No	Swap
Eom, Helwege, and Huang(2004)	?	?	?	Fixed Income Database	86-97	No	Treasury

Table 1 *Assumptions and findings in existing literature on the credit spread puzzle.* This table shows assumptions and findings in articles that test standard structural models of credit spreads in terms of matching corporate credit spreads. 'rep. firm' indicates if tests are done on a representative firm. 'def. freq' indicates if a representative firm is calibrated to fit historical default frequency, and if so 'period' is the historical period on which the default frequency is based. For the benchmark model considered ψ is the recovery rate, α is the deadweight cost of bankruptcy, and δ is the payout rate. The deadweight cost of bankruptcy is endogeneously determined in Chen(2010) and marked with '(est)' while the payout rate is estimated in several papers by calculating the total payout to equity and debt holders for individual firms and is marked with '(est)'. Under 'benchmark model' 'spread' refers to the spread examined and 'bps' states the model-implied spread in basis points in the benchmark model. Under 'data' 'bps' states the average actual spread, 'source' is the data source, 'period' is the time period over which spreads are averaged to calculate the actual spread, 'callable' indicates if callable bonds are used in the calculation of the actual spread, and 'riskfree' rates indicates if Treasury rates or swap rates are used as riskfree rates.

	Bonds	Volume	Transactions	Outstanding	Maturity	Coupon
			TRACE universe of bonds			
All	80114	51502	77043			
			+information in FISD			
All	54568	48255	73647	920	12.2	
AAA	8054	2379	3061	1460	8.1	
AA	16872	7247	11905	1163	9.2	
A	23102	13096	25211	1009	14.4	
BBB	12921	11665	16483	770	15.0	
Spec	11983	13868	16987	664	9.0	
			+standard except they may be callable			
All	25800	6438	17320	815	7.2	5.4
AAA	4970	1203	1949	832	7.2	4.6
AA	9042	1461	3957	799	7.6	4.9
A	13149	2332	7302	823	7.1	5.3
BBB	4918	710	1585	831	7.4	6.0
Spec	4215	732	2527	793	6.7	6.4
			+noncallable			
All	14556	4568	10962	987	4.7	5.0
AAA	1743	1097	1460	1770	4.8	4.6
AA	5410	1260	2923	1194	5.5	4.7
A	8474	1669	4815	877	4.5	5.1
BBB	3230	301	828	373	4.4	5.8
Spec	2260	241	935	223	3.9	6.0
			+issued by industrial firms			
All	1668	258	691	331	7.0	6.3
AAA	22	3	9	371	5.0	6.5
AA	562	52	126	374	3.2	4.7
A	695	57	222	392	4.8	5.8
BBB	733	101	194	320	8.2	6.7
Spec	498	44	139	206	12.3	8.1
			+firm has info in CRISP/COMPUSTAT			
All	389	104	395	466	7.1	6.4
AAA	7	1	3	389	8.9	7.8
AA	55	20	66	549	3.8	5.0
A	164	30	135	596	5.0	6.1
BBB	187	41	121	411	6.9	6.4
Spec	42	13	69	233	14.5	8.3

Table 2 *Data sample.* This table shows summary statistics of the initial and final data sample. 'TRACE universe of bonds' are all bonds in TRACE during the period 2002Q3-2012Q2. '+information in FISD' are bonds that have information about bond characteristics and rating in FISD. '+standard except they may be callable' are bonds that are senior unsecured fixed rate or zero coupon bonds and are not convertible, putable, perpetual, foreign denominated, Yankee, have sinking fund provisions, or have covenants. '+noncallable' are standard bonds that are not callable. '+issued by industrial firms' are standard noncallable bonds issued by industrial firms. '+firm has info in CRISP/COMPUSTAT' are standard noncallable bonds issued by industrial firms where the firm has information in CRISP/COMPUSTAT. 'Bonds' is the number of bonds, 'Volume' is the total transaction volume (in \$bn), 'Transactions' is the number of transactions (in 1,000), 'Outstanding' (in \$mm) is the average amount issued across transactions, 'Maturity' is the average bond time-to-maturity across transactions, and 'Coupon' is the average coupon across transactions.

	#firms	Mean	10th	25th	Median	75th	90th
Full sample period, 2002Q3-2012Q2							
Leverage ratio							
AAA/AA	12	0.11	0.04	0.07	0.10	0.14	0.19
A	33	0.24	0.14	0.16	0.20	0.30	0.40
BBB	52	0.41	0.19	0.24	0.31	0.56	0.80
Equity volatility							
AAA/AA	12	0.24	0.16	0.19	0.24	0.29	0.31
A	33	0.31	0.19	0.22	0.31	0.39	0.44
BBB	52	0.45	0.28	0.37	0.44	0.49	0.73
Asset volatility							
AAA/AA	12	0.21	0.14	0.17	0.22	0.25	0.27
A	33	0.24	0.15	0.18	0.24	0.30	0.33
BBB	52	0.28	0.13	0.21	0.29	0.33	0.38
Payout ratio							
AAA/AA	12	0.052	0.033	0.040	0.048	0.058	0.075
A	33	0.060	0.022	0.039	0.052	0.076	0.108
BBB	52	0.048	0.017	0.021	0.049	0.066	0.077
2002Q3-2007Q2							
Leverage ratio							
AAA/AA	11	0.11	0.03	0.04	0.10	0.14	0.19
A	29	0.22	0.13	0.15	0.20	0.30	0.35
BBB	46	0.35	0.19	0.22	0.28	0.41	0.72
Equity volatility							
AAA/AA	11	0.23	0.13	0.17	0.23	0.25	0.36
A	29	0.30	0.18	0.22	0.30	0.40	0.44
BBB	46	0.40	0.27	0.35	0.40	0.45	0.51
Asset volatility							
AAA/AA	11	0.20	0.13	0.15	0.20	0.24	0.30
A	29	0.24	0.15	0.18	0.23	0.30	0.34
BBB	46	0.27	0.11	0.22	0.29	0.33	0.34
Payout ratio							
AAA/AA	11	0.045	0.033	0.037	0.046	0.051	0.060
A	29	0.045	0.023	0.031	0.047	0.054	0.061
BBB	46	0.043	0.017	0.020	0.044	0.059	0.069
2007Q3-2012Q2							
Leverage ratio							
AAA/AA	9	0.12	0.08	0.08	0.10	0.15	0.18
A	21	0.26	0.15	0.17	0.21	0.28	0.49
BBB	23	0.57	0.27	0.39	0.64	0.75	0.80
Equity volatility							
AAA/AA	9	0.25	0.18	0.21	0.25	0.29	0.30
A	21	0.32	0.19	0.24	0.33	0.39	0.40
BBB	23	0.58	0.35	0.47	0.58	0.74	0.76
Asset volatility							
AAA/AA	9	0.22	0.15	0.18	0.23	0.26	0.27
A	21	0.24	0.16	0.17	0.24	0.31	0.33
BBB	23	0.30	0.16	0.20	0.29	0.37	0.47
Payout ratio							
AAA/AA	9	0.058	0.031	0.042	0.052	0.069	0.089
A	21	0.079	0.017	0.049	0.078	0.107	0.144
BBB	23	0.061	0.021	0.045	0.065	0.075	0.080

Table 3 *Firm summary statistics, investment grade bonds.* For each bond transaction, the leverage ratio, equity volatility, asset volatility, and payout ratio is calculated for the issuing firm on the day of the transaction. This table shows the distribution of firm values across transactions in the sample. Leverage ratio is the ratio of the book value of debt to the market value of equity plus the book value of debt. Equity volatility is the annualized volatility of daily equity returns from the last three years. Asset volatility is the unlevered equity volatility as explained in the text. Payout ratio is yearly interest payments plus dividends plus share repurchases divided by firm value. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE while firm variables are based on data from CRSP and Compustat.

