The Myth of the Credit Spread Puzzle *

Peter Feldhütter[†]
London Business School

Stephen Schaefer[‡]
London Business School

February 11, 2016

Abstract

We test the Merton model of credit risk using a much longer time series of data compared to previous studies and find that the model matches actual corporate bond spreads well. A crucial ingredient to the success of the model is that we use default rates for a long period of 92 years to calibrate the model. In simulations we show that such a long history of *ex post* default rates is essential to obtain estimates of *ex ante* default probabilities that have a reasonable level of precision; using default rates from shorter periods as in most existing studies will often lead to the conclusion that spreads in the Merton model are too low even if the Merton model is the true model. When we use default rates from 1920-2012, we show that the model matches investment grade spreads in any period we look at; 1920-2012, 1985-1995, and 1987-2012. Furthermore, we find that the model also captures the time series variation of the BBB-AAA spread over the period 1987-2012 well with a correlation between 84% and 92%. Using the calibration method described in the paper, our results likely hold not only for the Merton model but for a wide range of structural models.

Keywords: Credit spread puzzle, Structural models, Merton model, Corporate bond spreads, Default probabilities;

JEL: C23; G01; G12

* We are grateful to the Q-Group for the Jack Treynor Prize. We are also grateful for valuable comments and suggestions received from Antje Berndt (discussant), Harjoat Bhamra (discussant), Hui Chen (discussant), Mike Chernov, 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 NBER Asset Pricing 2015 conference, Utah Winter Finance Conference 2015, HKUST Finance Symposium 2014, WFA 2014, SFS Cavalcade 2014, ESSFM 2014 in Gerzensee, Fifth Risk Management Conference 2014, UNC's Roundtable 2013, UCLA Anderson School of Management, UC San Diego, Financial Conduct Authority, AQR, London Business School, Cass 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, Surrey Business School, 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

The structural approach to credit risk, pioneered by Merton (1974) and extended by Leland (1994) and others, represent the leading theoretical framework for studying corporate default risk and pricing corporate debt. While the models are intuitive and simple, many studies find that credit spreads calculated using structural models are lower than actual spreads, particularly for investment grade bonds, giving rise to the so-called credit spread puzzle.[1] Using a much longer history of data than previous studies, we use the Merton model as the lens through which to study this problem and find that the model is able to capture not only the level but also the time series variation of investment grade credit spreads. In other words that there is no credit spread puzzle. This result diverges from much of the literature and we explain carefully why we reach this different conclusion.

We start out by looking at average spreads within investment grade ratings and follow the approach in Chen, Collin-Dufresne, and Goldstein (2009) (CDG). A crucial input to their calculation of the spread is the default probability and CDG like others in the literature use average historical default rates based on Moody's default data starting from 1970. We confirm the results in CDG that with average default rates starting from 1970 investment grade spreads in the Merton model are too low compared to actual spreads.

Next, we show in simulations that a long history of *ex post* default rates is crucial in estimating *ex ante* default probabilities with a reasonable degree of precision. Using a "short" history of around 30 years, as in CDG and much of the literature, is likely to lead to the conclusion that the Merton model underpredicts spreads even if the model is correct. The reason is that the distribution of average default rates not only has a high level of dispersion, even when measured over several decades, but is also highly skewed. Most of the time we see few defaults but occasionally we see many defaults. Key to this result is that defaults are correlated as a result of the common dependence of individual firm values on systematic ("market") shocks. To see why correlation leads to skewness, we can think of a large number of firms with a default probability (over some period) of 5% and where their

[1]Papers finding that structural models underpredict credit spreads include Eom, Helwege, and Huang (2004), Leland (2006), Cremers, Driessen, and Maenhout (2008), Chen, Collin-Dufresne, and Goldstein (2009), Chen (2010), Huang and Huang (2012), and McQuade (2013).

defaults are perfectly correlated. We will see no defaults 95% of the time (and 100% defaults 5% of the time) so the realized default rate will underestimate the default probability 95% of the time. If an average default rate is calculated over two independent periods, the realized default rate will only underestimate the default probability $0.95^2 = 90.25\%$ of the time; thus a longer history reduces the skewness.

Since a long history of default rates is necessary to estimate default probabilities reasonably accurately, we revisit the results in Chen, Collin-Dufresne, and Goldstein (2009) using default rates from 1920 instead of from 1970. We find that when using default rates from 1920 the puzzle disappears, i.e. the Merton model matches average investment grade spreads.

Using default rates from 1920 instead of from 1970, the disappearance of the credit spread puzzle in recent decades could be due to rating agencies tightening their standards over time resulting in substantially lower expected default rates for a given rating.[2] However, the average BBB-AAA spread over 1920-2012 of 119bps is close to the average of 112bps over 1970-2012, suggesting that average ex ante default probabilities were actually very similar in the two periods. Furthermore, we test the Merton model without making any assumption on the stability of ratings by using default rates from 1920-2012 and spreads from the same period. Here, we find that the actual BBB-AAA spread of 119 basis points is remarkably close to the value we obtain in the Merton model (126 basis points). Thus, the model-implied Merton spread matches the actual BBB-AAA spread very well over almost a century.

Having established that the Merton model can match average spreads when calibrated to a long history of default rates, we put the model to a much harder test and ask if it is also possible to match the variation in spreads over time. We therefore implement the Merton model using firm-level data and study spreads of industrial firms over the period 1987-2012. Our dataset contains 286,234 individual bond yield observations. In this data we first confirm that the Merton model can capture average investment grade spreads; the average long-maturity BBB-AAA for example is 121bps in the Merton model and 113bps

[2] Since the periods for which we measure average spreads and default rates are long (30-90 years) our results are unlikely to be affected by the well-known phenomenon of the rating agencies rating "through the cycle".

in the data. For speculative grade bonds, average spreads in the Merton model are lower than those in the data which is consistent with an illiquidity premium for speculative grade bonds documented in Dick-Nielsen, Feldhütter, and Lando (2012). Then we calculate the average BBB-AAA spread on a monthly basis and we find that there is high degree of co-movement between actual and model-implied BBB-AAA spreads; the correlation is between 84% and 92% depending on maturity. Thus, the BBB-AAA spread – which has received most attention in the literature – is tracked well by the Merton model.

Our approach allows for cross-sectional and time-series variation in both firms' leverage and their asset volatility. At the same time we match long-run average default rates by finding the value of the default boundary (as a fraction of face value) such that the average model-implied default probability matches the average historical default rate (from 1920-2012). The fitted default boundary is then used to compute firm and time-specific spreads and default probabilities. This method combines two different approaches found in the literature. The first is to use a "representative" firm and match historical default rates. The second is to use individual firms, thus allowing for heterogeneity, while disregarding the level of model-implied default probabilities[3]. Departing from the first approach enables us to study time series variation in spreads, while matching default rates ensures consistency with historical experience.

The Merton model is only a convenient tool and our results have important implications for structural models in general. Huang and Huang (2012) show empirical that many structural models which appear very different in fact generate similar spreads once the models are calibrated to match historical default rates, recovery rates, and the equity premium (see also Chen, Collin-Dufresne, and Goldstein (2009)). The models tested in Huang and Huang (2012) include features such as stochastic interest rates, endogeneous default, stationary leverage ratios, strategic default, time-varying asset risk premia, and jumps in the firm value process. Our results together with the findings in Huang and Huang (2012) imply that not only the Merton model but a wide range of structural models of credit risk match historical

[3]See Cremers, Driessen, and Maenhout (2008), Chen, Collin-Dufresne, and Goldstein (2009), Chen (2010), and Huang and Huang (2012) for examples of the first approach and Eom, Helwege, and Huang (2004), Ericsson, Reneby, and Wang (2015), and Bao (2009) for the second approach.

corporate bond spreads once they are correctly calibrated to long-run default rates.

There is an extensive literature on testing the Merton model. Leland (2006), Cremers, Driessen, and Maenhout (2008), Chen, Collin-Dufresne, and Goldstein (2009), Chen (2010), and Huang and Huang (2012) use a representative firm to test variants of the Merton model and find that the model underpredicts spreads. In contrast to these papers, we calibrate to "long-term" rather than "short-term" default rates and report the time series variation of spreads. Eom, Helwege, and Huang (2004), Ericsson, Reneby, and Wang (2015), and Bao (2009) allow for heterogeneity in firms and variations in leverage ratios, but do not calibrate to historical default rates. Eom, Helwege, and Huang (2004) and Bao (2009) also do not look at the time series variation in credit spreads. When not calibrating to historical default rates, results become very sensitive to the choice of default boundary and there is no consensus in the literature as to the value of the boundary. In contrast we infer out the default boundary by matching historical default rates. It appears to us that much of the literature has not only not recognised the necessity of using a long history of default rates to estimate default probabilities with a reasonable degree of precision but has over-interpreted unreliable estimates of default rates based on a short (30-year) history. An exception is Bhamra, Kuehn, and Strebulaev (2010) who present a structural-equilibrium model with macro-economic risk. They simulate default rates over 5 and 10 years and find a wide distribution. In their model, uncertainty in default rates arises because of systematic cash flow risk, refinancing, and an endogenous default boundary that jumps when macro-economic conditions change according to a regime-switching model. We show that even with no refinancing, a constant default boundary, and average default rates measured over *several decades* there is nonetheless a large amount of variation in realized default rates.

The organization of the article is as follows: Section 2 examines the ability of the Merton model to match average spreads and highlights the importance of using a long history of default rates. Section 3 tests the ability of the Merton model to capture the time series variation of investment grade spreads using individual bond data for the period 1987-2012. Section 4 concludes.

2 The Merton model and average spreads

In this section we review the methodology used in the existing literature that finds a credit spread puzzle and argue why the puzzle does not exist. In line with much of the literature, we restrict our initial discussion to average spreads measured over long periods of time. Later in the paper we investigate the time series variation in credit spreads. We look at the credit spread puzzle through the lense of the Merton model, but as we will discuss later our results carry over to a wide range of structural models.

