The Myth of the Credit Spread Puzzle *

Peter Feldhütter[†]
London Business School

Stephen Schaefer[‡]
London Business School

June 30, 2016

Abstract

We ask whether standard structural models of credit risk are able to explain the average level and time series variation of credit spreads on corporate debt. Using a much longer time series of data compared to previous studies we find that the models match investment grade spreads well. A crucial ingredient to the success of the models is that we use default rates for a long period of 92 years to calibrate the models. 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. Because of high skewness in the distribution of realized default rates, using default rates from shorter periods as in most existing studies will often lead to the conclusion that model spreads are too low even if the model is the true model. When we use default rates from 1920-2012, we show that using a standard structural model we are able to match average investment grade spreads both in recent data and in the long run. Using the same approach we find that we are able to capture the time series variation of investment grade spreads over the period 1987-2012 with a correlation of 94%. Under the calibration method described in the paper, our results likely hold for a wide range of structural models.

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

JEL: C23; 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), Robert Merton, Jens-Dick Nielsen, Lasse Heje Pedersen, Scott Richardson, and seminar participants at the Q-Group Spring 2016 meeting, 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`

I declare that I have nothing to disclose for the purpose of reviewing or publishing this article in the Journal of Finance.

Peter Feldhutter
London Business School
Sussex Place
London, NW1 4SA
United Kingdom
Email: pfeldhutter@london.edu
Phone: +44-20-7000-8277

I declare that I have nothing to disclose for the purpose of reviewing or publishing this article in the Journal of Finance.

Stephen Schaefer
London Business School
Sussex Place
London, NW1 4SA
United Kingdom
Email: sschaefer@london.edu
Phone: +44-20-7000-8201

1. Introduction

The structural approach to credit risk, pioneered by Merton (1974) 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 of structural models find that "for investment-grade bonds of all maturities, credit risk accounts for only a small fraction of the observed corporate-Treasury yield spreads (what some researchers refer to as the "credit spread puzzle")" (Huang and Huang (2012)).[1] This has led researchers to try and resolve the puzzle by allowing for strongly time-varying Sharpe ratios, disaster risk, jump risk premiums, stochastic volatility premiums, time-varying recovery rates, and a counter-cyclical default boundary.[2]

In this paper we revisit the question whether standard structural models of credit risk, once calibrated to match historical default and recovery rates and the equity premium, are able to explain the average level of investment grade spreads without resorting to time-variation in 1) the price of risk for corporate assets, 2) the recovery rate, 3) the default boundary, or 4) other sources of priced risk.[3] The credit spread puzzle refers to the fact that the previous literature has answered this question in the negative. Using a much longer history of data than previous studies, we use the Merton (1974) and Black and Cox (1976) models as lenses through which to study the credit spread puzzle and find that the models are indeed able to explain the average level in investment grade credit spreads. In other words we find that there is no credit spread puzzle. Furthermore, we show that the Black-

[1]Examples of 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).

[2]Among others Chen (2010) and Bhamra, Kuehn, and Strebulaev (2010) consider a time-varying risk premium, Chen, Collin-Dufresne, and Goldstein (2009) consider a time-varying risk premium and a counter-cyclical default boundary, Cremers, Driessen, and Maenhout (2008) consider jump-risk premiums, McQuade (2013) consider a stochastic volatility premium, and Gabaix (2012) and Gourio (2013) consider disaster risk.

[3]Examples of models we consider standard include Merton (1974) (default at bond maturity), Black and Cox (1976) (default before bond maturity), Longstaff and Schwartz (1995) (stochastic interest rates), Leland (1994) (endogenous default, infinite bond maturity), Leland and Toft (1996) (endogenous default, finite bond maturity), Anderson and Sundaresan (1996) (strategic default), and models with idiosyncratic jumps in asset value.

Cox model can explain the time series variation of aggregate investment grade spreads with high precision. These results diverge from much of the literature and we explain carefully why we reach such different conclusions.

We start by looking at average spreads for investment grade debt implied by the Merton model 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 using average default rates starting from 1970 investment grade spreads (relative to AAA yields) in the Merton model are indeed too low compared to average actual spreads for 1985-1995.

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. The reason for the presence of skewness is that defaults are correlated across firms 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 defaults are perfectly correlated. In this case 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 the 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 over 1985-1995.

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 lower expected default rates for a given rating.[4] Our results suggest that this explanation should be discounted for two main reasons. First, the average BBB-AAA spread over 1920-2012 of 119bps is strikingly close to the average of 112bps over 1970-2012, suggesting strongly that average *ex ante* default probabilities in the two periods were actually very similar. Second, using default rates from 1920-2012 and spreads from the same period, we test the Merton model without making any assumption regarding the stability of ratings. Here, we find that the average actual BBB-AAA spread from 1920-2012 of 119 basis points is remarkably close to the value of 126 basis points we obtain in the Merton model when using the default rate from 1920-2012 as input.

Huang and Huang (2012) show empirically that many structural models which appear quite different in fact generate very similar spreads once the models are calibrated to the same default rates, recovery rates, and equity premium. In practice the default rates used for calibration (a) are averages over some period and (b) apply to classes of bonds (e.g., particular ratings) rather than individual bonds. This means that the similarity between the spread predictions from different models will apply to spreads averaged over similar periods of time and over bonds in the same class used for calibration. Indeed, when we study spreads that are averaged over ratings and maturities we find – consistent with Huang and Huang (2012) – that the results of the Merton and Black-Cox models are very similar and we prefer to use the Merton model because of the elegant simplicity of the model. However, when we look at spread data disaggregated either by time or *within* the class used for calibration – e.g., individual bonds – some models may provide a more accurate prediction of spreads than others and differences in model performance may emerge. To this point, we find that the Black-Cox model provides a somewhat better match to the term structure of historical default rates than the Merton model. When, later in the paper, we study spreads over time or at different points in the term structure we therefore use the Black-Cox model.

[4]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".

We implement the Black-Cox model using spreads for industrial firms over the period 1987-2012. Our dataset contains 205,624 individual bond yield observations. Our implementation approach is new to the literature in that it allows for cross-sectional and time-series variation in firm leverage and payout rate while matching historical default rates. Specifically, for each firm in the sample, we estimate the leverage ratio (debt over firm value) and payout rate (annual interest payments and dividends over firm value) period-by-period and compute a constant asset volatility by unlevering equity volatility period-by-period and calculating the average unlevered volatility. We then calibrate 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. In calibrating the default boundary we use a constant Sharpe ratio and match the equity premium, but once we have implied out the default boundary we compute firm and time-specific spreads using risk-neutral pricing, i.e., from the standard pricing formulae. 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.[5] Departing from the first approach enables us to study time series variation in spreads, while matching default rates ensures consistency with historical experience.

In this extensive data set, we first confirm that the Black-Cox model can capture average investment grade spreads; when we group all investment grade bonds with a maturity between 3-30 years, the average model spread is 105bps while the actual yield spread (relative to the swap rate) is 97bps. For speculative grade bonds, average spreads in the 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), although the underprediction appears to be larger than can be easily explained in this way. We also sort bonds according to the actual spread and find that actual and model-implied spreads are similar, except for bonds with a spread of more than 1000bps. For example, the average ac-

[5]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.

tual spread for bonds with an actual spread between 100-150bps is 121bps while the average model-implied spread is 124bps.

To study the time series we calculate average spreads on a monthly basis and we find that for bonds where most of trading occurs, namely for A and BBB rated bonds, there is a high degree of co-movement between actual and model-implied spreads. To illustrate the co-movement, Figure 1 shows the actual and model-implied monthly average investment grade spreads to the swap rate. Note that the model-implied spreads are "out-of-sample" predictions in that we do not use actual spreads as part of the calibration. Furthermore, for a given firm only changes in leverage and the payout rate – calculated using accounting data and equity values – lead to changes in the firm's credit spread. When regressing the monthly investment grade spread in the figure on the monthly model-implied spread, the slope coefficient is 0.93 with an R^2 of 88%. The corresponding regression for speculative grade spreads gives a slope regression of 0.86 but an R^2 of only 13%, showing that the model has a much harder time matching spreads for low-quality firms.

Although Figure 1 shows that *average* investment grade spreads are captured well on a monthly basis, the model does less well at the individual bond level. When regressing individual investment grade spreads on spreads implied by the Black-Cox model the R^2 is only 44%, so less than half the variation in investment grade spreads is explained by the model. For speculative grade spreads the corresponding R^2 is only 13%. It may be that although a wide range of structural models predict similar average spreads, their precision in capturing spreads on individual bonds varies.