	#firms	Mean	10th	25th	Median	75th	90th
Full sample period, 2002Q3-2012Q2							
Leverage ratio							
BB	21	0.53	0.23	0.33	0.63	0.69	0.75
B	9	0.66	0.31	0.45	0.74	0.83	0.93
C	4	0.87	0.78	0.82	0.88	0.93	0.96
Equity volatility							
BB	21	0.53	0.33	0.42	0.53	0.69	0.73
B	9	0.60	0.30	0.44	0.70	0.74	0.74
C	4	0.73	0.34	0.70	0.73	0.92	0.98
Asset volatility							
BB	21	0.29	0.15	0.22	0.30	0.35	0.39
B	9	0.23	0.08	0.15	0.23	0.29	0.33
C	4	0.17	0.04	0.05	0.19	0.25	0.29
Payout ratio							
BB	21	0.048	0.033	0.035	0.039	0.049	0.066
B	9	0.042	0.030	0.038	0.040	0.048	0.054
C	4	0.050	0.040	0.045	0.051	0.056	0.061
2002Q3-2007Q2							
Leverage ratio							
BB	18	0.51	0.22	0.31	0.46	0.73	0.91
B	6	0.64	0.31	0.34	0.55	0.92	0.95
C	3	0.89	0.76	0.87	0.92	0.93	0.96
Equity volatility							
BB	18	0.45	0.31	0.33	0.46	0.54	0.62
B	6	0.52	0.29	0.31	0.65	0.70	0.73
C	3	0.73	0.31	0.53	0.80	0.98	0.99
Asset volatility							
BB	18	0.25	0.05	0.19	0.28	0.33	0.34
B	6	0.18	0.06	0.09	0.20	0.22	0.37
C	3	0.14	0.04	0.05	0.13	0.21	0.26
Payout ratio							
BB	18	0.054	0.032	0.039	0.046	0.056	0.072
B	6	0.040	0.029	0.031	0.039	0.048	0.054
C	3	0.045	0.038	0.041	0.044	0.047	0.054
2007Q3-2012Q2							
Leverage ratio							
BB	10	0.55	0.27	0.35	0.66	0.69	0.72
B	7	0.68	0.34	0.73	0.75	0.77	0.80
C	3	0.86	0.78	0.80	0.85	0.92	0.97
Equity volatility							
BB	10	0.59	0.39	0.47	0.59	0.73	0.73
B	7	0.66	0.43	0.68	0.73	0.74	0.74
C	3	0.73	0.53	0.72	0.73	0.77	0.94
Asset volatility							
BB	10	0.31	0.20	0.26	0.32	0.36	0.40
B	7	0.28	0.17	0.26	0.27	0.31	0.33
C	3	0.19	0.04	0.11	0.20	0.28	0.30
Payout ratio							
BB	10	0.045	0.033	0.035	0.037	0.041	0.063
B	7	0.044	0.038	0.039	0.040	0.048	0.065
C	3	0.055	0.046	0.051	0.055	0.059	0.063

Table 4 *Firm summary statistics, speculative grade bonds.* For each bond transaction, the leverage ratio, equity volatility, asset volatility, and payout ratio is calculated for the issuing firm on the day of the transaction. This table shows the distribution of firm values across transactions in the sample. Leverage ratio is the ratio of the book value of debt to the market value of equity plus the book value of debt. Equity volatility is the annualized volatility of daily equity returns from the last three years. Asset volatility is the unlevered equity volatility as explained in the text. Payout ratio is yearly interest payments plus dividends plus share repurchases divided by firm value. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE while firm variables are based on data from CRSP and Compustat.

Short bond maturity

	AAA/AA	A	BBB	BB	B	C
Number of bonds	36	110	104	18	3	0
Mean number of bonds pr quarter	3.16	8.92	7.14	2.5	1.13	NaN
Mean number of transactions pr quarter	181	397.5	292.8	85.62	106.2	NaN
Age	5.49	7.23	6.08	6.74	4.74	NaN
Coupon	4.94	5.60	5.81	5.33	5.43	NaN
Amount outstanding ($mm)	430	465	396	360	395	NaN
Trade size (in 1,000)	328	245	284	545	179	NaN
Time-to-maturity	0.56	0.54	0.58	0.63	0.47	NaN

Medium bond maturity

	AAA/AA	A	BBB	BB	B	C
Number of bonds	46	125	119	24	3	0
Mean number of bonds pr quarter	7.08	18.8	15.5	3.18	1.17	NaN
Mean number of transactions pr quarter	641.5	1203	954	166.3	117.5	NaN
Age	3.86	6.26	4.98	8.58	3.74	NaN
Coupon	4.73	5.74	6.17	5.99	3.95	NaN
Amount outstanding ($mm)	529	613	372	237	232	NaN
Trade size (in 1,000)	296	185	305	540	195	NaN
Time-to-maturity	2.33	2.31	2.46	2.45	1.82	NaN

Long bond maturity

	AAA/AA	A	BBB	BB	B	C
Number of bonds	30	78	100	14	9	5
Mean number of bonds pr quarter	6.12	13.7	21.3	3.11	1.8	1.76
Mean number of transactions pr quarter	543.2	1044	1120	474	173.6	444
Age	7.16	7.30	5.83	12.00	14.08	16.90
Coupon	5.58	6.70	6.81	8.11	8.56	9.03
Amount outstanding ($mm)	590	628	449	244	179	186
Trade size (in 1,000)	286	251	377	192	123	102
Time-to-maturity	7.25	9.51	12.04	12.10	13.37	12.38

Table 5 *Bond summary statistics.* Only unsecured senior industrial bonds with a fixed coupon that are not callable, putable, perpetual, asset-backed, convertible, Yankee, foreign currency, and do not contain sinking fund provisions or covenants are used. All transactions in those bonds for which there are also firm variables in Crisp/Compustat are used. Short, medium, and long bond maturities are bonds with a maturity of 0-1, 1-4, and 4-30 years. This table shows summary statistics for the bond transactions. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE.