2.1 The Merton model

Asset value in the Merton model follows a Geometric Brownian Motion under the natural measure,

$$\frac{dV_t}{V_t} = (\mu - \delta)dt + \sigma dW_t^P \tag{1}$$

where δ is the payout rate to debt and equity holders, μ is the expected return on the firm's assets and σ 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 is below some default boundary, D, when the bond matures the firm cannot repay its bondholders and the firm defaults. For now, we can think of the default boundary as equal to the face value of debt as in Merton (1974), but later we relax this assumption. In case of default bondholders recovery a fraction R of the face value of debt.

Chen, Collin-Dufresne, and Goldstein (2009) derive a simple and compelling expression

[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).

for the credit spread as

$$(y-r) = -\Big(\frac{1}{T}\Big)\log\Big(1-(1-R)N\big[N^{-1}(\pi^P)+\theta\sqrt{T}\big]\Big) \tag{2}$$

where $\theta = \frac{\mu - r}{\sigma}$ is the Sharpe ratio on the assets of the firm. This formula is useful for two reasons. First, it depends on only three parameters: the natural default probability, Sharpe ratio, and recovery rate. Second, as we show in Appendix A, the relation between default probability and spread is approximately linear and therefore the spread computed using equation (2) with the average default rate as input is close to the average of the model spreads computed one-by-one, i.e.

$$\overline{y-r} \approx -\Big(\frac{1}{T}\Big)\log\Big(1-(1-R)N\big[N^{-1}(\overline{\pi^P})+\theta\sqrt{T}\big]\Big) \tag{3}$$

where $\overline{y-r}$ is the average of the model spreads computed one-by-one and $\overline{\pi^P}$ is the average default probability. Given an estimate of the Sharpe ratio θ and the recovery rate R, this allows us to compute the spread using equation (2) with the *average* default probability – proxied by the historical default rate – as input and compare this with the *average* historical spread.[5]

Much of the literature on the credit spread puzzle uses historical default rates published by rating agencies and categorized by rating. Rating agencies rate "through the cycle" and therefore the default probability of, say, a BBB-rated firm is typically higher in a recession than in a period of economic growth (see for example Altman and Rijken (2004)). Also, there is significant variation in default probability for firms with the same rating at any point in time. We show in Appendix A that for any sorting by rating and maturity we can test the relation in equation (2) using average historical spreads in place of $y-r$ and historical default rates in place of π^P. We stress that this test does not assume that the default probability in the sorting is constant across firms or over time. The literature has focused on sorting according to rating and there are at least two good reasons for this choice. First, average

[5]Note that although the relation in (2) between the spread and default probability is approximately linear, the relation between the spread and underlying parameters such as leverage and asset volatility is not. The relation between the average spread and the spread based on average parameters is therefore subject to a strong Jensen's inequality bias. See Feldhütter and Schaefer (2016).

spreads and default rates vary significantly across rating. Second, Moody's provide default rates as a function of rating as far back as 1920[6].

2.2 A long history of realized default rates is important

Using *ex-post* realized default rates in lieu of *ex ante* default probabilities when testing the Merton model through equation (2) is not unique to our paper. In fact, ex-post realized default rates play a crucial role in many empirical studies of credit spreads[7]. Even abstracting from the problem of secular variation in the default rate, there are two significant problems in obtaining reliable *ex-ante* default probabilities from realized default frequencies. The first is that the low level of default frequency, particularly for investment grade firms, leads to a sample size problem with default histories as short as those typically used in the literature (around 30 years). The second is that, even though the problem of sample size is potentially mitigated by existence of a large cross-section of firms, defaults are correlated across firms and so the benefit of a large cross-section in improving precision is greatly reduced.

While both these points may seem "obvious", their importance in the context of studies of credit pricing has been underestimated and, more to the point, the benefits of using a long history of defaults as a means of addressing both problems seems to have been overlooked.

There is a tradition in the literature of using realized default rates published by Moody's.[8] 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 the year 1970, Moody's define a cohort of BBB-rated firms and then report how many of these default over the next 10 years. The 10-year BBB default frequency for 1970 is the number of defaulted firms

[6]For any other sorting it would be necessary to calculate default rates using default databases and going back before 1979 has to our knowledge not been done, see for example Duffie, Eckner, Horel, and Saita (2009).

[7]See for example Leland (2006), Cremers, Driessen, and Maenhout (2008), Zhang, Zhou, and Zhu (2009), Chen, Collin-Dufresne, and Goldstein (2009), Chen (2010), Huang and Huang (2012), and McQuade (2013).

[8]All articles mentioned in Footnote 7 use Moody's estimates of default frequencies.

divided by the number in the 1970 cohort.[9] The average default rate of 4.39% is calculated as an average of the 10-year default rates for the cohorts formed at yearly intervals over the period 1970-1988[10].

To assess the statistical accuracy of the realized default rate of 4.39% as an estimate of the *ex ante* default rate we carry out a simulation. In an economy where the *ex ante* 10-year default probability is 4.39% for all firms, we simulate the ex post realized 10-year default frequency over 28 years. We assume that in year 1 we have 1,000 identical firms[11], where firm i's value under the natural measure follows the GBM in equation (1)

$$\frac{dV_t^i}{V_t^i} = (\mu - \delta)dt + \sigma dW_{it}^P. \tag{4}$$

We assume every firm has one T-year bond outstanding, and that default occurs if firm value is below bond face value at bond maturity, $V_T^i \leq F$. In the simulation $T = 10$. Using the properties of a Geometric Brownian Motion, the default probability is

$$p = P(W_{iT}^P - W_{i0}^P \leq c) \tag{5}$$

where $c = \frac{\log(F/V_0) - (\mu - \delta - \frac{1}{2}\sigma^2)T)}{\sigma}$. This implies that the unconditional default probability is $N(\frac{c}{\sqrt{T}})$ where N is the cumulative normal distribution. For a given default probability p we can always find c such that equation (5) holds, so in the following we use p instead of the underlying Merton parameters that give rise to p[12].

[9] 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.

[10] 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.

[11] 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. 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.

[12] Although only p is relevant for the simulation, we note that if we use median parameter values for BBB firms, $\sigma_A = 0.24$, $\delta = 0.045$, leverage=0.36, and $r = 0.05$ and a Sharpe Ratio of 0.22, the resulting default probability is 4.22%, very close to $p = 4.39\%$ used in the simulation.

We introduce systematic risk by assuming that

$$W_{iT}^{P} = \sqrt{\rho} W_{sT} + \sqrt{1-\rho} W_{iT} \tag{6}$$

where W_i is a Wiener process specific to firm i, W_s is a Wiener process common to all firms, and ρ is the pairwise correlation between percentage firm value changes. All the 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 (6).

In year 2 we form a cohort of 1,000 new firms. The firms in year 2 have characteristics that are identical to those of the previous firms at the point they entered the index in year 1. We calculate the realized 10-year default frequency of the year-2 cohort as we did for the year-1 cohort. Crucially, the common shock for years 1-9 for the year-2 cohort is the same as the common shock for years 2-10 for firms in the year-1 cohort. We repeat the same process 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. Finally, we repeat this entire simulation 100,000 times.

There are only two parameters in our simulation; the default probability p and the default correlation ρ. As mentioned we set $p = 4.39\%$. We assume $\rho = 0.25$; this is consistent with Cremers, Driessen, and Maenhout (2008) who find an average pairwise equity correlation of 25.4% for S&P 100 firms.

The top graph in Figure 1 shows the distribution of the realized default rate in the simulation study.[13] 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 the realized default rate can be far from the ex ante default probability. The reason is simply the presence of systematic risk in the economy which induces correlation in defaults among firms. If there is no systematic risk

[13] The distribution of the default frequency in Figure 1 is similar to the one derived analytically by Vasiček (1991). The reason that we cannot use Vasicek's result here is that the number reported by Moody's is the *average* of 10-year default rates while Vasicek's formula refers to the default rate over a single period. In a previous version of the paper we showed that if the correlation parameter in the Vasicek formula is set equal to the average correlation produced by Moody's overlapping cohorts this produces a good approximation to the distribution in Figure 1.

in the economy Table 1 shows that a 95% confidence interval for the realized default rate is $[4.11\%; 4.68\%]$.

We also see from Figure 1 that the default frequency is significantly skewed to the right, i.e., the modal value of around 2% is significantly below the mean of 4.39%. This means that the default frequency *most often* observed – e.g., the estimate from the rating agencies – is below the mean. This in turns implies that the number reported by Moody's (4.39%) is more likely to be below the true mean than above it and, in this case, if spreads reflect the true expected default rate, they will appear too high relative to the observed historical loss rate[14].

Moody's started to record default rates in 1920 so we can increase statistical precision 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. In the bottom graph in Figure 1 we therefore repeat the simulation where we maintain the ex-ante default probability at 4.39% but, instead of 28 years, we use 90 years in each simulated economy. We see that the distribution is tighter and more symmetric; the modal value is close to 4% and the width of the 95% confidence band of $[1.70\%; 8.61\%]$, although still wide, is almost half that when using 28 years of data[15]. Overall, a long history of realized default rates is necessary to estimate 10-year default probabilities with a reasonable degree of precision. We conclude that in those cases where historical default rates are employed in investigating the ability of structural models to price corporate bonds, using a long time series of defaults is of crucial importance and that estimates derived from periods of around 30 years are likely to be

[14]This is a point that Moody's KMV are aware of, see for example Kealhofer, Kwok, and Weng (1998) and Bohn, Arora, and Korablev (2005).

[15]Note that we hold p fixed at the historical default rate for the period 1970-1998, so $[1.70\%; 8.61\%]$ cannot be viewed as a confidence band for the 10-year BBB default probability for the long period because in such case we would need to set p equal to the default rate for the period 1920-2012–which is 7.11%–in the simulations. In this case the 95% band is $[3.10\%; 12.99\%]$.

unreliable.[16] We use these insights when investigating the credit spread puzzle next.[17]

2.3 The credit spread puzzle disappears when using a long history of default rates

Chen, Collin-Dufresne, and Goldstein (2009) (CDG) focus on the yield spread between BBB-rated bonds and AAA-rated bonds for bond maturities 4 and 10 years and – using equation (2) with a recovery rate of 44.9%, a Sharpe ratio of 0.22, and Moody's historical default rates from 1970-2001 – find that the Merton model underpredicts actual spreads. Figure 2 expands their results to all investment grade ratings and we see that the Merton model substantially underpredicts not only BBB-AAA spreads but all investment grade spreads.