There is an extensive literature on testing structural models. 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 different structural models and find that the models 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 variation 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 models are not calibrated to historical

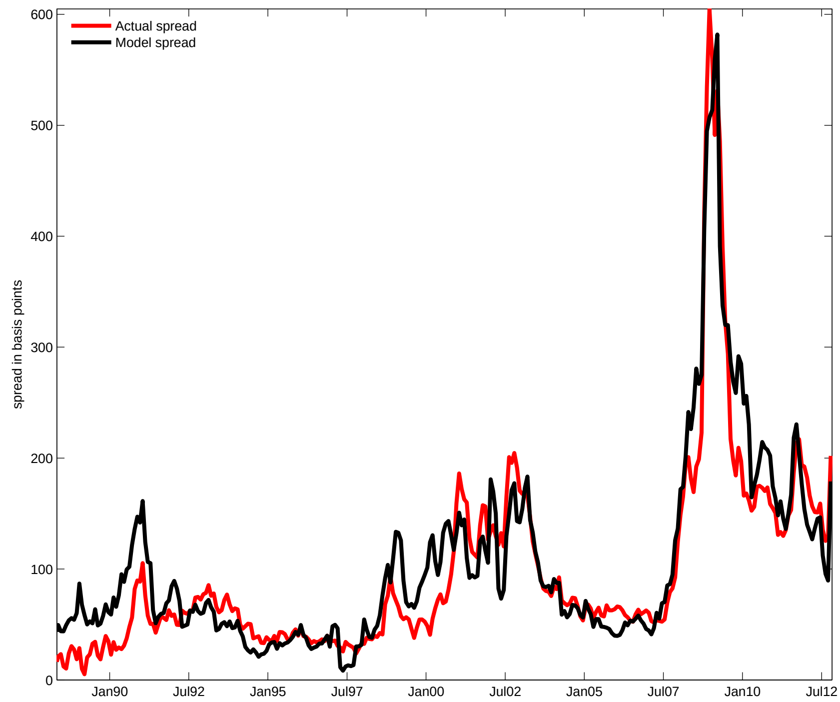

Fig. 1 *Time series variation in investment grade spreads.* This graph shows the time series of actual and model-implied investment grade corporate bond spreads. Each month all daily yield observations in bonds with an investment grade rating and with a maturity between 3-30 years are collected and the average actual spread (to the swap rate) and the average model-implied spread in the Black-Cox model are computed. The graph shows the time series of monthly spreads.

default rates, the results are 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 recognized the necessity of using a long history of default rates to estimate default probabilities with a reasonable degree of precision and, at the same time, has significantly over-interpreted unreliable estimates of default rates based on a short (30-year) history. Our use of long-run default rates is the central reason that our results differ from those in the literature. An exception to most of the literature on this point 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 and a constant default boundary, there is nonetheless a large amount of variation in realized average default rates even when these are measured over *several decades*.

The organization of the paper 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 Black-Cox model to capture both the average and the time series variation of investment grade spreads using individual bond data for the period 1987-2012. Section 4 concludes.

2 Structural models 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. In this section, 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.[6] If the asset value is below the face value of debt when the bond matures the firm cannot repay its bondholders and the firm defaults. 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 for the credit spread, s, as

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

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.

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

where $\overline{s}$ 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

[6] 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).

allows us to compute the average 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.[7]

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" with the result that 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 Fons, Cantor, and Mahoney (2002) and 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 spreads and default rates vary significantly across rating. Second, Moody's provides default rates as a function of rating as far back as 1920.[8]

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.[9] Even abstracting

[7]Note that although the relation in equation (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. This bias affects both spreads and default probabilities and in Appendix C we examine the effect of the bias on model-implied default probabilities derived in much of the literature.

[8]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).

[9]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).

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.[10] To explain how Moody's calculates 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 defines 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 divided by the number in the 1970 cohort.[11] 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.[12]

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[13], where

[10]All articles mentioned in Footnote 9 use Moody's estimates of default frequencies.

[11]Some firms have their rating withdrawn and Moody's has incomplete knowledge of subsequent defaults once firms are no longer rated. Moody's adjusts 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.

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

[13]We choose 1,000 firms each month because the average number of firms in Moody's BBB cohorts during

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

$$p = P(W_{iT}^P - W_{i0}^P \leq -\sqrt{T} \times DD) \tag{5}$$

where $DD = \frac{\log(V_0/F)+(\mu-\delta-\frac{1}{2}\sigma^2)T)}{\sigma\sqrt{T}}$ is the Distance-to-Default. This implies that the unconditional default probability is $N(-DD)$ where N is the cumulative normal distribution. For a given default probability p we can always find DD such that equation (5) holds, so in the following we use p instead of the underlying Merton parameters that give rise to p.[14]

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

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.

[14] Although only p is relevant for the simulation, we note that if we use median parameter values for BBB firms, $\sigma_A = 0.24$, $\delta = 0.044$, leverage=0.36, and $r = 0.05$ and a Sharpe Ratio of 0.22, the resulting default probability is 4.08%, close to $p = 4.39\%$ used in the simulation.

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 2 shows the distribution of the realized default rate in the simulation study.[15] 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 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 2 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.[16]

Moody's started to record default rates in 1920 and so we can increase statistical precision by extracting an estimate of the ex ante default probability from the average default

[15] The distribution of the default frequency in Figure 2 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 2.

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

frequency for the period 1920-2012 instead of 1970-1998. In the bottom graph in Figure 2 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 both 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.[17] 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 unreliable.[18] We use these insights when investigating the credit spread puzzle next.[19]

[17]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 $[3.10\%; 12.99\%]$.

[18]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 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 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.

[19]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.

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- 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 3 extends their results to all investment grade ratings and here 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 around 30 years of default data as CDG do, the realized default rate most often observed is below the average default probability and in this event 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 issue and Figure 3 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.[20]

The spread data used in Figure 3 are average spreads 1985-1995 from Duffee (1998)[21]. One concern may be that Moody's has 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 over time, but this is not what we observe in Moody's long-term BBB-AAA spread.[22] The average BBB-AAA spread for the period 1920-1995 is 123bps and for 1970-1995 it is 117bps. For the periods 1920-2012

[20]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.

[21]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).

[22]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 reports the average yield on both AAA and BBB bonds with a maturity of at least 20 years.

and 1970-2012 – which include the financial crisis – the corresponding figures are 119bps and 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 has 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 reports 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.[23] Using 20-year default rates from 1920-2012 to compute the 20-year BBB-AAA in the Merton model gives a spread of 125bps.[24] 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.*

2.4 Implication for structural models in general

So far, we have focused our attention on the Merton model primarily because it allows us to see the effect on the results of Chen, Collin-Dufresne, and Goldstein (2009) of using a long history of default rates. However, the model is only a convenient tool and our results have important implications for structural models in general. Huang and Huang (2012) (HH) show that many structural models which appear very different in fact generate similar

[23] 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.

[24] The 20-year default rate for 1920-2012 is 1.712% for AAA and 13.761% for BBB (see Moody's (2013)). 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. We use the 20-year default rate because this is the longest maturity for which Moody's report default rates. 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. 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).

spreads once the models are calibrated to the same 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.

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 same 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, as we have found, 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 calibrating models to relatively short default rate histories, e.g., from 1970 onwards, we show that a long history of default rates – much longer than the 30 years or so used in much of the recent literature – is necessary to estimate default probabilities accurately. With this insight in mind we show that:

1. Using *default rates starting from 1920* (instead of from 1970) to calibrate the Merton model, the model largely matches average investment grade spreads to AAA for 4-year and 10-year bond maturities.

Although the success of the model in explaining average spreads from recent decades by using long-run average default rates could be the result of a tightening of rating standards, we find this unlikely given that the average long-term BBB-AAA spread is very similar before and after 1970. We also provide further evidence that is immune to this criticism:

2. Moody's actual long-term BBB-AAA spread measured over 1920-2012 of 119bps is consistent with the model spread of 125bps calculated using the Merton model and average default rates from the same period.

Based on the findings of Huang and Huang (2012), these conclusions very likely apply to a wide range of structural models once these are calibrated to the same long-run historical default rates.

3. Structural models and time-series variation of spreads

In the previous section we showed that the Merton model is capable of capturing average investment grade spreads when calibrated to a long history of default rates. In this section we put structural models to a more demanding test and ask if they are also able to capture spreads (i) across a finer grid of bond maturities, (ii) when sorted (not according to rating but) according to the size of the actual spread and (iii) in particular 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 propose a new calibration method that fixes their average default probability over 1987-2012 to the average default rate over 1920-2012.

As discussed in the introduction, we find that the Black-Cox model provides a better fit to the term structure of default probabilities and we therefore use the Black-Cox model in analysing spreads disaggregated by time or maturity.

3.1 The Black-Cox model

As in the Merton model, asset value in the Black-Cox model follows a Geometric Brownian Motion under the natural measure,

$$\frac{dV_t}{V_t} = (\mu - \delta)dt + \sigma dW_t^P \tag{8}$$

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. The firm defaults the first time the asset value is below some fraction d of the face value of debt. One interpretation of the default boundary is that the bond has covenants in place that allow bondholders to take over the firm if firm value falls below the threshold. The cumulative default probability in the Black-Cox model at time T is (see Bao (2009))

$$\begin{aligned}\pi^P(T) &= N\Big[-\Big(\frac{(-\log(dL) + \big(\mu - \delta - \frac{\sigma^2}{2}\big)T}{\sigma\sqrt{T}}\Big)\Big] \\ &\quad + \exp\Big(\frac{2\log(dL)(\mu - \delta - \frac{\sigma^2}{2})}{\sigma^2}\Big) N\Big[\frac{\log(dL) + \big(\mu - \delta - \frac{\sigma^2}{2}\big)T}{\sigma\sqrt{T}}\Big]\end{aligned} \tag{9}$$

where $L = \frac{F}{V_0}$ is the leverage. Following Eom, Helwege, and Huang (2004), Bao (2009), Huang and Huang (2012), and others we assume that if default occurs, investors receive a fraction of the originally promised face value, but now with certainty. The credit spread, s, is given as

$$s = y - r = -\frac{1}{T}\log[1 - (1 - R)\pi^Q(T)] \tag{10}$$

where R is the recovery rate, T is the bond maturity, and π^Q is the risk-neutral default probability obtained by replacing μ with r in equation (9).

3.2 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.[25] 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). 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. Finally, we exclude bonds with a maturity less than three years because we find that dealer quotes are unreliable for short-maturity bonds.[26] In total we have 205,624 observations.[27]

Table 2 shows statistics for the corporate bond sample. The table shows that the number of bonds with a low rating of B or C is small; for example there are only 1-2 C-rated bonds in each of the maturity groups 13-17y, 17-23y, and 23-30y. One reason for this is that speculative grade bonds frequently contain call options which leads to their exclusion from our sample. In the following we report results for ratings B and C, but with the caveat that these results – particularly for long maturities – are based on few observations and therefore noisy.

To price a bond in the Black-Cox 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

[25]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.

[26]Dealer quotes are bid quotes and therefore yield quotes are upward-biased. When we compare quotes with actual transactions from TRACE, we find that for longer maturities this bias is small but for short maturities the bias is substantial. Results are available on request.