	2002Q3-2012Q2 Bond maturity			2002Q3-2007Q2 Bond maturity			2007Q3-2012Q2 Bond maturity		
	short	medium	long	short	medium	long	short	medium	long
$E[s(\theta_{it})]$									
AAA/AA	0	0	10	0	1	9	0	0	12
A	3	35	68	3	12	32	3	57	104
BBB	100	234	205	1	23	78	200	446	332
spec	113	98	285	164	78	212	0	136	357
$s(E[\theta_{it}])$									
AAA/AA	0	0	1	0	0	0	0	0	1
A	0	1	45	0	0	21	0	5	85
BBB	0	46	175	0	2	84	0	198	307
spec	0	7	265	0	7	210	0	5	324
$s(E[\theta_{it}])/E[s(\theta_{it})]$									
AAA/AA	0.00	0.00	0.05	0.00	0.00	0.04	0.00	0.00	0.06
A	0.00	0.03	0.66	0.00	0.02	0.66	0.00	0.08	0.81
BBB	0.00	0.20	0.85	0.00	0.11	1.07	0.00	0.44	0.93
spec	0.00	0.07	0.93	0.00	0.09	0.99	0.00	0.04	0.91

Table 6 *Convexity bias when calculating yield spreads in the Merton model using the representative firm approach.* It is a common approach to compare average actual spreads to model-implied spreads, where model-implied spreads are calculated using average firm variables. This introduces a bias because the spread in structural models is a non-linear function of firm variables. This table shows the magnitude of this bias. To calculate $E[s(\theta_{it})]$ we compute for every transaction in the sample the Merton spread, calculate the volume-weighted average Merton spread on a monthly basis and then average the monthly average spreads over the sample period. To calculate $s(E[\theta_{it}])$ we compute on a monthly basis the volume-weighted average firm variables (leverage ratio, asset volatility, payout ratio, bond maturity, riskfree rate at same as bond), average the monthly firm variables over the sample period, and then use the averaged firm variables to calculate the spread in the Merton model. All spreads are in basis points. Bond transactions are grouped into groups where the issued bond has remaining maturity 0-1y (short), 1-4y (medium), or 4-30y (long).

bond maturity	1y			3y			5y			7y			10y		
model parameter	L	σ_A	δ	L	σ_A	δ	L	σ_A	δ	L	σ_A	δ	L	σ_A	δ
AAA/AA	-	-	-	0.00	0.00	0.61	0.02	0.01	0.61	0.07	0.04	0.61	0.18	0.10	0.62
A	0.00	0.00	0.66	0.12	0.10	0.65	0.35	0.28	0.66	0.55	0.42	0.66	0.75	0.58	0.66
BBB	0.02	0.09	0.94	0.29	0.42	0.92	0.53	0.58	0.91	0.69	0.67	0.90	0.85	0.74	0.89
BB	0.04	0.13	0.98	0.40	0.50	0.98	0.65	0.66	0.98	0.80	0.74	0.97	0.92	0.81	0.97
B	0.03	0.12	0.99	0.30	0.42	0.98	0.51	0.55	0.98	0.66	0.62	0.97	0.80	0.68	0.97
C	0.72	0.61	1.00	0.93	0.80	0.99	0.97	0.85	0.99	0.98	0.87	0.98	0.99	0.89	0.98

Table 7 *The contribution of individual parameters to the convexity bias when using the representative firm approach.* It is a common approach to compare average actual spreads to model-implied spreads, where model-implied spreads are calculated using average firm variables. This introduces a bias because the spread in structural models is a non-linear function of firm variables. This table shows to what extent individual parameters in the Merton model contribute to the convexity bias. The individual firm variables are leverage ratio L, asset volatility σ_A, and firm payout rate δ. To explain how the numbers in the table are calculated, consider the number of 0.74 for asset volatility for a 10-year bond rated BBB. From Table 3 we find the 10% and 90% quantile of leverage ratio (0.19 and 0.80) and payout ratio (0.017 and 0.077) and set them to their average ($L = 0.495$ and $\delta = 0.047$). Using these averages we calculate the Merton model spread using the 10% respectively 90% quantile of asset volatility, 0.13 respectively 0.38, and the result is 156bps. We then compute the Merton model spread using the average of the 10% and 90% quantile of asset volatility and the result is 116bps. The ratio is between 116bps and 156bps is 0.74 and shown in the table. We proceed in the same way with all ratings, maturities, and firm variables. For AAA/AA-rated 1-year bonds, the spread was numerically 0 and no results are reported. We use a riskfree rate of 5%.

	2002Q3-2012Q2 Bond maturity			2002Q3-2007Q2 Bond maturity			2007Q3-2012Q2 Bond maturity		
	short	medium	long	short	medium	long	short	medium	long
$E[\mathbf{PD}(\theta_{it})]$									
AAA/AA	0.00	0.01	0.72	0.00	0.02	0.55	0.00	0.00	0.88
A	0.03	0.81	6.47	0.04	0.31	2.42	0.03	1.31	10.52
BBB	1.18	7.62	14.32	0.01	0.70	6.22	2.36	14.54	22.42
spec	1.29	3.58	22.70	1.87	2.14	16.35	0.00	6.45	29.05
$\mathbf{PD}(E[\theta_{it}])$									
AAA/AA	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.02
A	0.00	0.02	3.00	0.00	0.00	1.07	0.00	0.10	7.15
BBB	0.00	1.29	14.75	0.00	0.05	7.04	0.00	6.82	24.03
spec	0.00	0.14	27.42	0.00	0.14	19.91	0.00	0.13	34.85
$\mathbf{PD}(E[\theta_{it}])/E[\mathbf{PD}(\theta_{it})]$									
AAA/AA	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.02
A	0.00	0.03	0.46	0.00	0.01	0.44	0.00	0.07	0.68
BBB	0.00	0.17	1.03	0.00	0.07	1.13	0.00	0.47	1.07
spec	0.00	0.04	1.21	0.00	0.06	1.22	0.00	0.02	1.20

Table 8 *Convexity bias when calculating model-implied default probabilities in the Merton model using the representative firm approach.* It is a common approach to compare historical default frequencies to model-implied default probabilities, where model-implied default probabilities are calculated using average firm variables. This introduces a bias because the default probability in structural models is a non-linear function of firm value parameters. This table shows the magnitude of this bias. To calculate $E[\mathrm{PD}_T(\theta_{it})]$ we compute the volume-weighted average default probability implied by the Merton on a monthly basis and then average the monthly default probabilities over the sample period. To calculate $\mathrm{PD}_T(E[\theta_{it}])$ we compute the volume-weighted average firm variables (leverage ratio, asset volatility, payout ratio, bond maturity, riskfree rate at same maturity as bond), take the simple average of the monthly volume-weighted averages, and use the averaged firm variables to calculate the implied default probability in the Merton model. Bond transactions are grouped into groups where the issued bond has remaining maturity 0-1y (short), 1-4y (medium), or 4-30y (long).