In the previous section we showed that when using 31 years of default data as CDG do, the realized default rate most often observed is below the average default probability and it will appear as if there is a credit spread puzzle even in an economy where the Merton model is the correct model. Using a longer history of default rates mitigates this small-sample

[16] In the simulations we assume that the cohorts are different each year and so a given firm only appears in a single cohort. This is a simplification because when Moody's form cohorts of BBB-rated firms from year to year, there is a substantial overlap of firms from one cohort to the next. This overlap increases the correlation between the default frequencies in different cohorts in addition to that caused by systematic risk. If we allowed firms to stay in more than one cohort, the dispersion in the distribution of realised defaults would be larger. Also, the amount of systematic risk is assumed to be $\rho = 0.25$, but as Table 1 shows, even with low levels of systematic risk, there is considerable uncertainty regarding default rates. Finally, it is important to note that as the horizon over which default rates are calculated (10 years in the simulation) is increased the statistical uncertainty increases. Intuitively, over a 30-year horizon, there are six independent observations of 5-year default rates while there are only three independent observations of 10-year default rates.

[17] Although it is not the focus of their paper, Bhamra, Kuehn, and Strebulaev (2010) makes a related point. In their structural-equilibrium model with macro-economic risk, they simulate default rates over 5 and 10 years and find a wide distribution. In their model, uncertainty in default rates arises because of systematic cash flow risk, endogenous refinancing, and an endogenous default boundary that jumps when macro-economic conditions change according to a regime-switching model. We show that even with no refinancing, constant default boundary, and average default rates measured over *several decades* there is a large amount of variation in realized default rates.

issue[18] and Figure 2 shows the results from Chen, Collin-Dufresne, and Goldstein (2009) when we use default rates from the period 1920-2001 instead of 1970-2001. We see that once we use default rates from a longer period the puzzle goes away.

The spread data used in Figure 2 are average spreads 1985-1995 from Duffee (1998)[19]. One concern may be that Moody's have changed their rating methodology over time such that the quality of an investment grade bond is on average higher in the period 1985-1995 than in the early part of the 20th century. In this case BBB default rates would have gone down over time and it would be incorrect to use default rates from 1920 rather than from 1970 to estimate default probabilities when investigating spreads for the period 1985-1995. It would then also be the case that average spreads would have gone down substantially over time, but this is not what we observe in Moody's long-term BBB- and AAA-spreads.[20] The average BBB-AAA spread for the period 1920-1995 and 1970-1995 is 123bps respectively 117bps and for the period 1920-2012 and 1970-2012 is 119bps respectively 112bps.

We can also test the Merton model without making any assumptions on how ratings translate into default probabilities. We simply use default rates and spreads from the same period 1920-2012. Even if Moody's changed the "strictness" of their investment grade ratings, we can still use equation (2) because default rates and spreads are from the same (long) historical period.

Moody's report yields for bonds with a maturity of at least 20 years and their BBB-AAA spread for the period 1920-2012 is 119 basis points[21]. The 20-year default rate for the same

[18] Zhang, Zhou, and Zhu (2009) and Chen, Cui, He, and Milbradt (2015) also use default rates starting from 1920 in their calibration of structural models of credit risk.

[19] The spreads calculated in Duffee (1998) are commonly used in the literature, see for example Leland (2006), Chen (2010), Huang and Huang (2012), and McQuade (2013).

[20] We download the data from https://research.stlouisfed.org/fred2/data/BAA.txt and https://research.stlouisfed.org/fred2/data/AAA.txt. On a monthly basis since 1919 Moody's report the average yield on both AAA and BBB bonds with a maturity of at least 20 years.

[21] Some of the bonds are callable and Chen, Collin-Dufresne, and Goldstein (2009) estimate that the effect of the call options is to increase the spread by 7-15 basis points, so the option-adjusted spread is around 104-112 basis points. Because we do not make any adjustment this biases us towards finding a puzzle.

period 1920-2012 is 1.712% for AAA and 13.761% for BBB[22]. Throughout the paper, we use Chen, Collin-Dufresne, and Goldstein (2009)'s estimated Sharpe ratio of $\theta = 0.22$ and a recovery rate of 37.8% which is Moody's (2013)'s average recovery rate, as measured by post-default trading prices, for senior unsecured bonds for the period 1982-2012[23]. With a bond maturity of $T = 20$ the BBB and AAA spreads in the Merton model are – according to equation (2) – 167 and 42 basis points respectively, i.e. a BBB-AAA spread of 125 basis points. Thus the spread in the Merton model (125 basis points) based on default rates for the period 1920-2012 is remarkable close to the actual average spread (119 basis points) measured *over the same 92 years*[24].

2.4 Implication for other structural models

We focus our tests on the Merton model, but the model is only a convenient tool and our results have important implications for structural models in general.[25] Huang and Huang (2012) (HH) show that many structural models which appear very different in fact generate similar spreads once the models are calibrated to match historical default rates, recovery rates, and the equity premium. The models tested in HH include features such as stochastic interest rates, endogenous default, stationary leverage ratios, strategic default, time-varying asset risk premia, and jumps in the firm value process, yet all generate a similar level of credit spread. Chen, Collin-Dufresne, and Goldstein (2009) point out the significance of this finding and discuss the relation between the pricing kernel, the default time and LGD that would result in higher spreads.

[22]See Moody's (2013). We use the 20-year default rate because this is the longest maturity for which Moody's report default rates.

[23]Sharpe Ratios and recovery rates are similar in the first and second half of the 20th century: Chen, Collin-Dufresne, and Goldstein (2009) find very similar Sharpe ratios for the two periods 1927-2005 and 1974-1998. Furthermore, the average bond recovery rate in the period 1920-1996 is 40% according to Moody's, see Carty and Lieberman (1997).

[24]The 20-year default rate over 1920-2012 is based on cohorts from 1920-1992. We could compare this default rate with the average 20-year BBB-AAA spread 1920-1992 which is 125 basis points.

[25]We thank the referee for this very helpful suggestion.

From the no-arbitrage condition, the price of a corporate bond can be written as

$$P = \frac{1}{r}E[X_s] - cov(m_s, X_s) \tag{7}$$

where X_s is the bond cash flow in state s, m_s is the pricing kernel and r is the riskless rate. Calibrating different models to the same historical default rate and recovery rate will result in the first term in equation (7) being similar across models. Fixing the Sharpe ratio, recovery rate, and default rate will result in the second term being similar if different models result in similar values of the elasticity of the bond price to the firm's asset value. The implication of the striking results in HH is that for the wide range of models they considered, the elasticities are indeed similar. Therefore, all structural models calibrated to the historical recovery rate, the historical Sharpe ratio, and long-run default rates for the period 1920-2012 will generate spreads that are similar, both to each other and to the Merton model, and therefore consistent with historical spreads.

2.5 A summary of our results on average spreads

To sum up, while there will indeed appear to be a credit spread puzzle when using default rates only from 1970 onwards, we show that a long history of default rates is necessary to estimate default probabilities accurately. With this insight in mind we document that

1. when using *default rates starting from 1920* (instead of from 1970) the credit spread puzzle found in the existing literature goes away.

2. when using Moody's spreads and default data *both from 1920-2012* the BBB-AAA spread in the Merton model (125 basis points) matches the actual BBB-AAA spread (119 basis points) well.

3. these conclusions very likely apply to a wide range of structural models once these are calibrated to the same long-run historical default rates.

These results lead us to conclude that there is no credit spread puzzle.

3. The Merton model and time-series variation of spreads

In the previous section we showed that the Merton model is capable of capturing average spreads over longer periods of time when calibrated to a long history of default rates. In this section we put the Merton model to a harder test and ask if the model can also capture the time series variation of spreads. To examine the time series variation of spreads we need to model the time series variation of default probabilities. We do so by looking at individual firms over the period 1987-2012 and using our approach fixes their average default probability over 1987-2012 to the average default rate over 1920-2012.

3.1 Data

For the period January 1, 1997 to July 1, 2012, we use daily quotes provided by Merrill Lynch (ML) on all corporate bonds included in the ML investment grade and high-yield indices. These data are used by, among others, Schaefer and Strebulaev (2008) and Acharya, Amihud, and Bharath (2013). 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.[26] For the period April 1987 to December 1996 we use monthly data from the Lehman Brothers Fixed Income Database. This data is used by among others Duffee (1998), Huang and Huang (2012), and Acharya, Amihud, and Bharath (2013). We include only data that are actual quotes (in contrast to data based on matrix-pricing) and exclude bonds that are callable or contain sinking fund provisions. The Lehman database starts in 1973, but there are two reasons why we start from April 1987. First, there are few noncallable bonds before the mid-80s (see Duffee (1998)) and second, we do not have swap rates prior to April 1987. We use only bonds issued by industrial firms. The sample from ML contains 235,419 observations and the sample from Lehman 50,815 observations; in total we have 286,234 observations.

[26]For bond rating, we use the lower of Moody's rating and S&P's rating. If only one of the two rating agencies have rated the bond, we use that rating. We track rating changes on a bond, so the same bond can appear in several rating categories over time.

To price a bond in the Merton model we need the issuing firm's asset volatility, leverage ratio, and payout ratio along with the bond's recovery rate. *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.

An important parameter is the *asset volatility* and here 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:

$$\sigma^2 = (1-L)^2\sigma_E^2 + L^2\sigma_D^2 + 2L(1-L)\sigma_{ED}, \tag{8}$$

where σ is the volatility of assets, σ_D volatility of debt, σ_{ED} the covariance between the returns on debt and equity, and L is leverage ratio. If we assume that debt volatility is zero, asset volatility reduces to $\sigma = (1-L)\sigma_E$. 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 (8) in full. They find that for investment grade companies the two estimates of asset volatility are similar while for junk bonds there is a significant 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 (8) 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 $L > 0.75$[27]. This method has the advantage of being transparent and easy to replicate. All firm variables are obtained from CRSP and Compustat and details are given in Appendix B.

We set the recovery rate to 37.8% which is Moody's (2013)'s average recovery rate, as measured by post-default trading prices, for senior unsecured bonds for the period 1982-2012. 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.