[27]We do not use TRACE transactions data because we wish to use standard data sources used in earlier literature, such that our results are most easily comparable with previous research. In a prior draft of the paper we used TRACE data and found very similar results. This draft is available on request.

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:

$$\sigma^2 = (1 - L_t)^2 \sigma_{E,t}^2 + L_t^2 \sigma_{D,t}^2 + 2L_t(1 - L_t)\sigma_{ED,t}, \tag{11}$$

where σ is the volatility of assets, $\sigma_{D,t}$ volatility of debt, $\sigma_{ED,t}$ the covariance between the returns on debt and equity, and L_t is leverage ratio. If we assume that debt volatility is zero, asset volatility reduces to $\sigma = (1 - L_t)\sigma_{E,t}$. 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 (11) 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, $(1 - L_t)\sigma_{E,t}$, and multiply this lower bound with SS's estimate of the ratio of asset volatility computed from equation (11) to the lower bound. Specifically, we estimate $(1 - L_t)\sigma_{E,t}$ and multiply this by 1 if $L_t < 0.25$, 1.05 if $0.25 < L_t \leq 0.35$, 1.10 if $0.35 < L_t \leq 0.45$, 1.20 if $0.45 < L_t \leq 0.55$, 1.40 if $0.55 < L_t \leq 0.75$, and 1.80 if $L_t > 0.75$.[28] This method has the advantage of being transparent and easy to replicate. For a given firm we then compute the average asset volatility over the sample period and use this constant asset volatility for every day in the sample period. All firm variables are obtained from CRSP and Compustat and details are given in Appendix B.

Summary statistics for the firms in our sample are shown in Table 3. The average leverage ratios of 0.14 for AAA and AA, 0.28 for A, and 0.38 for BBB are similar to those found in other papers.[29] 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

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

[29]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.

average equity volatility for BBB firms of 0.38 is higher than the value of 0.33 given in SS. Asset volatilities are slightly increasing in rating and broadly consistent with the estimates in SS.

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 the swap rate for the same maturity as the bond. 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/safety premium that pushes their yields below riskfree rates (Feldhütter and Lando (2008), Krishnamurthy and Vissing-Jorgensen (2012), and Nagel (2014)).

3.3 Estimation of the default boundary

Recall from Section 3.1 that the firm defaults if the asset value falls below a default boundary given as a fraction d of the face value of debt F. 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.[30] Huang and Huang (2012) show that holding default probabilities fixed at historical averages eliminates the dependence of the spread on the default boundary as seen in equation (2). 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 on bond i with a time-to-maturity T issued by firm j on date t, we calculate the firm's T-year default probability $\pi^P(d, \sigma_j, L_{jt}, \delta_{jt}, \theta, r_t^T, T)$ using equation (9). Here, σ_j is the firm's constant asset volatility, L_{jt} and δ_{jt} are the time-t estimates of the firm's leverage ratio and payout rate, θ is the Sharpe ratio that is constant across firms and over time, and r_t^T is the T-year riskfree rate. For a given rating a and

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

maturity T - rounded up to the nearest integer year - we find all bond observations in the sample with the corresponding rating and maturity and calculate the average default probability $\overline{\pi}^P_{aT}(d,\theta)$. 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 major ratings (AAA, AA, A, BBB, BB, B, C) and horizons of 1-20 years (Moody's only reports default rates for up to a horizon of 20 years) we find the value of d that minimizes the sum of absolute errors between the historical and model-implied default rates by solving

$$\min_{\{d\}} \sum_{a=AAA}^{C} \sum_{T=1}^{20} |\overline{\pi}^P_{aT}(d,\theta) - \hat{\pi}^P_{aT}|. \tag{12}$$

We use Chen, Collin-Dufresne, and Goldstein (2009)'s estimated Sharpe ratio of $\theta = 0.22$ for all ratings/observations when we calculate the model-implied default probabilities 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 0.87.[31,32] By implying out the default boundary from historical default rates while allowing for heterogeneity in firms and leverage 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)).

Note that we only use the Sharpe ratio of 0.22 and historical default rates to imply out

[31] The max, min, and median of the yearly default boundary estimates are 1.40, 0.59, and 0.83 respectively. If we restrict our simple to investment grade bonds, the estimated default boundary is 0.88.

[32] 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 is 0.86 of the face value of debt with a Sharpe Ratio of 0.21.

the default boundary. Once we have estimated the default boundary, we compute yield spreads by risk-neutral pricing.

3.4 The term structure of default probabilities with the estimated default boundary

We estimate the default boundary by matching the default probabilities of firms issuing straight coupon bullet bonds to historical default rates. There may be at least three concerns with our estimation approach. First, firms issuing straight coupon bullet bonds may be different from the average firm with the same rating. Second, although the average historical default rate across maturities is matched, the term structure of default rates might not be matched accurately. Third, there may be systematic differences in the ability of the model to match default rates across ratings.

To address these concerns we compare Moody's historical default rates with average default probabilities computed using all rated firms in Compustat. Specifically, we extract from Capital IG the issuer senior debt rating assigned by Standard & Poor's. There are almost no rating observations before 1985, so our sample period is 1985-2012. Table 4 gives estimates of the main parameters – as in Table 3 – for this new sample. Compared to the sample of firms used to estimated the default boundary (in Table 3) the sample is 4.6 times as large and has a reasonably large number of speculative grade firms.

To compute average default probabilities in the same way Moody's calculates historical default rates, we carry out the following computation for each rating a and horizon T. We form a cohort of a-rated firms in 1985 and calculate the cohort's average T-year default probability. Likewise, we form cohorts of firms rated a in 1986,...,2012 and compute an average T-year default probability for each cohort. Finally, we calculate the average T-year default probability across the 1985, 1986, ..., 2012 cohorts.[33]

Figure 4 and Table 5 show the average model-implied default probabilities and Moody's historical default rates for 1920-2012. We show in the figure and table a 95% confidence

[33] We use the 10-year Treasury CMT rate as the riskfree rate because we do not have swap rates in the first years.

band for the ex ante average default probability obtained using the simulation procedure described in Section 2.2.[34] Also, Table 5 shows '*' and '**' if the model-implied default probability is outside the 95% and 99% confidence band respectively.

We see that in the Black-Cox model there is a statistically significant underestimation of 1-2 year AA and A default probabilities and overestimation of speculative grade default probabilities up to a horizon of six years. AAA-default probabilities up to three years are – although small – significantly overestimated. The term structure of model default probabilities is particularly close to historical default rates for A and BBB rated firms, which account for more than half of the US corporate bond market volume (measured as the number of transactions).[35]

It may seem surprising that the Black-Cox model captures the term structure of default rates reasonably well because there are a number of papers showing that structural models imply essentially zero default probabilities for investment grade firms and for horizons below 3-4 years (Zhou (2001), Leland (2004), Leland (2006), Cremers, Driessen, and Maenhout (2008), Zhang, Zhou, and Zhu (2009) and others). We show in Appendix C that results in the existing literature, documenting a failure of structural models to capture short-term default rates, are strongly biased due to a "convexity effect" arising from Jensen's inequality. The bias arises when using a representative firm with average leverage (and average asset volatility and payout rate) to calculate short-term default probabilities because the default probability using average leverage is substantially lower than the average default probability calculated using the distribution of leverage. Our results show that once we circumvent the convexity bias by looking at individual firms and aggregating default probabilities, the Black-Cox model captures short-term default rates much better than previously claimed.

[34]For each rating we use the average number of firms in Moody's yearly cohorts over 1970-2012 as the number of firms in each year. The default probability in the simulation is Moody's average default rate 1920-2012 and the correlation coefficient for systematic risk is $\rho = 0.25$.

[35]According to TRACE Fact Book 2012, 53% of all U.S. corporate bond transaction volume in 2012 was in A or BBB rated bonds (Tables C24 and C25).

3.5 The default boundary and the cost of bankruptcy

A default boundary of 0.87 together with a recovery rate of 0.378 implies a deadweight cost of bankruptcy of 57%. Taking into account that a firm typically has different classes of debt outstanding (such as secured/unsecured bank loans/corporate bonds) and bank debt has higher recovery rates than corporate bonds, the estimated deadweight cost of bankruptcy is around 48%.[36] Andrade and Kaplan (1998), Davydenko, Strebulaev, and Zhao (2012), and Korteweg (2010) estimate bankruptcy ccosts to be 10-30% of firm value. Glover (2016) argues that existing estimates of costs of bankruptcy are biased because firms with low costs choose high leverage, and therefore the sample of observed defaults is a biased sample. Using a dynamic capital structure model, Glover (2016) finds average costs of bankruptcy to be 45%, close to the value implied by our estimated default boundary.

Given the disparity in estimates of bankruptcy costs, we follow most of the literature and calibrate the model to historical default and recovery rates by use of the default boundary while leaving bankruptcy costs as a "free" parameter. The idea is that default and recovery rates are less noisy than estimates of bankruptcy costs.

If bankruptcy costs are significantly smaller than our estimate, the model is misspecified. Our results are equally important in this case, because combining our results with those of Huang and Huang (2012) imply that it is likely that any reasonable structural model with low costs of bankruptcy will – when calibrated to match historical recovery and default rates and the equity premium – match credit spreads when assuming no priced risks beyond a constant asset risk premium.