	Bond maturity			Bond maturity			Bond maturity		
	short	medium	long	short	medium	long	short	medium	long
Panel A: Asset volatility									
	$E(\sigma_A)$			$\hat{\sigma_A}$			Ratio		
AAA/AA	0.21	0.22	0.21	0.57	0.38	0.29	2.71	1.74	1.39
A	0.23	0.24	0.24	0.53	0.34	0.28	2.25	1.40	1.15
BBB	0.26	0.29	0.28	0.60	0.40	0.28	2.29	1.41	0.99
spec	0.28	0.26	0.23	0.63	0.39	0.20	2.29	1.51	0.90
Panel B: Calibrating to average expected default probabilities									
	E(Spread)			$\widehat{\text{Spread}}$			Ratio		
AAA/AA	0	0	10	0	0	13	1.27	1.27	1.27
A	3	35	69	4	29	77	1.29	0.85	1.12
BBB	100	234	205	130	216	171	1.30	0.92	0.84
spec	113	98	285	147	110	237	1.30	1.13	0.83
Panel C: Calibrating to 10-year average expected default probabilities									
	E(Spread)			$\widehat{\text{Spread}}$			Ratio		
AAA/AA	0	0		0	0		0.00	0.01	
A	3	35		0	6		0.00	0.18	
BBB	100	234		0	37		0.00	0.16	
spec	113	98		0	0		0.00	0.00	

Table 9 *Bias in spread predictions and asset volatility when fitting the Merton model to average default probabilities.* It is a common approach to apply average firm variables (leverage ratio, payout ratio) to a representative firm, and back out asset volatility σ_A by matching expected default probabilities in a structural model to realized historical default frequencies. This introduces a bias in the asset volatility estimate because the default probability in structural models is a non-linear function of firm value parameters as Table 8 shows. Panel A shows the average asset volatility, $E(\sigma_A)$, and the biased implied asset volatility, $\hat{\sigma_A}$, using this approach. Panel B shows the bias when predicting spreads *for the same horizon* as the horizon to which default probabilities are calibrated. Panel C shows the bias of fitting expected default probabilities to average 10-year default probabilities and predicting spreads at various time horizons. We assume that realized default frequency equals average default probability, and calculate average default probability by computing the volume-weighted average default probability implied by the Merton on a monthly basis and then average the monthly default probabilities over the sample period. The asset risk premium is set to 4%. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE.

Systematic risk	mean	Quantiles						
ρ		0.005	0.025	0.25	0.5	0.75	0.975	0.995
0%	4.39%	4.02%	4.11%	4.29%	4.39%	4.49%	4.68%	4.78%
5%	4.39%	1.72%	2.16%	3.37%	4.21%	5.20%	7.62%	9.01%
10%	4.39%	1.07%	1.50%	2.91%	4.03%	5.46%	9.36%	11.87%
15%	4.41%	0.68%	1.07%	2.55%	3.85%	5.67%	10.84%	14.29%
20%	4.39%	0.45%	0.77%	2.24%	3.64%	5.73%	12.27%	16.76%
25%	**4.38%**	**0.28%**	**0.56%**	**1.94%**	**3.45%**	**5.81%**	**13.50%**	**18.80%**
30%	4.39%	0.18%	0.39%	1.69%	3.25%	5.84%	14.86%	21.17%
35%	4.38%	0.11%	0.27%	1.43%	3.02%	5.86%	16.19%	23.62%
40%	4.40%	0.06%	0.18%	1.23%	2.85%	5.88%	17.44%	26.04%
45%	4.39%	0.03%	0.12%	1.02%	2.59%	5.78%	19.02%	28.39%
50%	4.36%	0.02%	0.07%	0.82%	2.37%	5.73%	20.02%	30.20%

Table 10 *Statistical uncertainty of realized BBB default frequencies published by Moody's.* In a simulation experiment we estimate the statistical uncertainty of the average realized cumulative 10-year default frequency of 4.39% of BBB firms published by Moody's using data for 1970-1998. We assume that in year 1 there is a cohort of 1,000 firms. In year 2 there is a new cohort of 1,000 firms formed. This goes on for 18 years. All firms are identical when the cohort is formed and their 10-year default probability is 4.39%. Part of firm volatility is systematic and part is idiosyncratic according to equation (7). For each cohort, we calculate the realized default frequency on a 10-year horizon and calculate the average default frequency across all cohorts. Overall, the realized default rate is based on 18,000 firms over a period of 28 years. We repeat this simulation 100,000 times and the table shows the distribution of realized default rates for different amounts of systematic risk ρ. The results for $\rho = 25\%$ used in the main text are highlighted.

	AA	A	BBB	BB	B
4-year maturity					
Actual spread	10	41	103	274	424
Model spreads computed using def. rates 1970-1998	8	13	43	232	544
	(3;38)	(6;51)	(23;122)	(140;434)	(361;818)
Model spreads computed using def. rates 1920-2012	18	33	73	212	438
	(10;36)	(21;60)	(49;117)	(156;300)	(342;566)
10-year maturity					
Actual spread	22	49	101	273	423
Model spreads computed using def. rates 1970-1998	5	16	61	238	423
	(−7;114)	(−1;143)	(23;233)	(125;441)	(258;572)
Model spreads computed using def. rates 1920-2012	30	48	95	224	363
	(10;82)	(22;108)	(55;173)	(154;323)	(275;457)
20-year maturity					
Actual spread	28	58	139	-	-
Model spreads computed using def. rates 1970-1998	3	26	76	199	247
	(−13;157)	(−1;187)	(24;230)	(98;287)	(138;300)
Model spreads computed using def. rates 1920-2012	43	60	98	185	241
	(14;117)	(25;137)	(53;178)	(124;249)	(181;284)

Table 11 *Merton model estimates of corporate spreads to AAA rates using default rates from different periods and taking into account statistical uncertainty about default rates.* Based on the average ex post realized default frequency over some period, we calculate the likelihood function of the ex ante default probability and a 95% confidence band for the ex ante default probability (as explained in Section 4). We then calculate a Merton model spread based on the realized default frequency and a 95% confidence band (using equation (8)). Spreads are relative to the AAA yield with the same maturity and we ignore any statistical uncertainty in the historical AAA default frequency (this implies that the 2.5% quantile of the spread of say AA to AAA can be negative because the 2.5% quantile of ex ante default probability of AA can be lower than the AAA realized default frequency). Actual spreads for investment grade bonds are based on Duffee(1998) while actual spreads for speculative grade bonds are based on Caouette, Altman, and Narayanan(1998).

	2002Q3-2012Q2 Bond maturity			2002Q3-2007Q2 Bond maturity			2007Q3-2012Q2 Bond maturity		
	short	medium	long	short	medium	long	short	medium	long
AAA/AA									
Actual spread	−4	−4	12	−2	−5	6	−5	−1	28
	(−18;21)	(−20;29)	(−4;92)	(−12;18)	(−18;10)	(−3;27)	(−22;26)	(−21;45)	(−6;102)
Model spread	0	0	0	0	0	0	0	0	0
	(0;0)	(0;0)	(0;18)	(0;0)	(0;0)	(0;15)	(0;0)	(0;0)	(0;18)
A									
Actual spread	7	23	67	7	15	57	8	83	87
	(−8;97)	(−2;195)	(16;156)	(−6;39)	(−4;156)	(14;146)	(−11;153)	(14;279)	(28;259)
Model spread	0	0	29	0	0	26	0	3	34
	(0;0)	(0;70)	(0;193)	(0;0)	(0;59)	(0;184)	(0;1)	(0;375)	(0;339)
BBB									
Actual spread	55	121	100	27	85	95	286	479	454
	(14;1028)	(17;912)	(43;341)	(10;90)	(9;267)	(42;281)	(38;3751)	(120;2663)	(163;986)
Model spread	0	26	104	0	11	100	3	859	359
	(0;1123)	(0;1099)	(18;239)	(0;0)	(0;68)	(17;213)	(0;1775)	(35;1709)	(27;752)
spec									
Actual spread	128	209	415	82	166	399	405	558	422
	(48;546)	(−90;727)	(50;2023)	(36;228)	(−130;528)	(109;1775)	(126;711)	(−32;965)	(38;3644)
Model spread	0	0	309	0	1	236	0	0	311
	(0;0)	(0;296)	(76;441)	(0;1)	(0;240)	(40;390)	(0;0)	(0;321)	(237;482)

Table 12 *Actual and Merton-model yield spreads.* This table shows actual and model-implied industrial corporate bond spreads. Bond transactions are grouped according to remaining maturity at transaction date; 0-1y(short), 1-4y(medium), and 4-30y(long). 'Actual spread' is the volume-weighted median actual spread to the swap rate while 'Model spread' is the volume-weighted median spread implied by the Merton model. Below the spreads in parantheses are the 10% and 90% volume-weighted quantiles. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE.