[27]These fractions are based on Table 7 in SS apart from 1.80 which we deem to be reasonable. Results are insensitive to other choices of values for $L > 0.75$. See also Correia, Kang, and Richardson (2014) for an assessment of different approaches to calculating asset volatility.

3.2 Calibration

Recall from Section 2.1 that the firm defaults if the asset value falls below some default boundary when the bond matures, $V_T < D$. We let the default boundary be a fraction d of the face value of debt F such that $D = dF$. Chen, Collin-Dufresne, and Goldstein (2009) show that the default probability is

$$\pi^P = N\Big[-\Big(\frac{1}{\sqrt{\sigma^2 T}}\Big)\Big(\log\big(\frac{V_0}{dF}\big) + \big(\mu - \delta - \frac{\sigma^2}{2}\big)T\Big)\Big] \tag{9}$$

while the risk-neutral probability of default π^Q is

$$\pi^Q = N\Big[-\Big(\frac{1}{\sqrt{\sigma^2 T}}\Big)\Big(\log\big(\frac{V_0}{dF}\big) + \big(r - \delta - \frac{\sigma^2}{2}\big)T\Big)\Big]$$

and the credit spread is given as

$$(y - r) = -\frac{1}{T}\log\big[1 - (1 - R)\pi^Q\big].$$

The level of the default boundary plays an important role in spread predictions: holding other parameters constant a higher default boundary leads to higher default probabilities and thus higher spreads. Direct estimates of the default boundary are difficult to obtain and a range of estimates has been used in the literature.[28] Huang and Huang (2012) show that holding default probabilities fixed at historical averages eliminates the dependence of the spread on the default boundary. In their implementation they fix the default boundary and imply out asset volatility by matching historical default rates. We follow their approach, but since the default boundary is harder to estimate than asset volatility, we imply out the default boundary instead of asset volatility.

Specifically, if we observe a spread observation of bond i with a time-to-maturity T issued by firm j on date t, we calculate the firm's T-year default probability $\pi^P(d, \Theta_{jt}, T)$ using equation (9). Here, Θ_{jt} contains the time-t estimate of asset volatility, payout rate, and leverage ratio.[29] For a given rating a and maturity T - rounded up to the nearest integer

[28] Davydenko (2013) estimate the default boundary using market values of debt and equity, but his estimates depend on difficult-to-estimate costs of default.

[29] For simplicity we omit that the default probability depends on the riskfree rate and Sharpe ratio because they are not firm dependent.

year - we find all N_{aT} bond observations in the sample with the corresponding rating and maturity and calculate the average default probability $\overline{\pi}^P_{aT}(d)$. We denote the corresponding historical default frequency as $\hat{\pi}^P_{aT}$ and use Moody's historical default frequencies for the period 1920-2012. For all investment grade ratings and horizons of 1-20 years (Moody's only report default rates for up to 20 years horizon) we find the value of d that minimizes the weighted sum of absolute errors between the historical and model-implied default rates by solving

$$\min_{\{d\}} \sum_{a=AAA}^{BBB} \sum_{T=1}^{20} N_{at} |\overline{\pi}^P_{aT}(d) - \hat{\pi}^P_{aT}|. \tag{10}$$

We use Chen, Collin-Dufresne, and Goldstein (2009)'s estimated Sharpe ratio of 0.22 and compute the default boundary for each calendar year from estimated values of $\overline{\pi}^P$ in that year. Our estimate of d is then the average of the 26 yearly estimates from 1987 to 2012. This average value is 1.036 which implies that the default boundary is close to the face value of debt, as in the Merton model, and we therefore set $d = 1$[30] [31]. By implying out the default boundary from historical default rates while allowing for heterogeneity in firms and leverage

[30]It may appear that, together, a default boundary of 1 and a recovery rate of 0.378 imply a large deadweight cost of bankruptcy, but this misses an important observation made by Bao (2009). Much corporate debt is bank debt that has a much higher recovery rate; Acharya, Bharath, and Srinivasan (2007) find that the average (median) recovery rate for bank debt is 81% (92%). Furthermore, much of total debt is bank debt; Houston and James (1996) find that 64% of total debt is bank debt. This implies that modest deadweight costs of bankruptcy measured as percentage of firm value can lead to low recovery rates on bonds. For example, suppose that the face value of debt is 100 of which bank debt is 64 and bond debt is 36, and 22% of firm value is lost at default. If the firm value is 100 at default and 22 is lost, this leaves 64 for the bank debt holders (because they are senior) and 14 for bond holders. Thus a 22% deadweight cost of bankruptcy leads, in this case, to a loss of 20 (36 - 16) or 61% of face value for bond holders.

[31]The Sharpe Ratio of 0,22 estimated in Chen, Collin-Dufresne, and Goldstein (2009) is based on the Treasury rate. Since we use swap rates as riskfree rates, the estimated Sharpe ratio should be calculated with respect to the swap rate. Over our sample period, 1987-2012, the average spread between 1-month LIBOR and 1-month Treasury yield is 42 basis points. If we assume that firm specific equity volatility is 40%, the Sharpe ratio adjusted for the 42 basis points is 0.21. The estimated default boundary in Section 3.2 is 1.036 of the face value of debt with a Sharpe Ratio of 0.22 and 1.025 with a Sharpe Ratio of 0.21. Since we set the boundary to 1, adjusting the Sharpe Ratio has no effect on our results and we therefore leave it at the often-used value of 0.22.

ratios, we combine two strands of literatures examining structural models of credit risk. The first strand matches historical default rates but uses an average "representative firm" (Cremers, Driessen, and Maenhout (2008), Chen, Collin-Dufresne, and Goldstein (2009), Chen (2010), and Huang and Huang (2012)). Departing from the representative firm assumption allows us to study the time series variation of spreads. The second strand allows for cross-sectional heterogeneity in firms, but does not target historical default rates (Eom, Helwege, and Huang (2004), Ericsson, Reneby, and Wang (2015), and Bao (2009)).

3.3 Summary statistics

Summary statistics for the firms in our sample are shown in Table 2. Since the sample contains few AAA-rated firms, particular in the later years, we combine AAA and AA into one rating group and refer to the combined group as AAA-rated bonds later in the paper. The average leverage ratios of 0.14 for AAA/AA, 0.28 for A, and 0.38 for BBB are similar to those found in other papers.[32] The payout ratio is similar across ratings, implying that higher coupon payments for riskier firms are offset by smaller dividend payments, with an average across ratings of 4.7%. Average equity volatility is monotonically increasing with rating, consistent with a leverage effect. The estimates are similar to those in SS for A-AAA ratings, while the average equity volatility for BBB firms of 0.38 is higher than the value of 0.33 estimated in SS. Asset volatilities are slightly increasing in rating consistent with the estimates in SS.

In Table 3 we see that the number of bonds with a low rating of B or C is small relative to the number of investment grade bonds. One reason for this is that speculative grade bonds frequently contain call options 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 place all speculative grade bonds in one group.

[32] 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.

3.4 Investment grade corporate bond credit spreads

Table 4 shows actual and model-implied bond spreads to AAA-rated bonds in our sample. To calculate an average spread, we first calculate a monthly average spread and then calculate the average spread across months. Credit spreads for investment grade ratings A and BBB for medium and long maturity bonds are on average well matched by the Merton model. For example, the average long maturity BBB-AAA spread is 121 basis points in the data and 113 in the Merton model. The empirical estimates of the BBB-AAA spread are slightly higher than the 100-110 basis points typically found in the literature and reflects the 2008-2009 crisis which saw the highest corporate bond spreads since the Great Depression; before 2008 actual spreads are close to those found in earlier studies. Overall, the correspondence between actual and model-implied investment grade spreads in the three sub-samples is good. These results confirm our finding in Section 2 that the Merton model matches average spreads once the model is calibrated to a long history of default rates[33].

We next examine whether the model also captures the time series variation of credit spreads. The spread between BBB and AAA-rated bonds has received most attention in the literature and we therefore focus on this spread. In each month, we calculate the average BBB yield spread and subtract the average AAA yield spread and Figure 3 plots the time series variation in this BBB-AAA spread. The model-implied spread tracks the actual spread well for all maturities: the correlation is 86% for long-maturity bonds, 92% for medium-maturity bonds, and 84% for short-maturity bonds. If we restrict the sample to the period from 1987 to 2007, the correlations are 82% (long-maturity), 75% (medium-maturity), and 31% (short-maturity) and so, apart from the results for short maturities, the correlations are not driven by the large spikes during 2008-2009. Overall, we find that the Merton model captures both the average level and time series variation of the BBB-AAA spread.

[33]If we calibrate the model to default rates from 1970-2012 the default boundary is 92.6% of the face value of debt and with this default boundary credit spreads will on average be 34.1% lower.

3.5 Speculative grade corporate bond credit spreads

For some maturities there are few speculative grade bonds in the sample compared to investment grade bonds (Table 3). Among short-maturity bonds for example there are no C-rated bonds and only 15 B-rated bonds. We therefore group all speculative grade bonds into one rating category in Table 4. The table shows that actual speculative grade bond spreads are higher than those implied by the Merton model; the average long-term speculative yield spread for example is 435 basis points in the data and 248 basis points in the Merton model.

To investigate in more detail the ability of the Merton model to capture speculative grade yield spreads, we compute the average difference between actual and model-implied yield spreads for rating notches. Each broad rating category has three rating notches and because we aggregate all yield spreads for bonds with a maturity between zero and 30 years we have sufficient observations to be able to compute the average spread difference by rating notch. Figure 4 shows the results and the vertical line separates investment-grade ratings on the left-hand side from speculative-grade ratings on the right-hand side. For investment grade ratings the average difference between actual and model-implied yield spreads is small for all rating notches and ranges from -16 to +14 basis points. For speculative grade ratings the actual spread is substantially larger than the model-implied spread and the difference ranges from 67 to 473 basis points[34]. Thus, as we move from a low investment grade rating to a high speculative grade rating the Merton model starts to underpredict spreads.