3.6 Average corporate bond credit spreads

We calculate average spreads by following the calculations in Duffee (1998). Specifically, we calculate a monthly average actual spread for a given rating a, maturity range M_1 to M_2, and month t. To easy notation, we index the combination of rating, maturity range and

[36]Rauh and Sufi (2010) find that for rated firms the fraction of bank debt to bank and bond debt is 34.10% ($\frac{0.119}{0.119+0.230}$ in Table 1). The average recovery rate – weighted by number of observations – in Altman (2010)'s Table 2 is 38.43% for bonds and 57.02% for bank debt. Combining these two findings gives a recovery rate of 45% and the costs of bankruptcy is $1 - \frac{0.45}{0.87} = 0.48$.

month by $h = (a, [M_1; M_2], t)$. For a given h, we find all N_h bond observations with rating a and individual bond maturities $T_1^h, T_2^h, ..., T_{N_h}^h$, where $M_1 \leq T_i^h < M_2$, observed on days $\tau_1^h, \tau_2^h, ..., \tau_{N_h}^h$ in month t. Denoting the corresponding yield observations as $y_1^h, y_2^h, ..., y_{N_h}^h$ and the swap rates as $sw(\tau_1^h, T_1^h), sw(\tau_2^h, T_2^h), ..., sw(\tau_{N_h}^h, T_{N_h}^h)$, the average yield spread for rating a, maturity range M_1 to M_2, and month t is

$$s^h = \frac{1}{N^h} \sum_{i=1}^{N^h} \left(y_i^h - sw(\tau_i^h, T_i^h) \right). \tag{13}$$

The average yield spread for a given rating and a given maturity interval is then the average of the monthly values.

Similarly, we calculate model-implied spreads by replacing the actual spread $y_i^h - sw(\tau_i^h, T_i^h)$ with the spread $\hat{s}(d, \sigma_{j_i^h}, L_{j_i^h, \tau_i^h}, \delta_{j_i^h, \tau_i^h}, R, T_i^h, sw(\tau_i^h, T_i^h))$ implied by the Black-Cox model given in equation (10), where $\sigma_{j_i^h}$ is the asset volatility of firm j that issued bond i, $L_{j_i^h, \tau_i^h}$ and $\delta_{j_i^h, \tau_i^h}$ are the leverage ratio and payout rate respectively of firm j on day τ_i^h, $R = 37.8\%$ is the recovery rate, and the default boundary $d = 0.87$ is estimated when matching model-implied default probabilities to historical default rates in the first step of the calibration approach. Note that the Sharpe ratio does not enter the calculation, as it is only used when calibrating the default boundary d.

Table 6 shows actual and model-implied bond spreads in our sample, calculated using the approach just described. We see that the average actual investment grade spread across maturity is 97bps while the average model-implied spread is 105bps, showing that the model matches actual spreads well. We also see that there is a good correspondence between model-implied and actual investment grade spreads when we look at the individual maturities 3-7, 7-13, 13-17, 17-23, and 23-30 years. Thus, the Black-Cox model captures aggregate investment grade spreads well. Turning to speculative grade spreads we see an underprediction of spreads across maturity with average actual spreads at 546bps and average model spreads at 351bps. We note though that the number of bonds on which average speculative grade spreads are based is substantially smaller than for the investment grade (Table 2), so our conclusions for speculative grade spreads are less firm.

When we look at individual investment grade ratings in Table 6 the Black-Cox model seems to underpredict spreads on AA-rated bonds and overpredict spreads on A-rated bonds.

One reason for the underprediction of AA-rated bonds may be that the model underpredicts default rates for AA, something we examine further in Section 3.8. For A-rated bonds the average actual spread across maturity is 66bps while it is 82bps in the model, and the upward-sloping term structure of spreads from 50bps at 3-7 years to 103bps at 23-30 years is matched by the model (61bps to 114bps). The average actual BBB spread across maturity is 148bps while the average model-implied spread is 153bps. The actual term structure of BBB spreads is fairly flat and although the model underestimates long-maturity spreads (and the difference is statistically significant) the flat term structure is reasonably matched. Thus, the average spread of bonds where most trading takes place in the US corporate bonds – bonds with a rating of A or BBB – is captured well by the Black-Cox model. For speculative grade bonds, underprediction increases as we move down the rating scale. For BB-rated bonds, there is no significant underprediction for maturities below 13 years, while model spreads are too low for longer maturities. Since the number of long-term bonds is much smaller, our conclusions regarding long-maturity BB bonds are again less firm.

Overall, the correspondence between actual and model-implied investment grade spreads is good with an average investment grade spread across maturity of 105bps in the model and 97bps in the data. These results confirm our finding in Section 2 that standard structural models match average investment grade spreads. In contrast we find that the model underpredicts speculative grade spreads, with the average spread across maturity being 546bps in the data and 351bps in the model.

3.7 Sorting according to actual yield spread

The literature has traditionally compared model-implied and actual credit spreads within rating categories. There are several reasons for this. First, Moody's provides yield data and default rates from 1920 and there is therefore a long history of default and yield organised according to rating. In fact, so far we are aware, the only publicly available data on aggregate default rates are organised by rating. Second, spreads organised according to rating show a large variation in the mean, with lower rated firms having higher average spreads; matching average bond spreads organised by rating has provided a hard test for structural models.

Although average default rates are available only by rating, we can nevertheless sort bonds in other ways than by rating and compare model-implied and actual spreads. If there is a substantial difference, the model is misspecified in some dimension. Since any useful sort should result in significant variation in spreads, the most obvious choice is to sort according by actual spreads.

Table 7 shows model spreads sorted according to the size of the actual spread. We see that except for spreads above 1000bps the model matches actual average spreads well. For example, the actual spread averaged across maturity for bonds with spreads between 100-150bps is 121bps while it is 124bps in the model. For bond spreads above 150bps we start to see an under-prediction at long maturities, although it is modest and only becomes strong when spreads are above 300bps. For maturities below 17 years average model spreads and average actual spreads are relatively close for spreads up to 300-1000bps.

Overall, the results when sorting by actual spread are similar to those sorted by rating, namely that spreads for low credit risk firms are matched well while spreads for the highest credit risk firms, particular for bonds with long maturity, are under-predicted.

3.8 Relation between yield spread and historical default rate

We use historical default rates in calibrating the default boundary and so, averaged across all ratings and maturities, the model-implied average default probability matches the average historical default rate. For a particular maturity and rating, however, the average model-implied default probability may differ from the historical default rate, for example as a result of model misspecification. For example, the analysis in Section 3.4 showed that the Black-Cox model tends to overestimate short-term speculative grade default rates and underestimate short-term investment grade default rates. To address this problem we can compare actual spreads with model-implied spreads computed using equation (3) where we set the default probability equal to the historical default rate. This is the approach we have taken in Section 2 and has also been used by Huang and Huang (2012) and other papers in the literature.

In Figure 5 “data” plots average spreads from Table 6 against historical default rates from Table 5. The line marked “Merton model” plots credit spreads computed from equation (3) for the range of default probabilities shown on the horizontal axis. For investment grade bonds (the first four ’Data’ dots from the left in each graph) the model matches actual spreads for 5- and 10-year maturities and for 15- and 20-year maturities the model in fact implies higher spreads than actual spreads. This suggests that the modest underprediction for AAA and AA rated bonds found in the previous sections may be due to the model underpredicting default probabilities rather than an underprediction of the compensation for risk in spreads. Although the lowest rating categories have relatively few observations, the figure shows a pattern of underprediction for these ratings that is consistent with the evidence in the previous section.[37]

Figure 5 also shows the relation between default probability and spread in the Black-Cox model. We see that spreads in the Merton and Black-Cox models are very similar, which as observed earlier is to be expected when the models are calibrated to the same recovery rates, default rates, and the equity premium.

Overall, the results in this section show that the primary reason for the (slight) underprediction in the Black-Cox model of AAA and AA spreads is due to the model-implied default probabilities being lower than historical default rates. If historical default rates reflect ex ante default probabilities, the model is misspecified through either the default boundary or asset value dynamics.

3.9 Time series variation in yield spread

Having established that standard structural models can match the size of investment grade credit spreads, we next examine whether they can also captures the time series variation of credit spreads. In each month, we calculate the average yield spread for a given rating according to equation (13) in both the model and in the data (where the spread is relative to the swap rate), and investigate the time series of monthly spreads. Model spreads are calculated using the Black-Cox model as in Section 3.3 where the default boundary $d = 0.87$

[37] Culp, Nozawa, and Veronesi (2015) find in their Figure 2A that the Merton model substantially underpredicts spreads for a given default probability. Our results are not directly comparable with theirs because they look at bond maturities below two years while we look at bond maturities above three years. However, to a large extent the difference in results can be explained by recovery rates: Culp, Nozawa, and Veronesi (2015) do not match historical recovery rates in their figure and their low Merton spreads reflect counterfactually high recovery rates.

is estimated from matching model-implied default probabilities to historical default rates.

To provide an overall assessment of the model's ability to capture investment grade spreads, we group all investment spreads and all maturities between 3-30 years and plot the actual and model-implied spreads in Figure 1. Note that for a given firm, asset volatility is constant over the sample period and therefore it is only changes in the leverage ratio and the payout rate that causes changes in a given bond's spread (besides the effect of changes in the riskfree rate and bond maturity). We see that the model-implied spread tracks the actual spread very well.

Figure 6 shows the time series variation for 5- and 10-year spreads for the four ratings AA to BB where there are most monthly observations (see Table 6 for number of observations). The figure shows that the time series variation of A and BBB (and to a lesser extent BB) spreads is tracked well by the Black-Cox model. In contrast, the model spread for AA spread is too low and shows too little variation over time. The better fit of A and BBB is partly due to the fact that for these ratings the time series is based on a larger number of bonds and firms and therefore more noise is eliminated in the aggregation.

To test more formally the ability of the Black-Cox model to capture the time series variation in spreads, we regress the monthly time series of the actual spread, s_t, on the model-implied spread, $\hat{s}_t$,

$$s_t = \alpha + \beta \hat{s}_t + \epsilon_t \tag{14}$$

and report the β and the R^2 of the regression in Table 8. The table shows that – for bonds with maturities between 3-30 years – the regression of actual investment grade spreads on model-implied investment grade spreads gives a slope coefficient of 0.93 and an R^2 equal to 88% showing that once investment grade spreads are aggregated model-implied spreads track actual spreads with high precision. The R^2's for A and BBB-rated spreads are high and between 48-81% across maturities. The regression coefficients are lower than one, but in the case of BBB close to and often not statistical significantly different from one. For speculative grade and AAA/AA ratings the Black-Cox model's ability to capture the time series variation starts to deteriorate. More noise due to fewer observations is one factor contributing to the deteriorating fit.