Sample years	mean	Quantiles						
		0.005	0.025	0.25	0.5	0.75	0.975	0.995
20								
simulation	4.39%	0.13%	0.31%	1.50%	3.08%	5.87%	15.88%	23.49%
approximation	4.39%	0.14%	0.33%	1.56%	3.13%	5.84%	15.65%	22.93%
30								
simulation	4.39%	0.32%	0.61%	2.02%	3.51%	5.80%	13.15%	18.17%
approximation	4.39%	0.39%	0.71%	2.14%	3.59%	5.77%	12.61%	17.37%
50								
simulation	4.40%	0.70%	1.10%	2.60%	3.89%	5.65%	10.57%	13.74%
approximation	4.39%	0.85%	1.27%	2.73%	3.95%	5.56%	10.04%	12.94%
70								
simulation	4.39%	1.01%	1.45%	2.88%	4.03%	5.50%	9.38%	11.76%
approximation	4.39%	1.20%	1.64%	3.03%	4.08%	5.42%	8.89%	11.05%
90								
simulation	4.39%	1.25%	1.69%	3.08%	4.12%	5.41%	8.64%	10.51%
approximation	4.39%	1.45%	1.90%	3.21%	4.16%	5.31%	8.22%	9.97%
200								
simulation	4.39%	1.99%	2.41%	3.53%	4.27%	5.12%	7.06%	8.15%
approximation	4.39%	2.21%	2.61%	3.63%	4.29%	5.04%	6.75%	7.72%

Table A1 *Accuracy of closed-form approximation of the realized default loss distribution.* This table shows quantiles from the simulation of realized default losses and from a closed-form approximation. The quantiles from the simulation are obtained by simulating as explained in Section 4. The quantiles in the approximation are from the Vasicek (1991) loan loss probability distribution with a correlation coefficient given in equation (14). The default probability is $p = 4.39\%$ and correlation is $\rho = 0.25$.

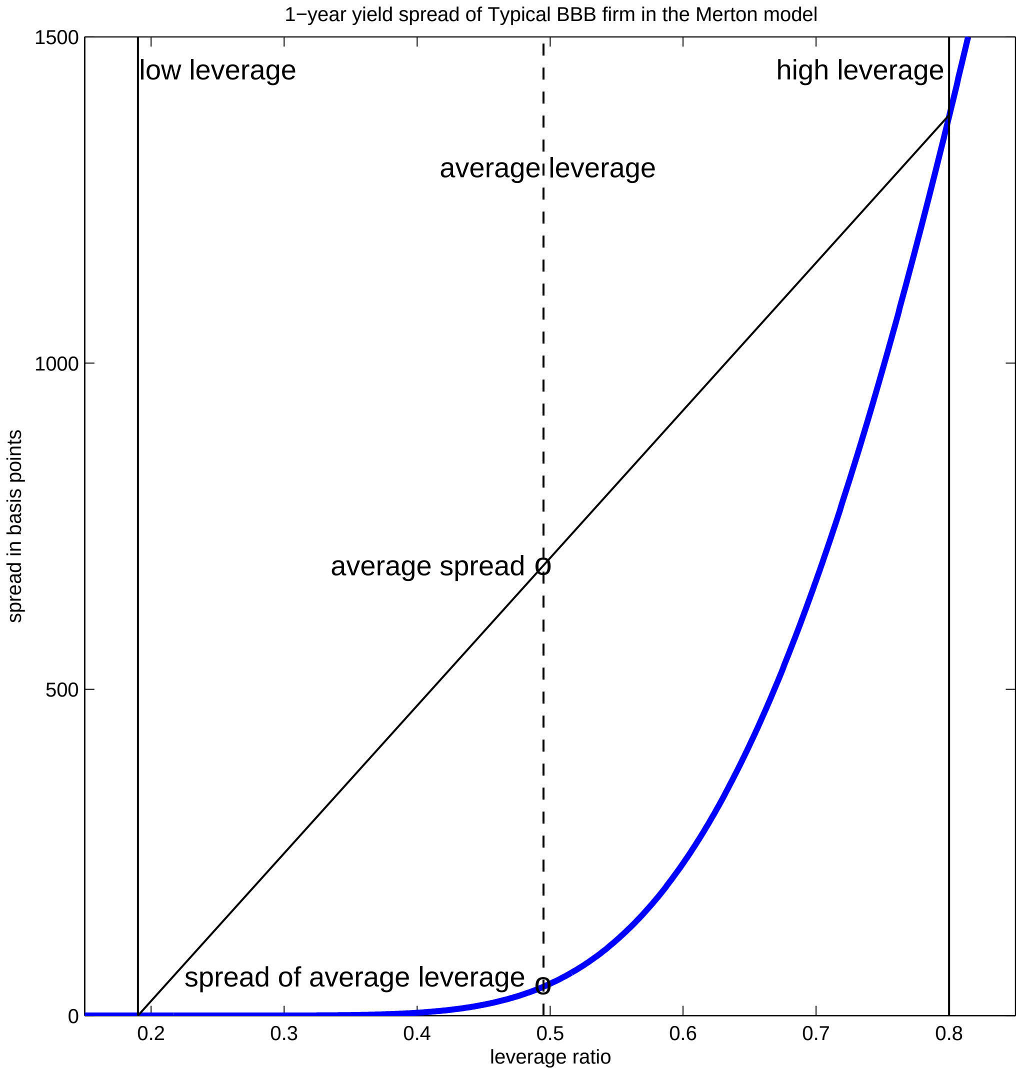

Fig. 1 *Convexity bias when calculating the spread in a structural model using average leverage and comparing it to the average spread.* It is a common approach to compare average actual spreads to model-implied spreads, where the model-implied spreads are calculated by using average firm variables. This introduces a bias because the spread in structural models is a non-linear function of firm variables. The figure illustrates the bias in case of two BBB-rated firms, one with a low leverage ratio and one with a high leverage ratio. Asset volatility is 28%, dividend yield 4.8%, recovery rate 49.2%, and riskfree rate 5%. 'Low leverage' is the 10% quantile (0.19) in the distribution of leverages in our sample while 'high leverage' is the 90% quantile (0.80). The quantities are from Table 3.