3.6 Spread between the AAA yield and the riskless rate

Traditionally, Treasury yields have been used as riskfree rates, but recent evidence shows that swap rates are a better proxy than Treasury yields. A major reason for this is that Treasury bonds enjoy a convenience yield that pushes their yields below riskfree rates (Feldhütter and Lando (2008), Krishnamurthy and Vissing-Jorgensen (2012), and Nagel (2014)). Hull, Predescu, and White (2004) find that the riskfree rate is 63bps higher than Treasury yields

[34]The results are not sensitive to the choice of sample period. If we split the sample period into three subperiods of equal length, the average difference between actual and model-implied spreads is positive in 8 out of 18 instances for investment grade bonds (six rating notches times three time periods) while the difference is positive for 17 of 18 instances for speculative grade bonds.

and 7bps lower than swap rates, Feldhütter and Lando (2008) find riskfree rates to be 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. However, on occasion, swap rates also deviate significantly from riskfree rates. Feldhütter and Lando (2008) find that during 2002-2003 swap rates were influenced by supply and demand factors and were pushed 40bps below the riskfree rate. More recently, long-term swap rates became extraordinary low following the Lehman default in September 2008. For example, the 30-year swap spread to Treasury was around 50bps in the two years before the Lehman default, but went negative and stayed negative until the end of the sample period. This implies that although recent papers find swap rates to be more appropriate riskfree rates than Treasury yields in general, swap rates are poor proxies for riskfree rates in certain periods, and it is difficult to identify a single rate that is a good proxy for the riskfree rate at all times.

Figure 5 plots the time series variation in the spread between AAA-yields and both the swap rate and the Treasury yield along with the model-implied AAA bond spread. When we use Treasury yields as the riskfree rate, the AAA-riskfree spread is substantially underpredicted by the Merton model for all maturities. For example, the average long-term AAA-Treasury spread is 77bps and only 10bps in the Merton model.

There is also an underprediction of long-term AAA spreads when using the swap rate as the riskfree rate, but the difference is relatively small. For long-maturity bonds, the difference in average spreads is 14bps (the actual spread to the swap rate is 25 bps versus 11bps in the Merton model), while for medium and short-maturity bonds the difference between the average empirical and model-implied spread is 12bps respectively 14bps. Figure 5 shows that for most of the sample period, AAA-spreads to the swap rate are not far from zero. The median spread to swap is 10bps for short maturity, 6bps for medium maturity, and 16bps for long maturity. This implies that when the swap rate is used as the riskfree rate, high-quality corporate bond spreads–particularly at shorter maturities–are close to zero most of the time and only deviate substantially around the recessions of 2001 and 2008-2009. It has previously been regarded as a failure of the Merton model that high-quality short-term spreads are close to zero, but these results suggest that most of the time this corresponds to the spreads to the swap rate that we actually observe.

3.7 The role of bond illiquidity

Our finding that the Merton model captures corporate bond credit spreads well may appear inconsistent with the findings in Dick-Nielsen, Feldhütter, and Lando (2012) (DFL) and Bao, Pan, and Wang (2011) that there are times where corporate bond spreads contain a significant illiquidity premium[35]. However, our results are in line with these findings and to see this we focus on the contribution of illiquidity to yield spreads in long-maturity bonds in DFL's Table 4. There we see that in normal times (Panel A) the contribution of illiquidity to investment grade spreads to AA is less than 3bps (here and in the following we use DFL's estimates for AA-rated bonds because most of the bonds in our AA/AAA category are AA-rated). During the 2008-2009 crisis DFL find an illiquidity premium in AA-rated (BBB-rated) bonds to be 65bps (98bps), but the difference of 33bps is minor relative to the size of spreads during this period, so again the contribution of illiquidity to investment grade spreads is small, because the illiquidity premium is "differenced out". Overall, the results in DFL suggests that our finding that investment spreads relative to AAA are captured by the Merton model is due to small liquidity premiums outside the 2008-2009 crises and due to "differencing out" liquidity premiums in the 2008-2009 crisis. Consistent with this view, we find that actual AAA-swap spreads are close to model-implied spreads outside the 2008-2009 crisis, but the Merton spread is substantially below actual AAA-swap spreads during 2008-2009.

The contribution of illiquidity to the spread between speculative grade bonds and AA-rated bonds is estimated in DFL to be 83bps outside 2008-2009, which suggests that a significant part of the underprediction for speculative grade spreads is due to an illiquidity premium. DFL's finds the illiquidity part of the spec-AA spread to be a sizeable 178bps during the 2007-2008 which is consistent with the spread underprediction in Table 4 of 539-358=181bps during 2007-2012.

[35] See also Bao and Pan (2013).

4. Conclusion

We test the Merton model of credit risk using U.S. corporate bond spread data from 1920-2012, 1987-2012, and 1985-1995 and we use both average spreads and spreads from individual firms. In all cases we find that model spreads match actual investment grade credit spreads well. As a further test of the model, we find that the time series of the model-implied BBB-AAA spread tracks the actual BBB-AAA spread well with a correlation between 84% and 92% depending on bond maturity.

Crucial for our conclusions is that we calibrate the model to empirical default rates that are calculated using a long period, 1920-2012. In simulations we show that such a long history of default rates is essential to estimate expected default probabilities with a reasonable degree of reliability; using default rates from short periods of around 30 years as frequently done in the literature will often lead to the conclusion that spreads in the Merton model are too low even if the Merton model is the true model.

Our results show that the credit spread puzzle in the Merton model - the perceived failure of the model to explain levels of credit spreads particularly for investment grade bonds - has less to do with deficiencies of the model than with the way in which it has been implemented. While we only test the Merton model, our results have strong implications for a wide range of structural models. Huang and Huang (2012) show that many structural models that are calibrated to match a given recovery rate and default probability generate similar spreads. Among the features that these structural models include are stochastic interest rates, endogenous default, stationary leverage ratios, and strategic default. Our results combined with Huang and Huang (2012)'s findings suggest that all these structural models can match the average level of investment grade credit spreads when calibrated – as we argue they should be – to long-run default rates.

A The use of average default rates to calculate average spreads

In this Appendix we elaborate on the discussion in the main text regarding the use of the formula (2) to compute the average spread in the Merton model from the average default rate. Formula (2) is:

$$(y - r) = -\left(\frac{1}{T}\right) \log\left(1 - (1 - R)N\left[N^{-1}(\pi^P) + \theta\sqrt{T}\right]\right)$$

where $y - r$ is the corporate bond spread, R the recovery rate, θ the Sharpe ratio and π^P the default probability.

A.1 Bias

We suggest in the main text that the spread calculated from formula (2) using the *average* default probability is close to the average spread. Chen, Collin-Dufresne, and Goldstein (2009) show that this is the case in a simple numerical example. We use simulation to provide a more extensive analysis. The simulations are carried out as follows.

For each investment grade rating a and maturity t between one and 20 years we simulate 10,000 default probabilities from a normal distribution truncated at 0 and 1. The mean μ_{at} in the normal distribution is the long-run default rate for 1920-2012 obtained from Moody's (2013). In principle we could estimate the standard deviation from the individual model estimates of default probability used in our calibration to obtain the default boundary. However, because the distribution of bond maturities varies considerably over time, this results in estimates of the standard deviation that exhibit a great deal of variation by maturity. We therefore add a second step and, for each rating a, smooth the variation in the estimated standard deviation across maturities by fitting a second-order polynomial to their logs. For each maturity we use the exponential of the fitted value as the estimated standard deviation. We fit to logs to ensure positive standard deviations and we use observations from April 1987 to March 1995 to be consistent with Duffee (1998)'s data used in Figure 1. This procedure also allows us to estimate missing values.

For each rating and maturity we then calculate the spread, s_1, using formula (2) and the average simulated default probability (together with a recovery rate of 37.8% and a Sharpe ratio of 0.22). We also calculate spreads using the individual default probabilities and compute the average spread, s_2. Table A1 shows the difference $s_2 - s_1$. This difference is a measure of bias induced by calculating the spread using the average default rate.

We see that there is a modest upward bias when using the average default probability to calculate the average spread. For example, the 10-year BBB spread is upward biased by around five basis points which is small relative to a total spread of 142 basis points (not reported). The bias is less than eight basis points for all ratings and maturities and therefore not economically important for our analysis in Section 2. Note that our analysis in Section 3 is not affected by this bias since there we compare average actual spreads with the directly calculated average model-implied spreads.

Overall the results show that the bias when using formula (2) in conjunction with average default probabilities is small.

maturity	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AAA	0.3	0.2	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AA	0.3	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A	2.1	2.5	3.0	3.6	4.2	4.9	5.5	6.2	6.5	7.0	7.3	7.4	7.2	6.8	6.1	5.6	4.6	3.6	2.7	1.9
BBB	1.6	1.7	1.8	2.1	2.4	2.7	3.4	3.9	4.4	4.8	5.1	5.2	5.1	4.8	4.3	3.7	3.0	2.3	1.7	1.2

Table A1 *Bias (in basis points) when calculating spread using average default rate instead of calculating average spread.* For each rating r and maturity t, we simulate 10,000 default probabilities from a normal distribution $N(\mu_{rt}, \sigma_{rt})$ (truncated at 0 and 1), calculate the spread s_1 using the average simulated default probability and calculate the spread s_2 by computing the average spread where each spread is derived using an individual simulated default probability. The calculated spread is based on the Merton model using the formula $(y-r) = -\left(\frac{1}{T}\right)\log\left(1-(1-R)N\left[N^{-1}(\pi^P)+\theta\sqrt{T}\right]\right)$ where T is maturity, $R = 0.378$ is the recovery rate and $\theta = 0.22$ is the Sharpe ratio. The average μ_{it} in the simulations is the long-run historical default rate 1920-2012 obtained from Moody's(2013). The standard deviation σ_{rt} in the simulation is the standard deviation of default probabilities derived as explained in Appendix A. The table shows the the difference $s_1 - s_2$ (in basis points) and this measures the resulting bias when using the average historical default rate to calculate an average Merton spread.

B Firm data

To compute bond prices in the Merton model we need the issuing firm's leverage ratio, payout ratio, and asset volatility. This Appendix gives details on how we calculate these quantities using CRSP/Compustat.

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. If the payout ratio is larger than 0.13, three times the median payout in the sample, we set it to 0.13.

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.