3.10 The role of bond illiquidity

Our finding that structural models capture investment grade 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.[38] However, our results are broadly in line with these findings and to see this we focus on the contribution of illiquidity to yield spreads for long-maturity bonds in DFL's Table 4.

In DFL's Table 4 we see that in normal times (Panel A) the contribution of illiquidity to investment grade spreads is less than 5bps while it is 84bps for speculative grade spreads. This is consistent with our finding that average investment grade spreads are captured by structural models while speculative grade spreads are underpredicted. A comparison of the average underprediction of speculative grade spreads of 195bps in Table 6 with the illiquidity premium of 84bps reported in DFL suggests that the underprediction of speculative grade spreads is not explained by an illiquidity premium alone.

During the 2008-2009 crisis DFL find an illiquidity premium in AAA, AA, A, BBB, and speculative grade bonds to be 8bps, 65bps, 75bps, 98bps, and 243bps respectively. If we restrict the analysis in Section 3.6 to the crisis period in DFL (2007:Q2-2009:Q2) we find for maturities 3-30 years that the average difference between actual and model-implied spreads for AAA, AA, A, BBB, and speculative grade bonds is 18bps, 63bps, -50bps, -14, and 303bps respectively (when calculated as in Table 6). For AAA, AAA, and speculative grade bonds the model underprediction is comparable to the liquidity premium found in DFL. However, for A and BBB rated bonds the model overpredicts spreads in the crisis period, suggesting the presence of some misspecification. In the next section we explore potential sources of model misspecifiation.

3.11 Spread predictions on individual bonds

Our main result is that structural models with a constant 1) Sharpe ratio, 2) recovery rate, and 3) default boundary, and no priced risks beyond diffusion risk, and when calibrated

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

to match historical default rates can match the average spread of portfolios of investment grade bonds. This result does not necessarily imply that the model can match spreads on individual bonds with a high degree of precision, because average spreads may well mask significant individual pricing errors. While our interest mainly lies in testing if the model can capture average spreads, we nevertheless carry out an exploratory analysis on the ability of the Black-Cox model to capture the cross-section of spreads (for a more extensive analysis see Bao (2009)).

The first column in Table 9 shows the R^2's from regressing actual spreads on model-implied spreads (and a constant) at the individual bond level. For investment grade bonds the R^2 is 44%, substantially below the R^2 of 88% when we aggregate spreads at a monthly level in Table 8. For speculative grade bonds the explanatory power in the regression at the individual bond level is only 13%, showing that the Black-Cox model has only limited ability to price speculative grade bonds.

To give an indication on how the model or parameter estimates might be improved, we correlate the pricing error – the difference between actual and model-implied spread – with variables used in the estimation. Table 9 shows the results. The pricing errors for investment grade bonds have a correlation of -0.39 with leverage and -0.33 with the payout rate. This suggests that estimates for individual bonds could be improved either by estimating leverage and payout rate in a different way or indeed by changes to the model. If asset volatility is stochastic, we expect to see a negative correlation between pricing errors and both day-to-day estimates of asset volatility and equity volatility. These correlations are modest and range from -0.13 to 0.11, so stochastic volatility seems to be less important in improving cross sectional accuracy.

4. Conclusion

We test the Merton and Black-Cox models of credit risk using U.S. corporate bond spread data from 1920-2012, 1970-2012, 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, we find that the time series of the model-implied investment grade spread tracks the actual investment grade spread well with a correlation of 94%.

Crucial for our conclusions is that we calibrate the models 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 implied by structural models are too low even if the models are correct.

Our results show that the credit spread puzzle - the perceived failure of structural models to explain levels of credit spreads particularly for investment grade bonds - has less to do with deficiencies of the models than with the way in which they have been implemented. While we only test the Merton and Black-Cox models, 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 the historical recovery rates, default rates, and the equity premium 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 Rating through the cycle and the use of average default rates to calculate average spreads

We suggest in Section 2 that the spread calculated from formula (3) using the *average* default probability is close to the average spread. Formula (3) is:

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

where $\overline{y-r}$ is the average corporate bond spread, R the recovery rate, θ the Sharpe ratio and $\overline{\pi}^P$ the average default probability.

Chen, Collin-Dufresne, and Goldstein (2009) and Huang and Huang (2012) find that formula (3) provides a close approximation to the average spread and the example in Table A1 supports their conclusion. The example assumes that rating agencies "rate through the cycle" and that an investment grade firm in a recession period has the same default probability as a speculative grade firm in an expansion. We see that using the average default probability across the expansion and recession periods as input in the calculation of the model spread for the investment grade firm yields a spread of 95bps while the average model spread is 93bps. Thus, even though there is significant variation in the default probability of the investment grade firm, there is only a small bias in using the spread computed using the average default probability as an input to equation (3) as a proxy for the average spread. Likewise, we see for the speculative grade firm that the difference between the spread using the average default probability and the average spread is only 4bps out of a total spread of 201bps.

We provided a more extensive analysis through simulation in an earlier version of the paper and found the bias to be small in that analyis as well. The results are available on request.

	Investment grade		Speculative grade	
	Recession	Expansion	Recession	Expansion
Default probability	6%	2%	16%	6%
Spread (bps)	129	56	272	129
Average def.prob.	4%		11%	
Average spread (bps)	93		201	
Spread using av. def.	95		205	

Table A1 *Bias (in basis points) when calculating spread using average default rate and comparing with average spread when rating agencies "rate through the cycle".* This example illustrate how the fact that rating agencies "rate through the cycle" affects results when comparing the average spread with a spread calculated using the Merton model with an average default rate as input. The Merton formula 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 R is the recovery rate, π^P is the natural default probability, θ is the Sharpe ratio, and T is the bond maturity. We use a recovery rate of $R = 37.8\%$ which is the average recovery rate for 1982-2012 and a Sharpe ratio of $\theta = 0.22$ from Chen, Collin-Dufresne, and Goldstein (2009).

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.

C Convexity bias when using a "representative firm" to calculate default probabilities

Our finding in the Section 3.4 that the Black-Cox model matches default probabilities for BBB-rated firms even for horizons as short as one year is surprising, since it is an established stylized fact in the literature that short-run default probabilities in structural models with only diffusion-risk are much too low. Papers showing that default probabilities at short horizons are too low include among others Zhou (2001), Leland (2004), Leland (2006), Cremers, Driessen, and Maenhout (2008), Zhang, Zhou, and Zhu (2009), and McQuade (2013).

The reason that we arrive at a different conclusion is that we allow for cross-sectional variation in asset volatility and both cross-sectional and time series variation in leverage and payout rates. In contrast, the existing literature uses a "representative firm" with average asset volatility, leverage, and payout rate within a given rating category. Using a representative firm leads to a bias due to Jensen's inequality because the default probability is typically convex in asset volatility and leverage (while it is close to linear in the payout rate). This convexity bias in the case of leverage is illustrated in Figure A1. The convexity bias when using a representative firm to calculate spreads is known to the literature, but importantly the impact of the convexity bias on the short-run default probabilities has not been recognized in the literature.[39]

To document the impact of the convexity bias on short-run default probabilities we focus on heterogeneity in leverage and carry out a simulation of 100,000 firms. For each firm we use an asset volatility of 23%, a payout rate of 3.3%, and a Sharpe ratio of 0.22. The firms differ only in their leverage ratios and we draw 100,000 values from a normal distribution with mean 0.29 and a standard deviation of 0.18.[40] The chosen values are median values for BBB firms and the standard deviation of leverage in the simulation is equal to the empirical standard deviation of BBB firms in the sample. Finally, the riskfree rate is 5%. For each firm we calculate the cumulative default probability for different maturities. Panel A in Table

[39]Bhamra, Kuehn, and Strebulaev (2010) present a structural-equilibrium model with macro-economic risk and simulate default rates over 5 and 10 years and find there is a substantial effect of allowing for firm heterogeneity. They do not look at default probabilities below five years, our main focus.

[40]If a simulated leverage ratio is negative we set it to zero. This implies that the average leverage ratio is slightly higher than 0.29, namely 0.2937 in our simulation.

A2 shows the average default probability and the correct asset volatility of 23% that is used for all firms and at all maturities.

Zhou (2001), Leland (2004), Leland (2006), and McQuade (2013) use values of the leverage ratio, payout rate, and asset volatility averaged over time and firms to calculate model-implied default probabilities for a representative firm and then compare these with historical averages. To see the extent of the convexity bias in the Black-Cox model when using their approach, we calculate the term structure of default probabilities in Panel B of Table A2 for a representative firm with a leverage ratio equal to the mean in our simulation of 0.29. There is a downward bias in default probabilities relative to the correct values given in Panel A and the bias becomes more pronounced at shorter maturities. For example, the one-year default probability of the representative firm in Panel B is 0.00% while the true average default probability in Panel A is 0.42%. The aforementioned papers compare the default probability of the representative firm with the average historical default rate and since the historical default rate reflects the average default probabilities in Panel A, their results for particularly short-maturity default probabilities are strongly biased.

Cremers, Driessen, and Maenhout (2008), Zhang, Zhou, and Zhu (2009), and Huang and Huang (2012) let a representative firm match historical default rates by backing out asset volatility. To examine how the convexity bias influences the implied asset volatility we proceed as follows. For a given maturity, we compute the asset volatility that allows the representative firm to match the average default probability in the economy at that given maturity. Panel C shows the implied asset volatilities and we see that there are two problems with this approach. The first problem is that asset volatility is biased: all firms in the economy have an asset volatility of 23% and yet the implied asset volatility ranges from 27.0% at the 10-year horizon to 43.3% at the one-year horizon. The finding that implied asset volatility in the diffusion-type structural models is too high, particular at shorter horizons has been seen as a failure of the models, but this example shows that the high implied asset volatility arises mechanically from the use of a representative firm. The second problem is that it is not possible to match the term structure of default probabilities without counterfactually changing the asset volatility maturity-by-maturity.