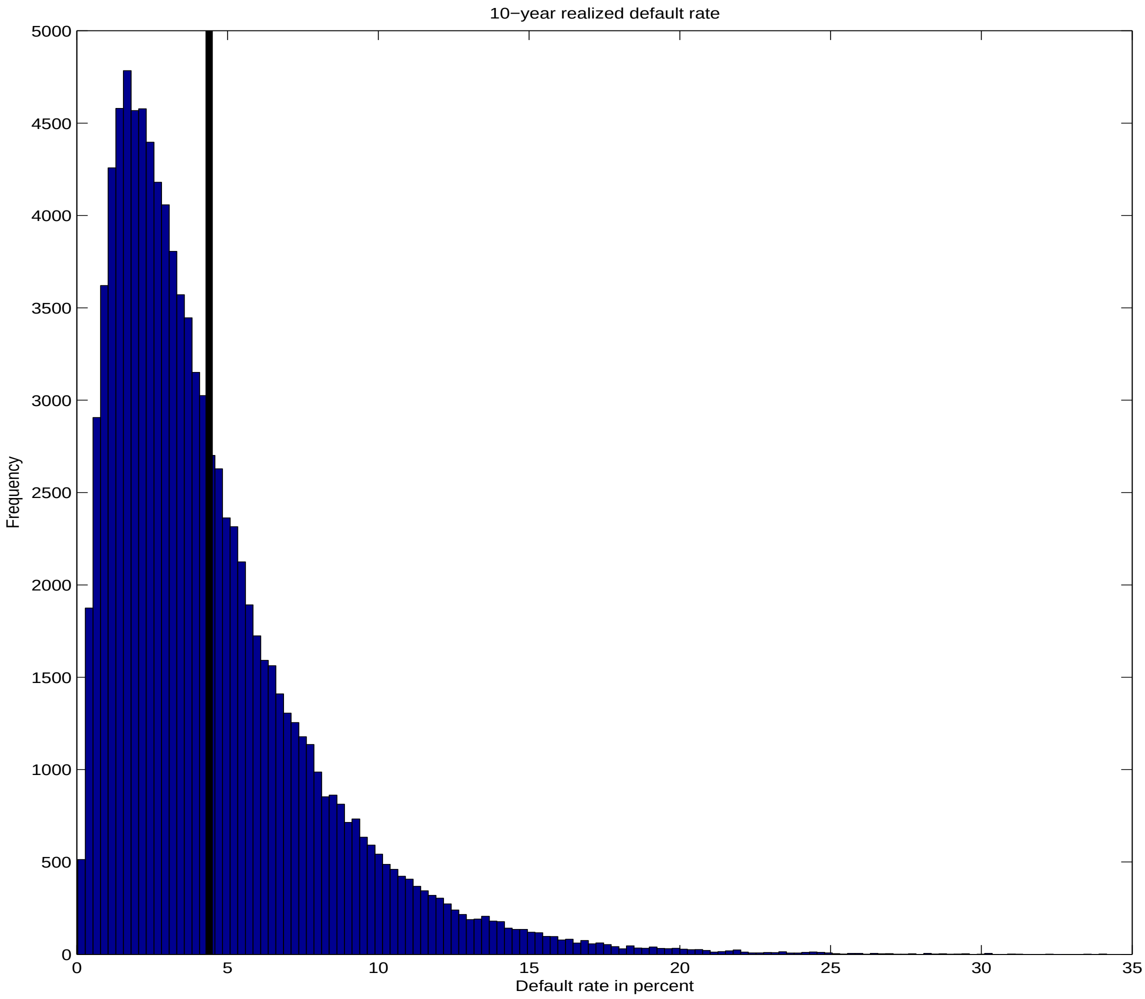

Fig. 2 *Distribution of realized default frequency.* In a simulation experiment we estimate the distribution of the average realized cumulative 10-year default frequency over 28 years given an ex ante default probability of 4.39%. We assume that in year 1 there is a cohort of 1,000 firms. In year 2 there is a new cohort of 1,000 firms formed. This goes on for 18 years. All firms are identical when the cohort is formed and their 10-year default probability is 4.39%. Part of firm volatility is systematic and part is idiosyncratic according to equation (7). For each cohort, we calculate the realized default frequency on a 10-year horizon and calculate the average default frequency across all cohorts. Overall, the realized default rate is based on 18,000 firms over a period of 28 years. We repeat this simulation 100,000 times and the graph shows the distribution of realized 10-year cumulative default rates. The solid line is the ex ante default probability of 4.39%.