References

Acharya, V., Y. Amihud, and S. Bharath (2013). Liquidity risk of corporate bond returns: conditional approach. *Journal of Financial Economics 110*(2), 358–386.

Acharya, V., S. Bharath, and A. Srinivasan (2007). Does industry-wide distress affect defaulted bonds? Evidence from creditor recoveries. *Journal of Financial Economics 85*, 787–821.

Altman, E. and H. Rijken (2004). How rating agencies achieve rating stability. *Journal of Banking and Finance 28*, 2679–2714.

Bao, J. (2009). Structural Models of Default and the Cross-Section of Corporate Bond Yield Spreads. *Working paper*.

Bao, J. and J. Pan (2013). Bond Illiquidity and Excess Volatility. *Review of Financial Studies 26*, 3068–3103.

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

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*.

Carty, L. V. and D. Lieberman (1997). Historical Default Rates of Corporate Bond Issuers, 1920-1996. *Special Comment. New York: Moody's Investors Services*, 1–32.

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

Chen, H., R. Cui, Z. He, and K. Milbradt (2015). Quantifying Liquidity and Default Risks of Corporate Bonds over the Business Cycle. *Working Paper*.

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.

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

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.

Davydenko, S. A. (2013). When do firms default? A Study of the Default Boundary. *Working paper*.

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., A. Eckner, G. Horel, and L. Saita (2009). Frailty correlated default. *Journal of Finance 64*, 2089–2123.

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., J. Reneby, and H. Wang (2015). Can structural models price default risk? Evidence from bond and credit derivative markets. *Quarterly Journal of Finance 5*(3).

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

Feldhütter, P. and S. Schaefer (2016). Structural models and the term structure of credit spreads. *Work in progress*.

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

Houston, J. and C. James (1996). Bank Information Monopolies and the Mix of Private and Public Debt Claims. *Journal of Finance*, 1863–1889.

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.

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. (1994). Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance 49*, 157–196.

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

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.

Moody's (2013). Annual Default Study: Corporate Default and Recovery Rates, 1920-2012. *Moody's Investors Service*, 1–64.

Nagel, S. (2014). The Liquidity Premium of Near-Money Assets. *Working paper*.

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.

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.

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 1 *Statistical uncertainty of realized BBB default frequencies published by Moody's.* The table shows the distribution of the realized cumulative 10-year default frequency when the mean cumulative 10-year default rate is equal to 4.39%, the rate for BBB-rated issuers published by Moody's using data for 1970-1998. We assume that in year 1 there is a cohort of 1,000 firms and that in year 2 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 (see equation (6)). The degree of systematic risk is determined by ρ. For each cohort, we calculate the realized default frequency on a 10-year horizon and then 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 levels of systematic risk ρ. The results for $\rho = 25\%$ used in the main text are highlighted.

	#firms	Mean	10th	25th	Median	75th	90th
			Leverage ratio				
AAA	15	0.14	0.07	0.08	0.09	0.15	0.24
AA	69	0.14	0.05	0.09	0.12	0.17	0.23
A	201	0.28	0.13	0.17	0.24	0.35	0.48
BBB	228	0.38	0.19	0.26	0.36	0.47	0.59
BB	119	0.45	0.18	0.29	0.45	0.59	0.72
B	54	0.51	0.21	0.32	0.47	0.70	0.81
C	7	0.80	0.63	0.73	0.85	0.93	0.96
all	458	0.33	0.11	0.18	0.29	0.44	0.60
			Equity volatility				
AAA	15	0.20	0.16	0.17	0.19	0.23	0.26
AA	69	0.27	0.17	0.21	0.25	0.32	0.38
A	201	0.31	0.20	0.24	0.30	0.37	0.42
BBB	228	0.38	0.24	0.28	0.35	0.44	0.56
BB	119	0.46	0.25	0.31	0.42	0.54	0.71
B	54	0.50	0.29	0.33	0.46	0.68	0.79
C	7	0.69	0.31	0.64	0.72	0.76	1.01
all	458	0.35	0.21	0.25	0.32	0.41	0.54
			Asset volatility				
AAA	15	0.17	0.14	0.15	0.16	0.20	0.24
AA	69	0.23	0.14	0.18	0.23	0.27	0.32
A	201	0.23	0.14	0.17	0.22	0.28	0.34
BBB	228	0.25	0.15	0.19	0.24	0.31	0.37
BB	119	0.28	0.15	0.19	0.25	0.34	0.43
B	54	0.27	0.13	0.19	0.25	0.33	0.42
C	7	0.21	0.04	0.06	0.19	0.31	0.44
all	458	0.25	0.15	0.18	0.23	0.30	0.35
			Payout ratio				
AAA	15	0.034	0.012	0.015	0.027	0.048	0.069
AA	69	0.043	0.017	0.030	0.043	0.053	0.066
A	201	0.047	0.020	0.031	0.043	0.056	0.078
BBB	228	0.050	0.018	0.029	0.045	0.065	0.093
BB	119	0.046	0.020	0.027	0.040	0.057	0.080
B	54	0.045	0.019	0.030	0.040	0.058	0.073
C	7	0.064	0.040	0.049	0.056	0.093	0.101
all	458	0.047	0.019	0.029	0.043	0.059	0.081

Table 2 *Firm summary statistics.* For each bond yield observation, the leverage ratio, equity volatility, asset volatility, and payout ratio are calculated for the issuing firm on the day of the observation. 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, calculated as explained in the text. Payout ratio is yearly interest payments plus dividends plus share repurchases divided by firm value. Firm variables are computed using data from CRSP and Compustat.

Short bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	22	71	197	140	60	15	0	447
Mean number of bonds pr quarter	3.17	4.11	9.13	6.34	3.29	1.19	NaN	22
Mean number of quotes pr quarter	6.548	66.46	136.8	112.8	69.71	18.57	NaN	356.7
Age	3.13	5.70	5.58	6.79	6.57	4.34	NaN	6.07
Coupon	7.53	5.65	6.01	7.28	7.67	5.35	NaN	6.58
Amount outstanding ($mm)	287	414	374	367	314	242	NaN	367
Time-to-maturity	1.24	1.23	1.21	1.23	1.27	1.15	NaN	1.23

Medium bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	22	85	256	202	82	22	3	587
Mean number of bonds pr quarter	4.55	4.94	12.9	10.6	5	1.63	1	33.6
Mean number of quotes pr quarter	9.895	80.53	159.9	155.1	99.1	20.81	37	480.8
Age	2.42	4.60	5.00	6.15	5.19	3.98	3.97	5.29
Coupon	7.38	5.79	6.53	7.39	7.62	8.63	11.86	6.94
Amount outstanding ($mm)	278	379	288	349	334	219	207	328
Time-to-maturity	2.98	2.90	2.95	3.03	2.99	3.07	3.44	2.98

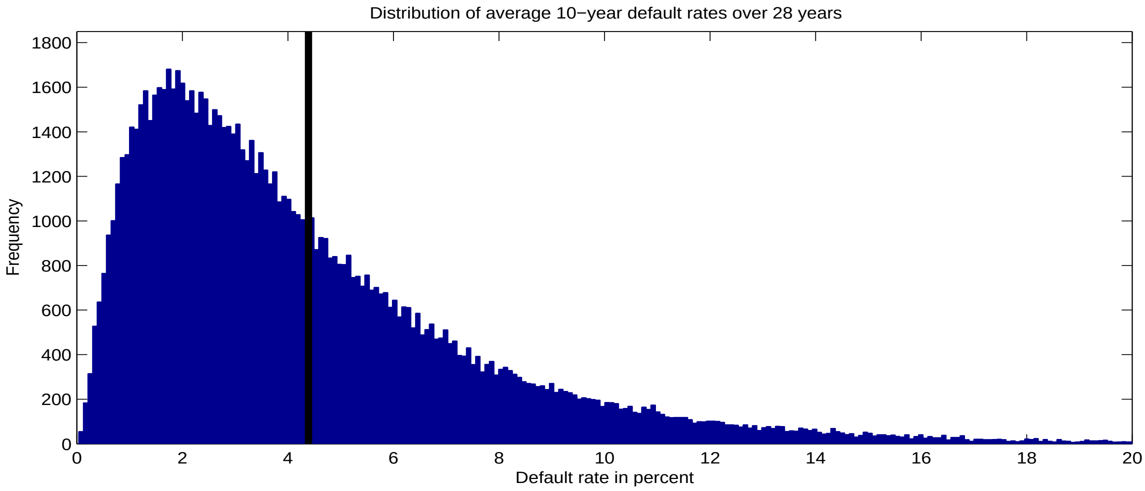
Long bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	39	203	666	629	226	69	8	1445
Mean number of bonds pr quarter	7.25	20.3	65.4	51	17.3	4.27	1.64	158
Mean number of quotes pr quarter	36.53	157.1	480.2	473.4	197.9	60.86	64.85	1382
Age	9.73	5.32	5.41	5.28	4.13	5.01	6.85	5.24
Coupon	7.72	6.95	7.44	7.75	7.79	9.01	10.46	7.64
Amount outstanding ($mm)	521	296	281	314	450	261	298	319
Time-to-maturity	13.32	8.32	11.25	11.10	8.34	9.33	10.75	10.43

Table 3 *Bond summary statistics.* The sample consists of noncallable bonds with fixed coupons issued by industrial firms. Short, medium, and long bond maturities are bonds with a maturity of 0-2, 2-4, and 4-30 years. This table shows summary statistics for the data set. Bond yield quotes cover the period 1987Q2-2012Q2.

	1987Q2-2012Q2			1987Q2-1997Q1			1997Q2-2007Q1			2007Q2-2012Q2		
	Bond maturity			Bond maturity			Bond maturity			Bond maturity		
	short	medium	long	short	medium	long	short	medium	long	short	medium	long
A												
Actual spread	39	32	40	16	19	22	23	28	38	103	66	74
Model spread	14	36	51	6	26	20	6	20	50	38	81	111
BBB												
Actual spread	141	117	121	57	63	64	130	93	118	311	253	230
Model spread	109	112	113	31	42	46	45	77	109	341	296	240
spec												
Actual spread	321	537	435	276	469	253	222	576	562	495	588	539
Model spread	219	306	248	86	338	162	173	144	276	476	475	358

Table 4 *Actual and Merton-model yield spreads.* This table shows actual and model-implied industrial corporate bond yield spreads. Spreads are grouped according to remaining bond maturity at the quotation date; 0-2y(short), 2-4y(medium), and 4-30y(long). 'Actual spread' is the average actual spread to the swap rate minus the average actual AAA/AA spread to the swap rate. 'Model spread' is the difference in average Merton model spreads between the bonds in a given maturity/rating bucket and bonds rated AAA or AA. The average spread is calculated by first calculating the average spread across bonds in a given month and then calculating the average of these spreads over months.