Cremers, Driessen, and Maenhout (2008) and Zhang, Zhou, and Zhu (2009) use a representative firm to imply out asset volatility by matching long-term default rates and then use this asset volatility to calculate the term structure of default probabilities. We replicate this approach by implying out the asset volatility that makes the representative firm's default probability match the average default probability for the 10-year bond in the economy and then calculate the term structure of default probabilities for this representative firm. The implied asset volatility is 27.0% and the term structures are in Panel D. The difference between the implied asset volatility of 27.0% and the true value of 23% reflects a moderate convexity bias at the 10-year horizon, but since the bias becomes more severe at shorter horizons, the strong downward bias in default probabilities reappears as maturity decreases. Thus, the bias in short-term default probabilities persists when using a representative firm and imputing asset volatility by matching a long-term default rate[41].

In summary, we show that the term structure of default probabilities in the Black-Cox model is downward biased, and more so at short maturities, when using a representative firm. This is likely to be true for any standard structural model: default probabilities are strongly bias downwards at short maturities. Existing evidence (showing that default probabilities at short horizons are much too low) in Zhou (2001), Leland (2004), Leland (2006), Cremers, Driessen, and Maenhout (2008), Zhang, Zhou, and Zhu (2009), and McQuade (2013) is subject to this strong bias and therefore not reliable.

[41]Our results clarify those in Bhamra, Kuehn, and Strebulaev (2010). Within the framework of their structural-equilibrium model, they compare a representative firm with a cross section of firms and find that the slope of the term structure of default probabilities is flatter for the cross section of firms. In their experiment, the cross section of firms have an average default probability that is more than three times as large as the default probability of the representative firm (their Table 3, Panels B and C). Since the term structure of default probabilities becomes flatter for a representative firm at the same time as default risk increases, it is not clear if it is cross sectional variation or the rise in default probability that drives the flattening of the term structure. Since we hold the 10-year default probability fixed in Panels A and D, it is clear in our analysis that the flatter term structure is driven by cross-sectional variation in leverage alone.

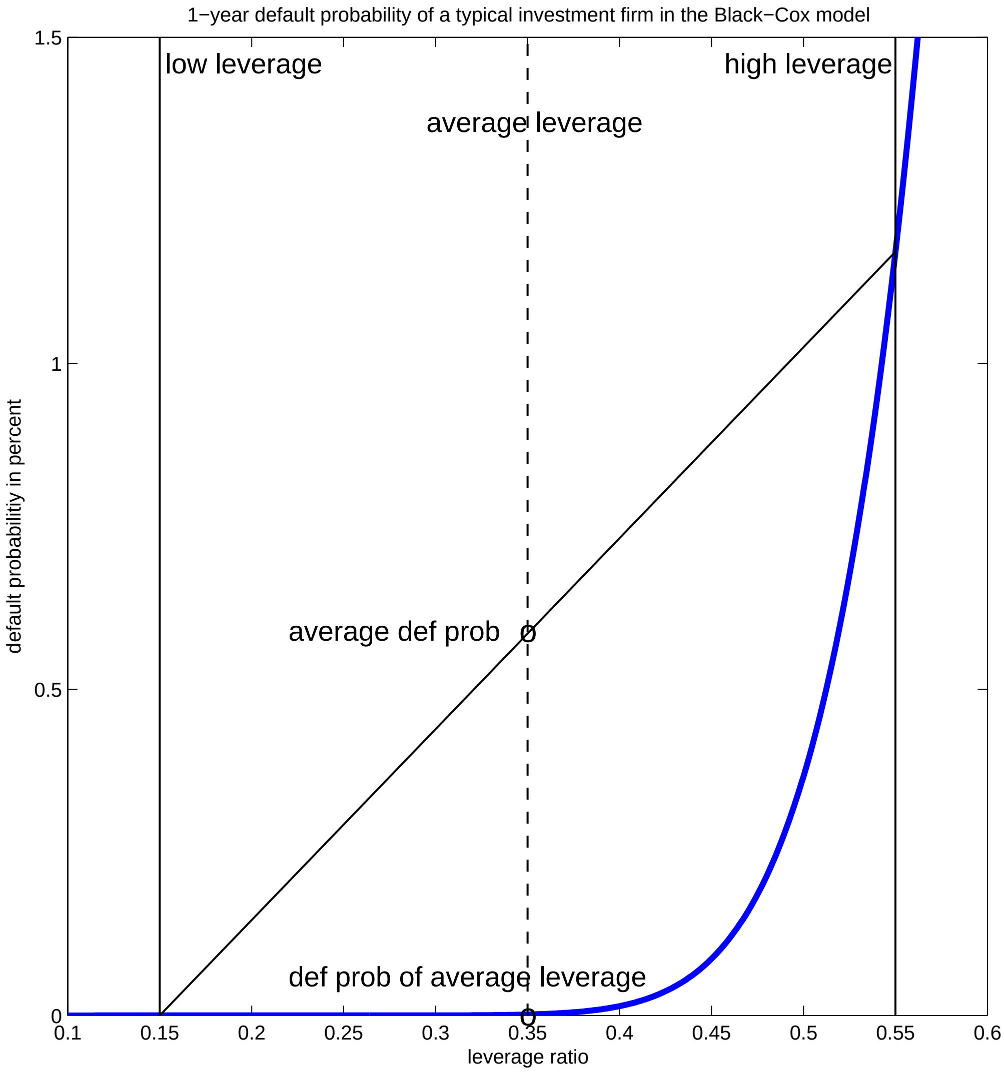

Fig. A1 *Convexity bias when calculating the default probability in the Black-Cox model using average leverage and comparing it to the average default probability.* It is common in the literature to compare historical default rates to model-implied default probabilities, where the latter are calculated using average firm variables. This introduces a bias because the default probability in structural models is a non-linear function of firm variables. The figure illustrates the bias in case of two firm observations with the same rating, one with a low leverage ratio and one with a high leverage ratio. The two observations can be two different firms at the same point in time or the same firm at two different points in time. Asset volatility is 5%, dividend yield 3.7%, Sharpe ratio 0.22, and riskfree rate 5%.

maturity	1	2	3	4	5	6	7	8	9	10
Panel A: True economy (there is variation in leverage ratios)										
average default probability	0.13	0.59	1.24	1.98	2.75	3.50	4.23	4.92	5.58	6.20
asset volatility	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
Panel B: Representative firm (average leverage ratio used)										
default probability	0.00	0.00	0.05	0.20	0.49	0.89	1.37	1.90	2.46	3.03
asset volatility	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
Panel C: Representative firm, average def. prob. at bond maturity is matched										
default probability	0.13	0.59	1.24	1.98	2.75	3.50	4.23	4.92	5.58	6.20
implied asset volatility	44.3	37.3	34.3	32.6	31.4	30.5	29.9	29.3	28.9	28.6
Panel D: Representative firm, average def. prob. at 10-year bond maturity is matched										
default probability	0.00	0.03	0.24	0.74	1.47	2.34	3.30	4.28	5.25	6.20
implied asset volatility	28.6	28.6	28.6	28.6	28.6	28.6	28.6	28.6	28.6	28.6

Table A2 *Convexity bias when calculating default probabilities in the Black-Cox model using the representative firm approach.* It is common in the literature to compare average actual default rates to model-implied default probabilities, where model-implied default probabilities are calculated using average firm variables. This introduces a bias because the default probability and spread in the Merton model is a non-linear function of firm variables. This table shows the magnitude of this bias. Panel A shows, for maturities between one and 10 years, the average default probability for 100,000 firms that have different leverage ratios but are otherwise identical. Their common asset volatility is 25% and payout rate 3.7%. Their leverage ratios are simulated from a normal distribution with mean 0.28 and standard deviation 0.18 (truncated at zero). The riskfree rate is 5%. Panel B shows the default probability of a representative firm where the average leverage ratio is used. In Panel C, for each maturity – one at a time – an asset volatility is computed such that, for a representative firm with a leverage ratio equal to the average leverage ratio, the default probability is equal to the average default probability in the economy (given in the first row in Panel A and again in Panel C). This is done separately for each maturity. The panel shows the resulting implied asset volatility. Panel D shows the results of a calculation similar to that in Panel C except here the asset volatility used to compute the default probability for each maturity is the value that matches the average *10-year* default probability in the economy.

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.

Altman, E. (2010). Default Recovery Rates and LGD in Credit Risk Modeling and Practice: An Updated Review of the Literature and Empirical Evidence. *Unpublished Manuscript*, 1–40.

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

Anderson, R. and S. Sundaresan (1996). Design and Valuation of Debt Contracts. *Review of Financial Studies 9*(1), 37–68.

Andrade, G. and S. Kaplan (1998). How costly is financial (not economic) distress? Evidence from highly leveraged transactions that become distressed. *Journal of Finance 53*, 1443–1493.

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.

Black, F. and J. Cox (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *Journal of Finance 31*, 351–367.

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.

Culp, C., Y. Nozawa, and P. Veronesi (2015). Option-Based Credit Spreads. *Working paper*.

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

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., 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.

Fons, J., R. Cantor, and C. Mahoney (2002). Understanding Moody's Corporate Bond Ratings And Rating Process. *Special Comment. New York: Moody's Investors Services*, 1–16.

Gabaix, X. (2012). Variable rare disasters: an exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics 127*, 645–700.

Glover, B. (2016). The Expected Cost of Default. *Journal of Financial Economics 119*, 284–299.

Gourio, F. (2013). Credit risk and disaster risk. *American Econ. Journal: Macroeconomics 5*(3), 1–34.

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

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.

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.

Korteweg, A. (2010). The Net Benefits to Leverage. *Journal of Finance 65*, 2137–2170.