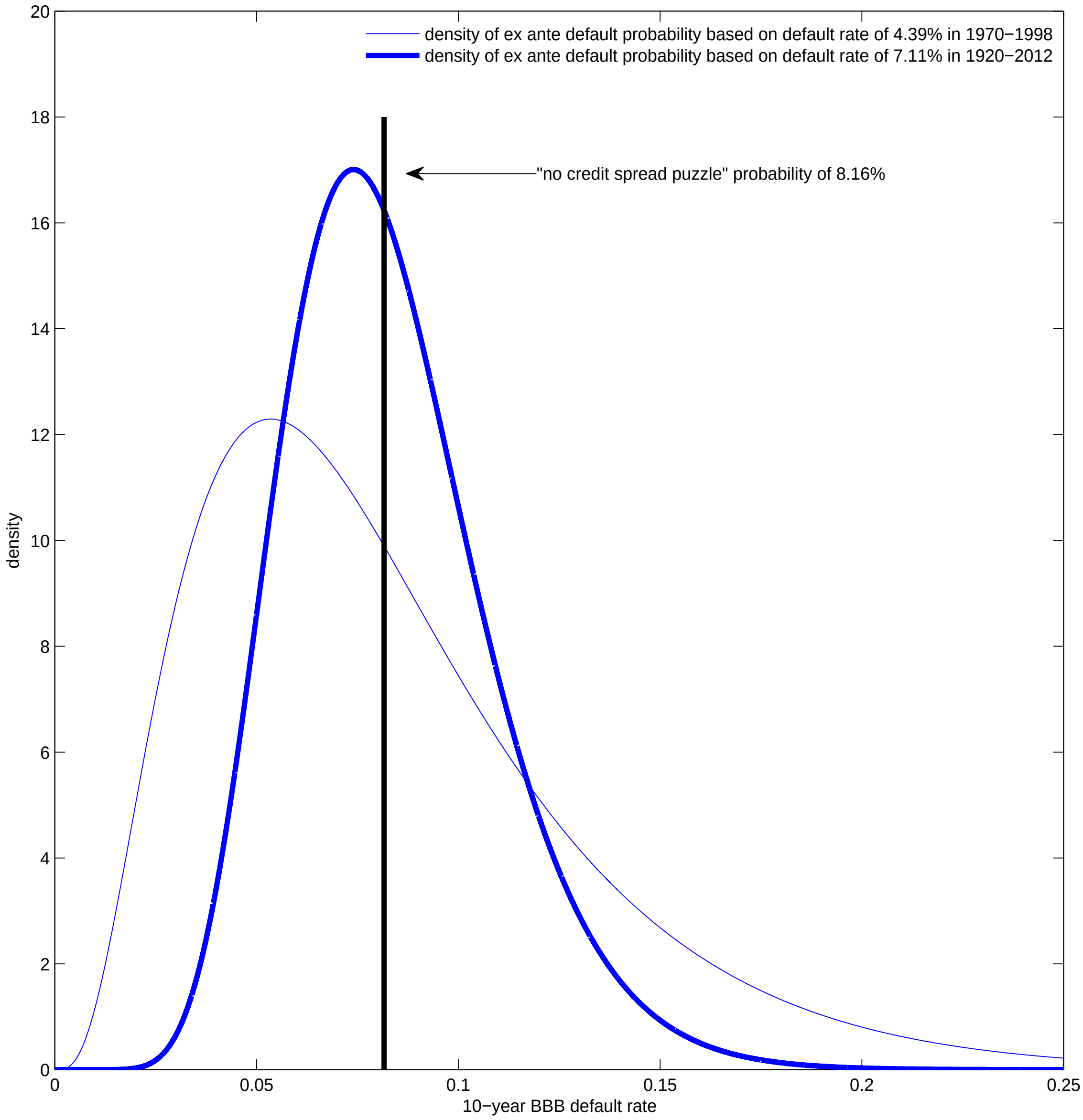

Fig. 3 *Density of ex ante default probability given the observed default rate history.* Based on the average ex post realized default frequency over some period, we can calculate the likelihood function of the ex ante default probability as explained in Section 4. The graph shows the likelihood function of the 10-year cumulative BBB default probability based on a realized rate of 4.39% over the period 1970-1998 and a rate of 7.112% over the period 1920-2012 as published by Moody's. The solid black line is the ex ante default probability of 8.16% that would make default probabilities and spreads consistent with each other in the Merton model.

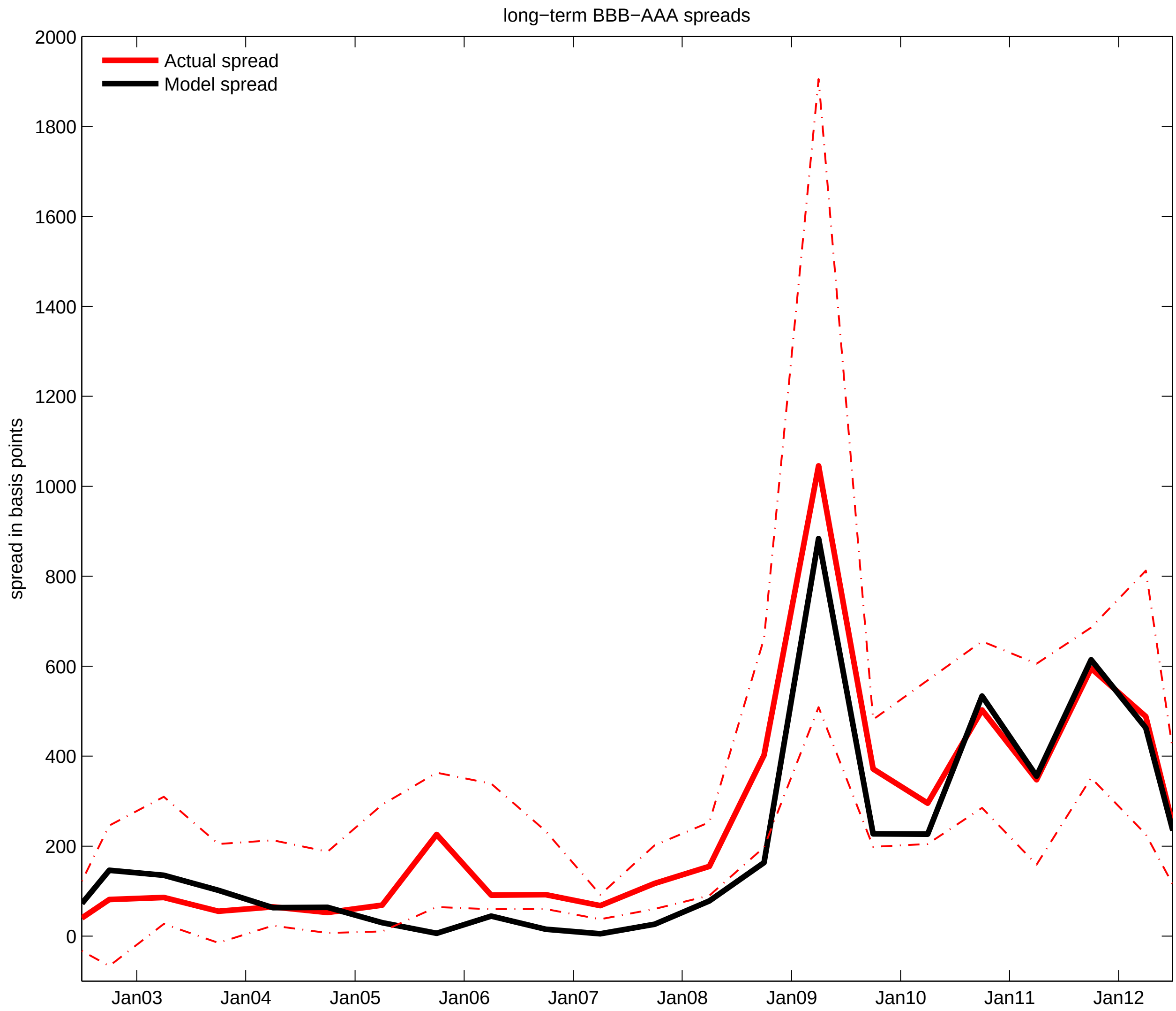

Fig. 4 *Long-term BBB-AAA corporate bond yield spreads.* This graph shows the time series variation of actual and model-implied long-term AAA-BBB spreads. On a semi-annual basis all transactions in bonds where the bond maturity is more than four years at transaction date and the rating is AAA/AA or BBB are collected. The graph shows the volume-weighted median BBB spread minus the volume-weighted median AAA/AA spread. Volume-weighted 10pct and 90pct quantiles are bootstrapped as explained in the text. The figure also shows the model-implied Merton spread, found by calculating the model-implied AAA-BBB spread computed in the same way as the actual spread. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE.