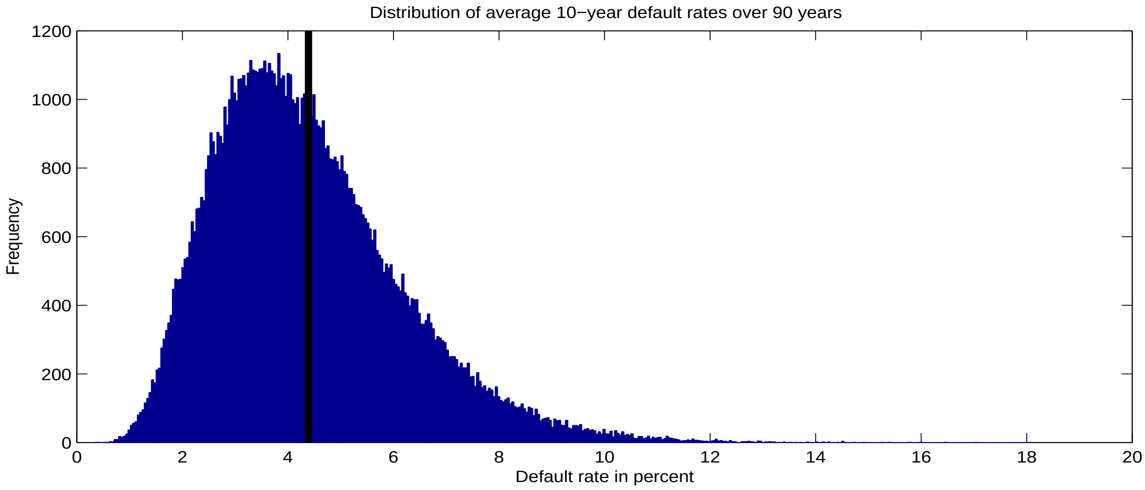


Fig. 1 *Distribution of realized default rate.* The figure shows the results of a simulation of the realized cumulative 10-year default rate when the mean cumulative 10-year default rate is equal to 4.39%, the rate for BBB-rated issuers published by Moody's using data for 1970-1998. We assume that in year 1 there is a cohort of 1,000 firms and that in year 2 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 (see equation (6)). The degree of systematic risk is determined by ρ. For each cohort, we calculate the realized default rate on a 10-year horizon and then calculate the average default rate 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 top graph shows the distribution of realized 10-year cumulative default rates. The solid line is the ex ante default probability of 4.39%. The bottom graph shows the distribution when the realized default rate is based on 90 years (but the ex ante default probability remains equal to 4.39%.

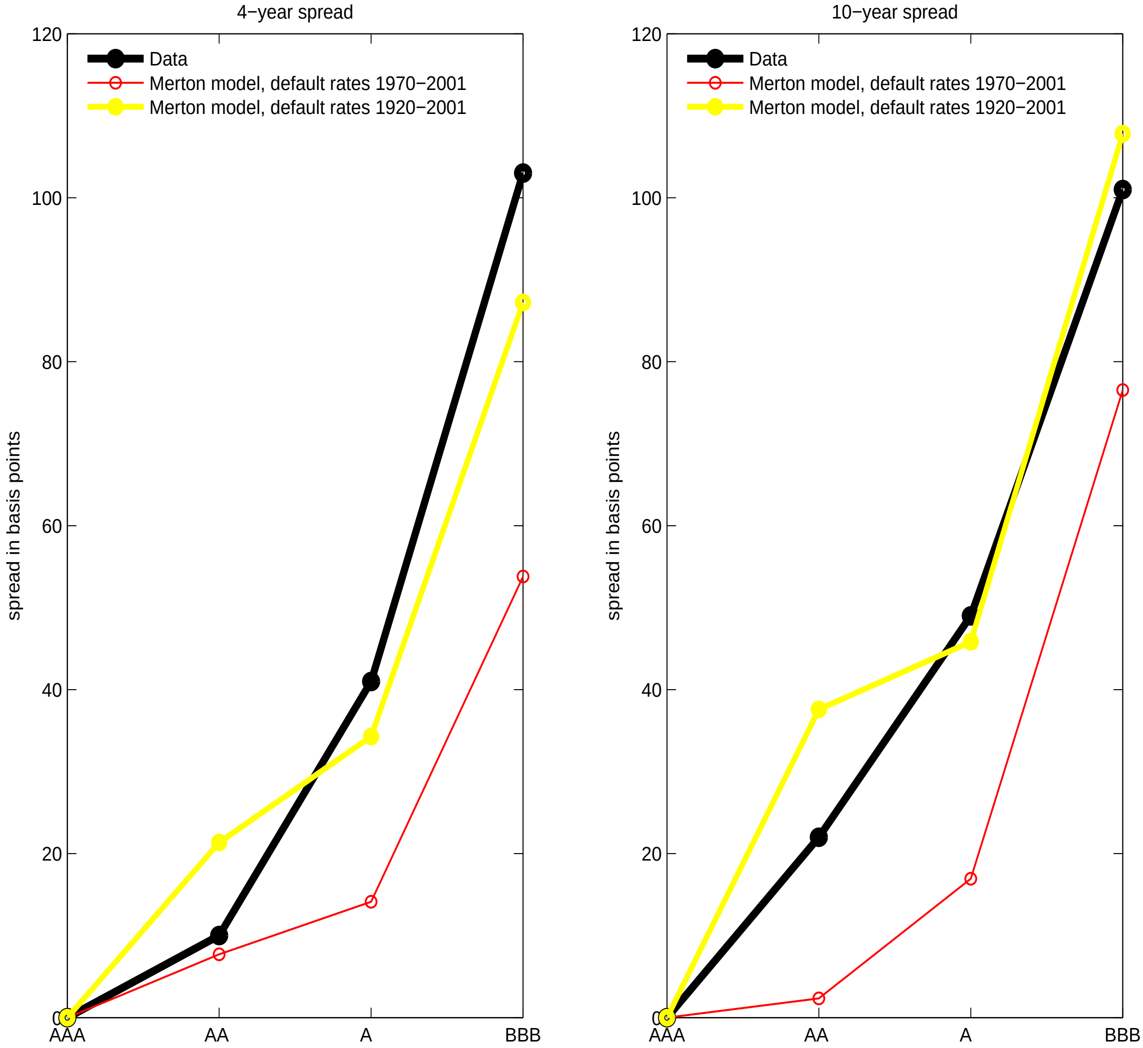


Fig. 2 *Corporate spreads to AAA-bond yields in the Merton model using default rates from different periods.* This figure shows actual and model-implied spreads to AAA yields. The thin red line shows spreads in the Merton model based on Moody's default rates from the period 1970-2001 and corresponds exactly to the calculations in Chen, Collin-Dufresne, and Goldstein(2009). The thick yellow line shows spreads in the Merton model where Moody's default rates from the period 1920-2001 are used. Actual spreads are from Duffee(1998).