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. (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.

Longstaff, F. and E. S. Schwartz (1995). A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *Journal of Finance 50*, 789–819.

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*.

Rauh, J. and A. Sufi (2010). Capital Structure and Debt Structure. *Review of Financial Studies 23*, 4242–4280.

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.

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

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 is 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.

3-7-year bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	21	109	327	289	116	36	14	753
Mean number of bonds pr quarter	7.77	8.55	24.1	19.9	9.22	2.61	1.27	63.2
Mean number of quotes pr quarter	16.87	117.1	257.7	270.1	170.5	42.94	30.66	826
Age	2.05	4.73	5.19	5.89	4.37	3.07	2.50	5.08
Coupon	7.19	6.30	7.00	7.57	7.62	9.61	11.72	7.36
Amount outstanding ($mm)	265	329	277	321	414	300	231	322
Time-to-maturity	4.74	4.73	4.87	4.77	4.92	5.25	5.23	4.85

7-13-year bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	16	94	288	276	100	26	11	680
Mean number of bonds pr quarter	5.57	9.41	22.6	22.6	7.57	2.25	1.38	58.8
Mean number of quotes pr quarter	19.13	75.52	164.2	195.9	88.45	28.05	29.56	497.6
Age	7.19	6.95	5.67	4.88	2.86	5.21	9.46	5.26
Coupon	7.21	7.29	7.30	7.84	7.76	9.23	9.95	7.64
Amount outstanding ($mm)	400	308	265	294	535	222	270	317
Time-to-maturity	10.37	8.96	9.26	9.14	8.74	8.94	10.14	9.12

13-17-year bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	2	15	58	51	17	9	1	129
Mean number of bonds pr quarter	1	3.8	6.91	5.26	3.81	1.5	1	12.8
Mean number of quotes pr quarter	31.39	8.657	52.87	49.8	22.32	13.88	24.6	106.3
Age	14.91	3.34	8.78	6.07	11.29	12.00	9.31	8.54
Coupon	8.31	8.29	8.09	8.09	9.19	7.36	9.22	8.18
Amount outstanding ($mm)	690	197	323	234	182	155	183	307
Time-to-maturity	15.01	14.93	15.10	14.69	14.69	16.17	14.07	14.94

17-23-year bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	2	15	63	58	26	4	1	139
Mean number of bonds pr quarter	1	4.13	7.35	5.76	4.18	1.24	1	13.5
Mean number of quotes pr quarter	9.22	9.45	61.66	45.8	28.8	29.12	41.43	110
Age	11.17	1.66	6.97	6.85	9.52	9.76	11.31	7.38
Coupon	8.05	7.87	8.26	7.84	8.13	7.15	8.90	8.03
Amount outstanding ($mm)	649	224	364	207	315	206	502	296
Time-to-maturity	17.53	18.92	19.50	20.52	19.19	20.27	22.43	19.90

23-30-year bond maturity

	AAA	AA	A	BBB	BB	B	C	all
Number of bonds	8	24	104	99	24	1	2	196
Mean number of bonds pr quarter	2.43	8.89	15.3	11	12	1	1	27.1
Mean number of quotes pr quarter	5.46	25.14	71.77	86.98	30.6	2	17.13	144.3
Age	1.32	1.85	2.98	3.47	3.54	2.64	9.37	3.21
Coupon	7.14	8.01	7.77	7.24	9.38	7.88	8.89	7.62
Amount outstanding ($mm)	342	199	271	487	283	325	499	370
Time-to-maturity	28.69	27.93	27.03	26.65	26.77	27.38	24.30	26.88

Table 2 *Bond summary statistics.* The sample consists of noncallable bonds with fixed coupons issued by industrial firms. This table shows summary statistics for the data set. Bond yield quotes cover the period 1987Q2-2012Q2.

	#firms	Mean	10th	25th	Median	75th	90th
			Leverage ratio				
AAA	14	0.15	0.07	0.08	0.09	0.15	0.24
AA	61	0.13	0.05	0.09	0.12	0.17	0.22
A	187	0.27	0.13	0.17	0.23	0.35	0.47
BBB	216	0.37	0.19	0.26	0.36	0.47	0.59
BB	106	0.45	0.18	0.29	0.45	0.59	0.72
B	44	0.50	0.21	0.32	0.47	0.70	0.80
C	6	0.80	0.63	0.73	0.85	0.93	0.96
all	423	0.33	0.11	0.18	0.29	0.44	0.60
			Equity volatility				
AAA	14	0.20	0.16	0.17	0.18	0.22	0.26
AA	61	0.27	0.17	0.21	0.25	0.32	0.38
A	187	0.31	0.20	0.24	0.30	0.37	0.42
BBB	216	0.38	0.24	0.28	0.35	0.44	0.56
BB	106	0.46	0.25	0.31	0.42	0.54	0.71
B	44	0.50	0.29	0.33	0.46	0.68	0.79
C	6	0.69	0.31	0.64	0.72	0.76	1.01
all	423	0.36	0.21	0.26	0.33	0.41	0.54
			Asset volatility				
AAA	14	0.18	0.15	0.15	0.19	0.19	0.24
AA	61	0.23	0.19	0.22	0.23	0.25	0.28
A	187	0.24	0.17	0.22	0.24	0.28	0.29
BBB	216	0.25	0.17	0.20	0.24	0.28	0.33
BB	106	0.27	0.17	0.22	0.25	0.29	0.39
B	44	0.28	0.16	0.19	0.26	0.39	0.41
C	6	0.24	0.17	0.17	0.22	0.22	0.42
all	423	0.25	0.17	0.22	0.24	0.28	0.30
			Payout ratio				
AAA	14	0.034	0.012	0.015	0.026	0.047	0.070
AA	61	0.043	0.017	0.030	0.043	0.053	0.066
A	187	0.047	0.019	0.030	0.043	0.057	0.078
BBB	216	0.050	0.018	0.028	0.044	0.066	0.094
BB	106	0.046	0.020	0.027	0.039	0.057	0.081
B	44	0.045	0.019	0.029	0.040	0.058	0.073
C	6	0.064	0.040	0.049	0.056	0.093	0.101
all	423	0.047	0.018	0.029	0.043	0.059	0.082

Table 3 *Firm summary statistics, firms in Compustat with straight bullet bonds outstanding.* 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.

	#firms	Mean	10th	25th	Median	75th	90th
Leverage ratio							
AAA	19	0.12	0.02	0.03	0.07	0.15	0.34
AA	95	0.15	0.04	0.07	0.11	0.18	0.29
A	385	0.20	0.06	0.10	0.17	0.27	0.39
BBB	650	0.28	0.08	0.15	0.25	0.37	0.51
BB	998	0.39	0.13	0.23	0.37	0.54	0.70
B	1014	0.53	0.21	0.35	0.53	0.72	0.86
C	162	0.72	0.38	0.58	0.77	0.89	0.95
all	2087	0.35	0.08	0.16	0.30	0.50	0.71
Equity volatility							
AAA	19	0.26	0.17	0.21	0.26	0.30	0.35
AA	95	0.28	0.19	0.22	0.27	0.32	0.37
A	385	0.31	0.21	0.25	0.30	0.36	0.42
BBB	650	0.37	0.24	0.28	0.34	0.42	0.52
BB	998	0.48	0.31	0.37	0.45	0.56	0.68
B	1014	0.65	0.38	0.48	0.61	0.77	0.94
C	162	0.88	0.53	0.62	0.80	1.05	1.24
all	2087	0.45	0.24	0.30	0.40	0.55	0.72
Asset volatility							
AAA	19	0.23	0.18	0.22	0.23	0.26	0.26
AA	95	0.24	0.20	0.21	0.24	0.26	0.29
A	385	0.26	0.20	0.22	0.24	0.28	0.35
BBB	650	0.28	0.20	0.22	0.27	0.32	0.38
BB	998	0.32	0.21	0.25	0.30	0.38	0.44
B	1014	0.34	0.20	0.25	0.32	0.41	0.51
C	162	0.33	0.18	0.25	0.31	0.40	0.50
all	2087	0.30	0.20	0.23	0.28	0.35	0.43
Payout ratio							
AAA	19	0.027	0.008	0.014	0.023	0.035	0.050
AA	95	0.026	0.005	0.011	0.020	0.034	0.053
A	385	0.032	0.007	0.014	0.026	0.043	0.067
BBB	650	0.037	0.009	0.016	0.030	0.049	0.078
BB	998	0.039	0.009	0.018	0.033	0.054	0.078
B	1014	0.049	0.010	0.024	0.044	0.068	0.092
C	162	0.067	0.022	0.045	0.068	0.089	0.108
all	2087	0.039	0.008	0.017	0.032	0.054	0.081

Table 4 *Firm summary statistics, firms in Compustat with a rating.* For each firm in Compustat for which there is an *S&P* rating in Capital IQ, the leverage ratio, equity volatility, asset volatility, and payout ratio are calculated on December 31 in each year 1985-2015. 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.