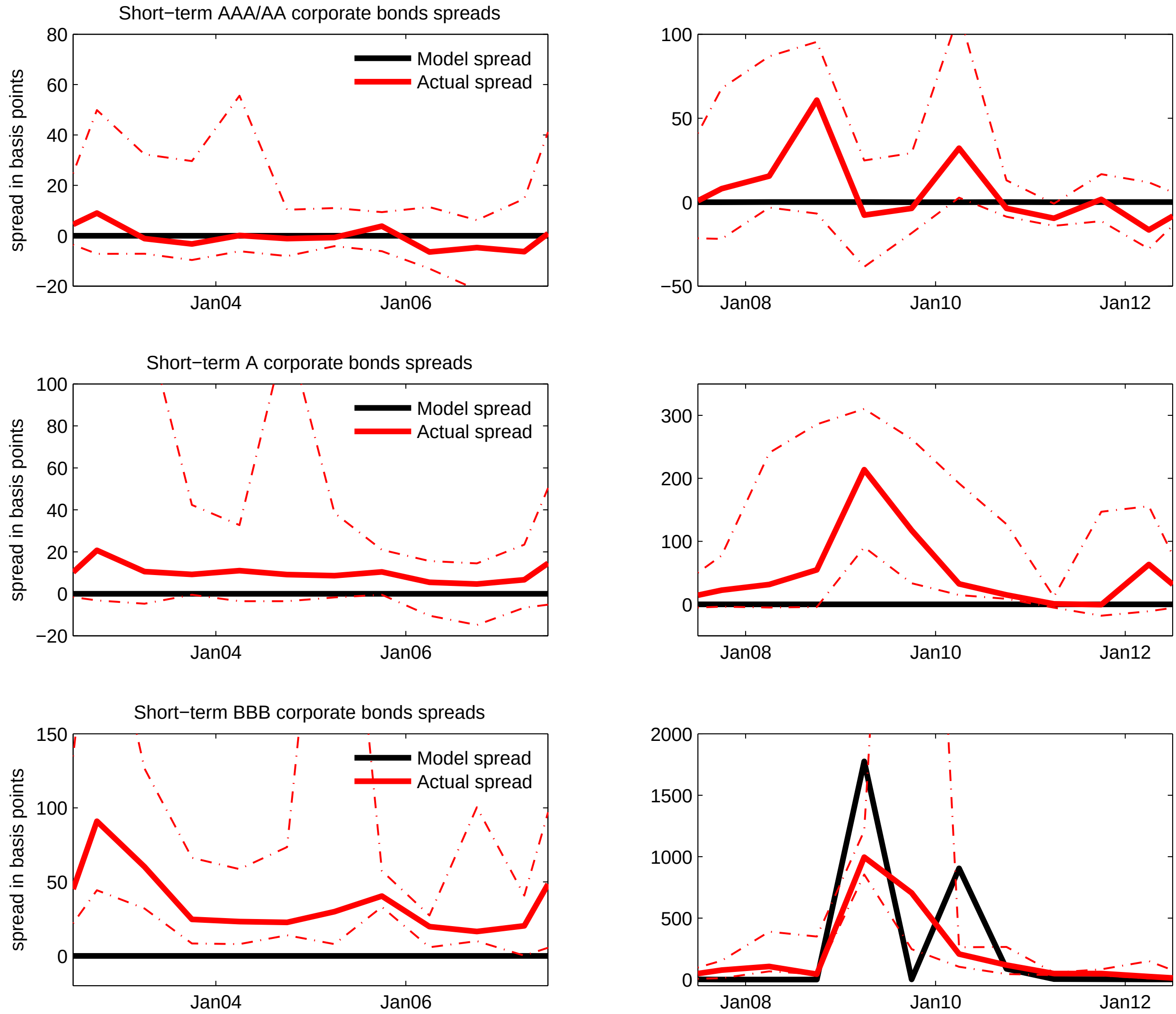

Fig. 5 *Short-term corporate bond yield spreads.* This graph shows the time series variation of actual and model-implied short-term industrial corporate bond spreads. On a semi-annual basis all transactions in bonds maturing within one year on transaction day are collected. The Figure shows - for ratings AAA/AA, A, and BBB - the volume-weighted median actual spread along with volume-weighted 10pct and 90pct quantile. The figure also shows the model-implied Merton spread, found by calculating the model-implied spread for each transaction and computing the volume-weighted median. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE.

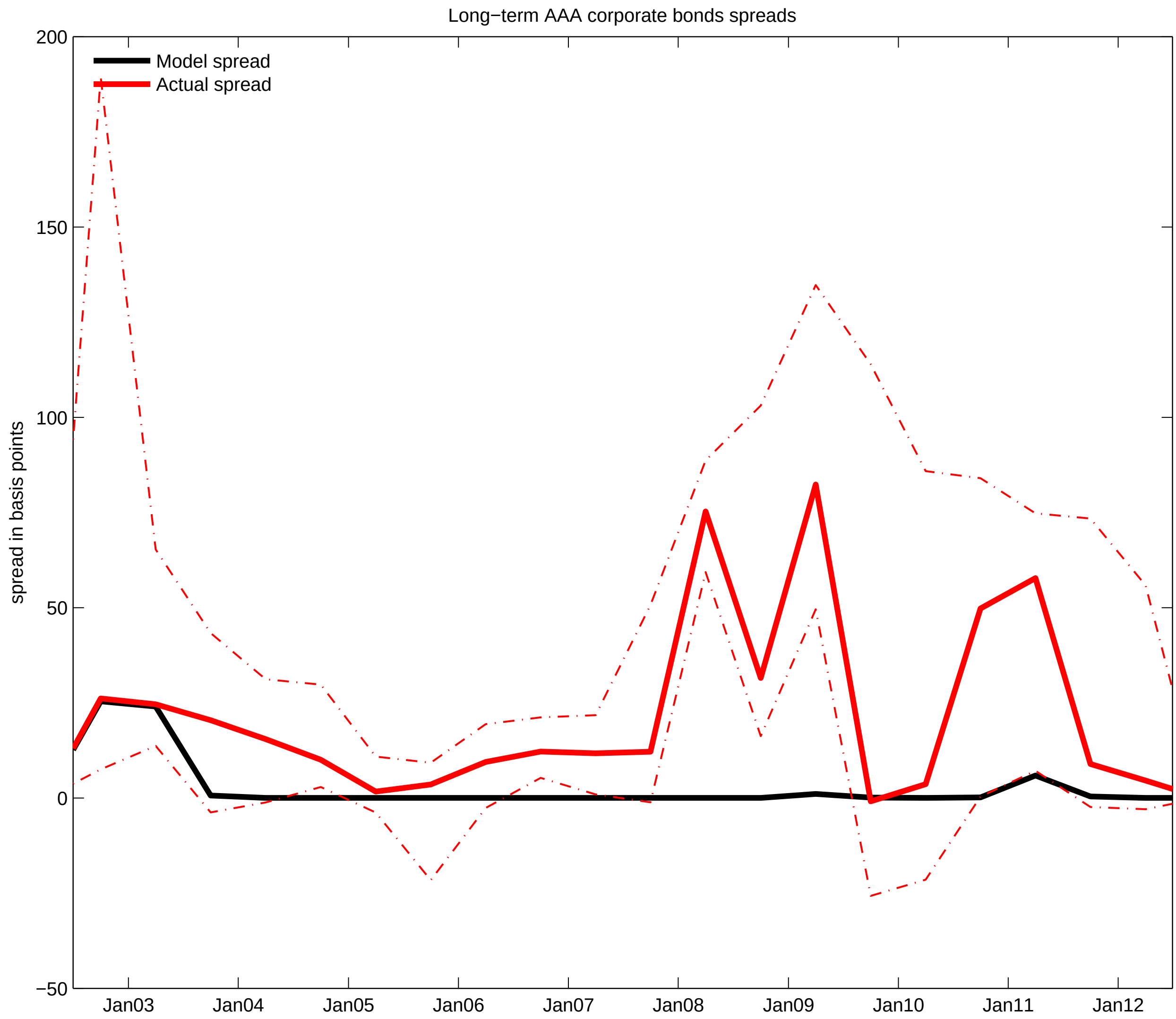

Fig. 6 *Long-term AAA corporate bond yield spreads.* This graph shows the time series variation of actual and model-implied long-term industrial AAA/AA corporate bond spreads. On a semi-annual basis all transactions in bonds with maturity between 4 and 30 years on transaction day are collected. The figure shows the volume-weighted median actual spread along with volume-weighted 10pct and 90pct quantiles. The figure also shows the model-implied Merton spread, found by calculating the model-implied spread for each transaction and computing the volume-weighted median. Bond transactions cover the period 2002Q3-2012Q2 and are from TRACE.