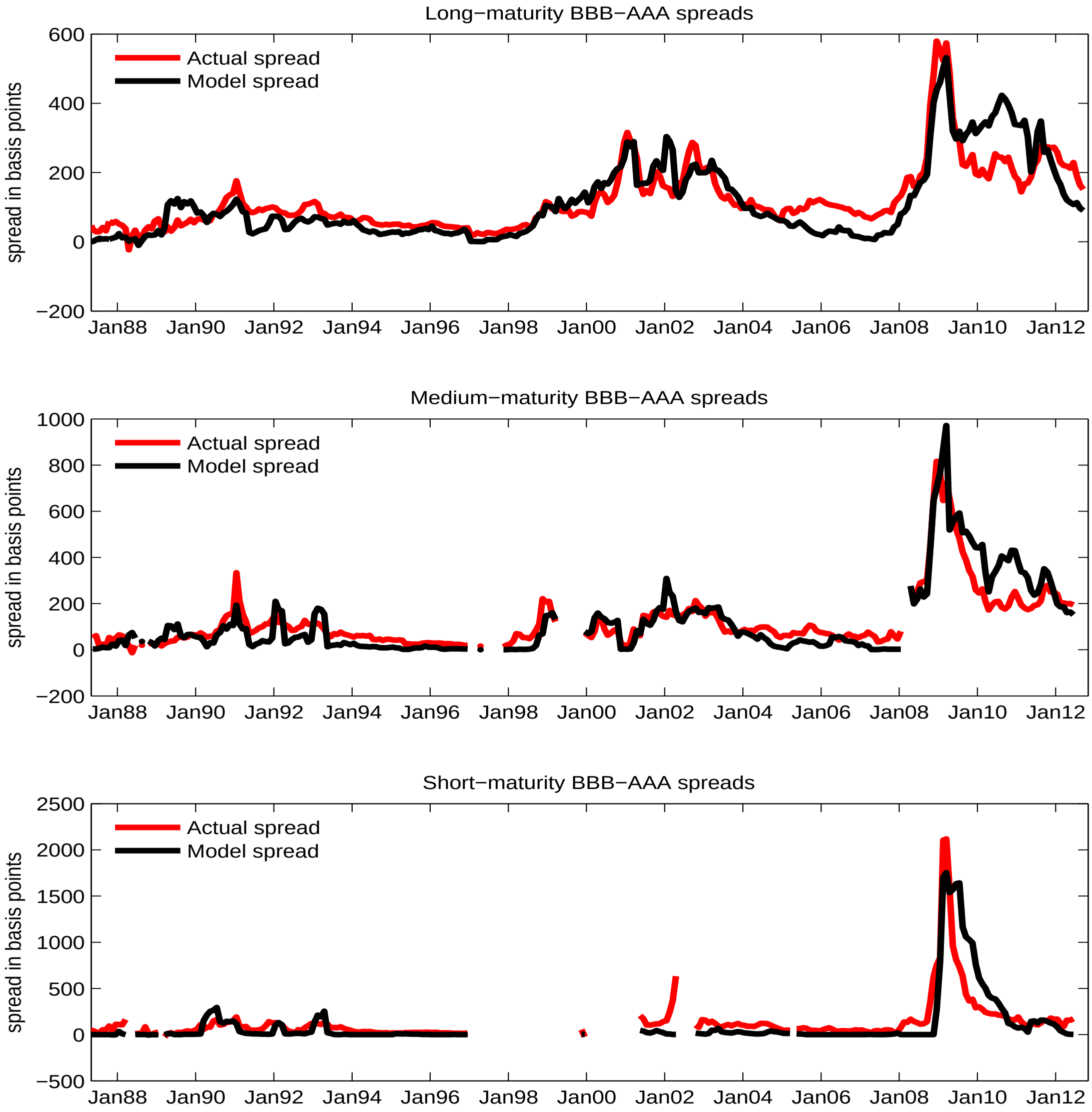


Fig. 3 *BBB-AAA corporate bond yield spreads.* This graph shows the time series of actual and model-implied BBB-AAA spreads. Each month all daily yield observations in bonds rated AAA/AA and bonds rated BBB are collected, and the graphs shows the average BBB spread to the swap rate minus the average AAA/AA spread to the swap rate. The results are shown for maturities between 4 and 30 years (long-maturity), 2 and 4 years (medium-maturity), and 0 to 2 years (short-maturity). 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.

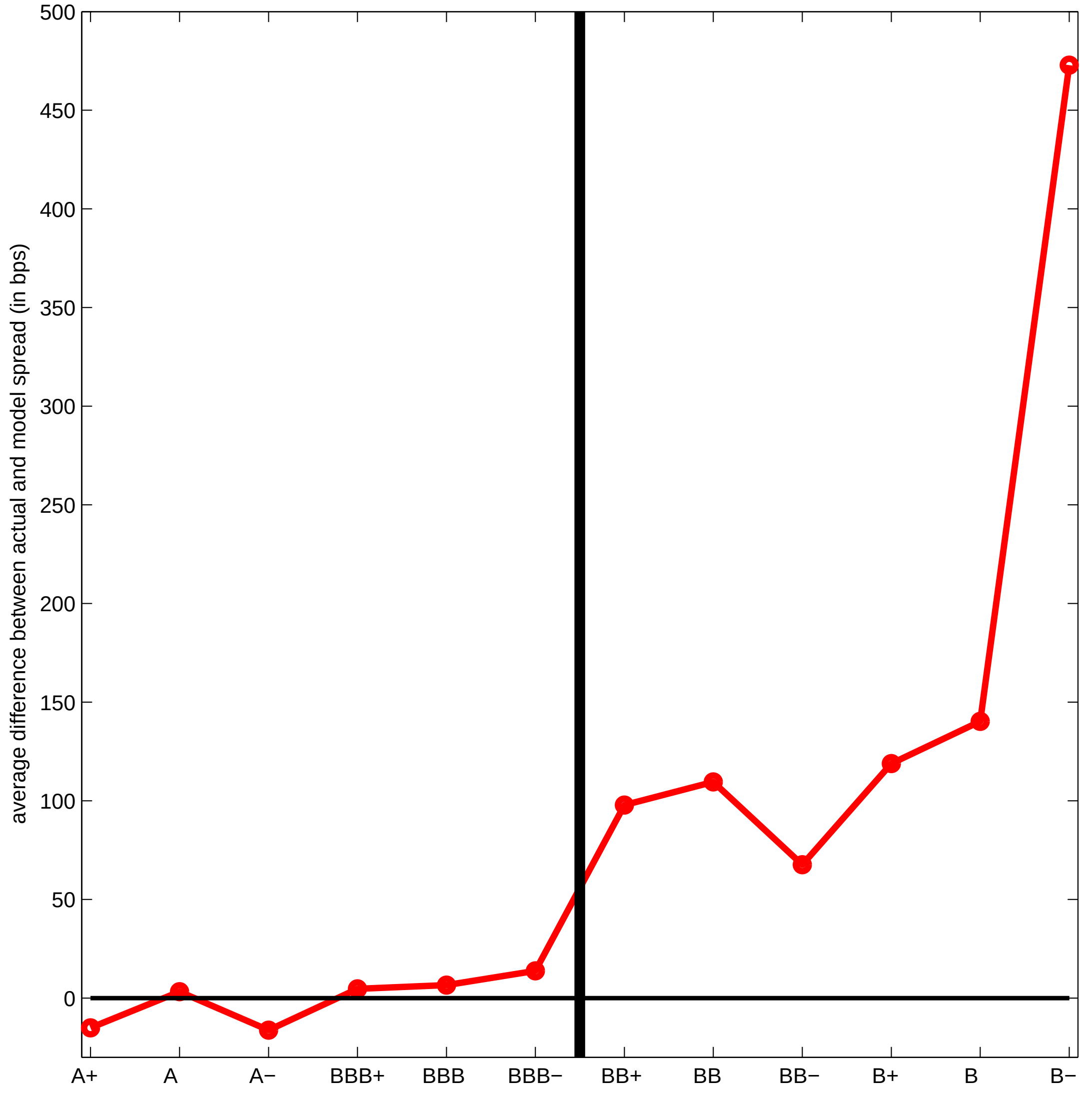


Fig. 4 *Average difference between actual and Merton-model credit spreads for rating notches.* On a monthly basis, the average actual yield spread to the swap rate for a given rating notch is calculated over all observations in that month with a maturity between 0 and 30 years and with a current rating corresponding to this rating notch. The overall average spread is calculated by calculating the average of monthly spreads. This is done for Merton model-implied spreads as well and the graph shows the difference (actual minus model) for each rating notch. Ratings to the left of the vertical line are investment grade ratings while those to the right are speculative grade.

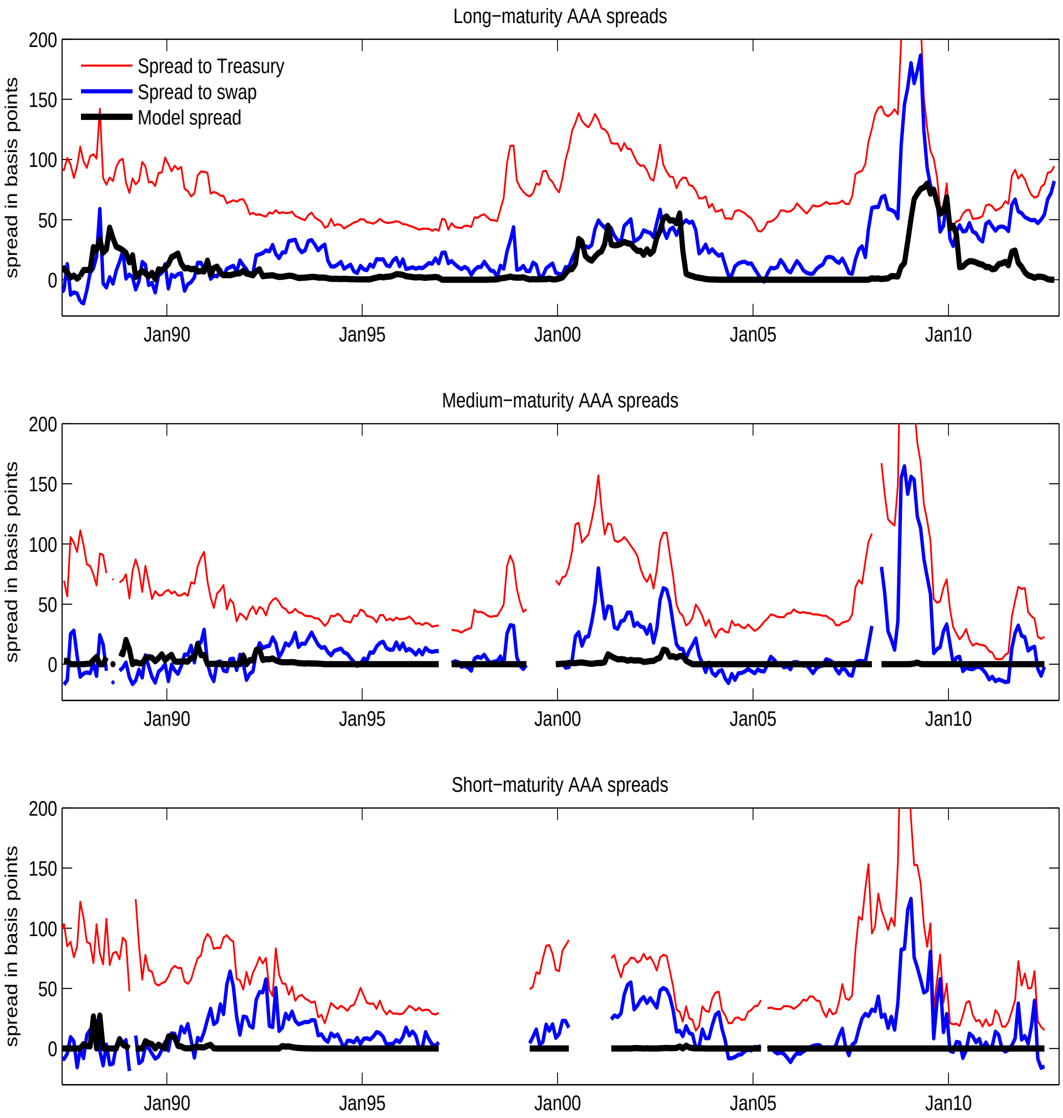


Fig. 5 *AAA corporate bond yield spreads.* This graph shows the time series of actual and model-implied AAA spreads. Each month all daily yield observations in bonds rated AAA or AA are collected (we call this the AAA yield) and the graphs shows the average AAA spread to (i) the swap rate and (ii) the Treasury yield. The results are shown for maturities between 4 and 30 years (long-maturity), 2 and 4 years (medium-maturity), and 0 to 2 years (short-maturity). The figure also shows the model-implied Merton spread.