horizon (years)	1	2	3	4	5	6	8	10	12	15	20
AAA											
Model	0.00**	0.03	0.08	0.13	0.18	0.24	0.36	0.48	0.60	0.79	1.09
Actual	0.00	0.01	0.03	0.09	0.17	0.25	0.52	0.87	1.16	1.38	1.71
95% c.b.	(0.00;0.00)	(0.00;0.03)	(0.00;0.09)	(0.01;0.22)	(0.04;0.43)	(0.06;0.63)	(0.12;1.29)	(0.21;2.20)	(0.27;2.97)	(0.27;3.77)	(0.27;5.12)
AA											
Model	0.01**	0.06**	0.15	0.26	0.39	0.52	0.78	1.06	1.34	1.76	2.42
Actual	0.07	0.22	0.35	0.54	0.83	1.17	1.83	2.50	3.34	4.52	5.85
95% c.b.	(0.03;0.14)	(0.10;0.40)	(0.15;0.67)	(0.22;1.07)	(0.33;1.67)	(0.45;2.35)	(0.67;3.80)	(0.84;5.35)	(1.08;7.29)	(1.3;10.1)	(1.5;13.8)
A											
Model	0.01**	0.11**	0.29*	0.53	0.81	1.13	1.83	2.57	3.31	4.40	6.06
Actual	0.10	0.31	0.64	0.99	1.38	1.78	2.66	3.62	4.61	5.99	7.93
95% c.b.	(0.05;0.18)	(0.16;0.56)	(0.31;1.14)	(0.46;1.83)	(0.61;2.58)	(0.75;3.41)	(1.05;5.26)	(1.34;7.29)	(1.63;9.46)	(2.0;12.6)	(2.3;17.6)
BBB											
Model	0.24	0.81	1.57	2.41	3.30	4.19	5.92	7.54	9.01	10.97	13.65
Actual	0.29	0.86	1.54	2.29	3.10	3.90	5.46	7.11	8.72	10.87	13.76
95% c.b.	(0.17;0.46)	(0.49;1.37)	(0.84;2.52)	(1.20;3.84)	(1.56;5.28)	(1.91;6.72)	(2.51;9.69)	(3.1;12.9)	(3.7;16.1)	(4.3;20.5)	(4.8;27.1)
BB											
Model	1.84	4.94*	7.92*	10.61	13.00	15.12	18.71	21.63	24.06	27.01	30.68
Actual	1.35	3.27	5.46	7.75	9.92	11.97	15.71	19.27	22.47	26.65	32.06
95% c.b.	(0.92;1.88)	(2.16;4.64)	(3.49;7.91)	(4.8;11.3)	(6.1;14.7)	(7.2;17.9)	(9.1;23.9)	(10.8;29.6)	(12.4;34.8)	(14.1;41.7)	(15.9;51.1)
B											
Model	7.14**	14.72**	20.58**	25.22*	28.97	32.09	36.97	40.64	43.52	46.87	50.84
Actual	3.80	8.71	13.72	18.16	22.06	25.54	31.41	35.89	39.58	44.22	49.14
95% c.b.	(2.85;4.90)	(6.4;11.4)	(9.9;18.1)	(12.9;24.1)	(15.5;29.5)	(17.6;34.3)	(21.3;42.5)	(23.9;49.0)	(25.9;54.3)	(28.2;60.8)	(29.9;68.6)
C											
Model	20.14**	32.93**	40.84**	46.33*	50.44	53.66	58.43	61.85	64.44	67.36	70.73
Actual	14.02	23.81	31.21	36.86	41.40	44.78	49.63	53.88	58.02	63.76	71.34
95% c.b.	(11.6;16.6)	(19.3;28.6)	(25.0;37.8)	(29.2;44.9)	(32.5;50.7)	(34.7;55.0)	(37.8;61.6)	(40.6;66.9)	(43.2;72.0)	(47.4;78.6)	(52.9;86.4)

Table 5 *Average default probabilities in the Black-Cox model and historical default rates.* We merge firm data from CRSP/Compustat with ratings from Standard & Poors and for every firm and every year 1985-2012 we calculate a 1-, 2-,...,19-,20-year default probability in the Black-Cox model. 'Model' shows the average default probabilities. 'Actual' shows Moody's average historical default rate 1920-2012. '95% c.b.' shows the 95% confidence band for the historical default rate calculated following the approach in Section 2.2. '*' and '**' show when the model-implied default probability is outside the 95% and 99% confidence band respectively.

		3-30y	3-7y	7-13y	13-17y	17-23y	23-30y
Inv	Actual spread	97	89	87	96	104	127
	Model spread	105	99	101	98	95	117
	Difference	8	10	14	2	−10	−9
		(1.70)	(1.43)	(1.93)	(0.31)	(1.13)	(1.73)
	Observations	295	294	293	214	262	279
Spec	Actual spread	546	560	417	471	360	364
	Model spread	351	348	375	311	206	127
	Difference	−195**	−212**	−42	−159**	−154**	−238*
		(3.02)	(2.76)	(0.87)	(2.87)	(2.69)	(2.19)
	Observations	289	276	229	98	135	64
AAA	Actual spread	16	4	6	26	2	14
	Model spread	7	2	1	1	8	24
	Difference	−8	−1	−6*	−25**	5	11
		(1.17)	(0.45)	(2.42)	(6.22)	(0.37)	(0.64)
	Observations	140	70	70	76	17	66
AA	Actual spread	25	17	34	31	25	43
	Model spread	8	2	13	15	14	18
	Difference	−16**	−15**	−21**	−16*	−10	−25**
		(3.60)	(3.59)	(3.75)	(2.54)	(1.86)	(3.29)
	Observations	289	279	264	83	88	83
A	Actual spread	66	50	65	63	78	103
	Model spread	82	61	96	88	88	114
	Difference	16**	11	31**	24	10	10
		(2.85)	(1.44)	(3.58)	(1.92)	(1.28)	(0.83)
	Observations	294	294	293	185	210	231
BBB	Actual spread	148	141	141	139	138	165
	Model spread	153	153	157	120	113	136
	Difference	6	11	16	−19	−24*	−29**
		(0.83)	(0.98)	(1.68)	(1.44)	(2.13)	(3.46)
	Observations	295	291	257	177	203	210
BB	Actual spread	377	370	290	387	297	216
	Model spread	325	296	321	209	199	92
	Difference	−53	−74	31	−177**	−98**	−124**
		(1.53)	(1.86)	(0.98)	(2.64)	(5.10)	(7.74)
	Observations	259	240	216	71	104	55
B	Actual spread	674	723	427	488	355	212
	Model spread	407	447	421	382	224	108
	Difference	−267*	−276*	−5	−106	−132**	−104**
		(2.56)	(2.09)	(0.09)	(1.71)	(5.96)	(5.32)
	Observations	243	203	134	54	76	10
C	Actual spread	1422	1211	1948	661	1611	1266
	Model spread	878	1044	761	502	363	339
	Difference	−544	−167	−1187*	−159**	−1248*	−927*
		(1.49)	(0.36)	(2.04)	(4.30)	(2.05)	(2.42)
	Observations	96	65	42	7	15	9

Table 6 *Actual and model yield spreads.* This table shows actual and model-implied corporate bond yield spreads. Spreads are grouped according to remaining bond maturity at the quotation date. 'Actual spread' is the average actual spread to the swap rate. 'Model spread' is the average Black-Cox model spreads of the bonds in a given maturity/rating bucket. The average spread is calculated by first calculating the average spread of bonds in a given month and then calculating the average of these spreads over months. 'Difference' is the difference between the model spread and the actual spread. In parenthesis are the t-statistics, Newey-West corrected with 12 lags, for the monthly differences; '*' implies significance at the 5% level and '**' at the 1% level. 'Observations' is the number of monthly observations. The bond yield spreads are from the period 1987-2012.

		3-30y	3-7y	7-13y	13-17y	17-23y	23-30y
<20bps	Actual spread	7	7	8	13	8	9
	Model spread	13	12	17	11	26	12
	Difference	6^{*}	6	9	-2	18^{**}	3
		(2.04)	(1.58)	(1.79)	(0.36)	(4.09)	(0.52)
	Observations	279	279	214	106	103	89
20-40bps	Actual spread	30	29	30	31	31	33
	Model spread	33	27	50	27	43	27
	Difference	3	-2	19^{**}	-4	12	-5
		(0.70)	(0.43)	(2.82)	(0.86)	(1.42)	(0.86)
	Observations	280	272	233	129	123	113
40-70bps	Actual spread	55	54	55	55	57	56
	Model spread	62	62	78	51	58	48
	Difference	7	8	23^{**}	-4	1	-8
		(1.36)	(0.89)	(3.08)	(0.49)	(0.19)	(0.92)
	Observations	288	277	262	148	145	168
70-100bps	Actual spread	84	85	84	84	86	84
	Model spread	91	104	110	86	80	78
	Difference	7	19	26^{*}	2	-6	-6
		(0.93)	(1.47)	(2.34)	(0.18)	(0.58)	(0.64)
	Observations	282	264	246	141	180	172
100-150bps	Actual spread	121	121	120	119	121	121
	Model spread	124	131	131	127	100	110
	Difference	3	11	11	8	-21^{*}	-11
		(0.39)	(0.74)	(0.88)	(0.42)	(2.51)	(0.77)
	Observations	269	260	254	127	210	164
150-200bps	Actual spread	172	172	172	170	171	173
	Model spread	160	152	177	157	118	136
	Difference	-12	-20	4	-13	-52^{**}	-37^{*}
		(0.94)	(1.49)	(0.27)	(0.49)	(4.80)	(2.01)
	Observations	256	222	211	65	139	146
200-300bps	Actual spread	242	242	243	245	240	237
	Model spread	233	230	289	194	166	185
	Difference	-9	-12	46	-51	-74^{**}	-52^{**}
		(0.65)	(0.65)	(1.57)	(1.41)	(3.93)	(2.60)
	Observations	271	222	220	82	101	139
300-1000bps	Actual spread	492	523	456	537	440	411
	Model spread	460	522	478	366	288	250
	Difference	-32	-1	22	-171^{**}	-152^{**}	-161^{**}
		(0.77)	(0.02)	(0.55)	(7.52)	(6.40)	(3.66)
	Observations	269	244	221	100	93	60
>1000bps	Actual spread	1748	1746	1789	1180	2470	3569
	Model spread	854	941	708	497	405	407
	Difference	-894^{**}	-805^{*}	-1081^{**}	-683^{**}	-2065^{**}	-3161^{**}
		(3.84)	(2.55)	(3.40)	(23.30)	(5.91)	(84.66)
	Observations	132	89	59	13	8	2

Table 7 *Actual and model yield spreads sorted according to actual spread.* This table shows actual and model-implied corporate bond yield spreads. Spreads are grouped according to the size of the actual spread and the remaining bond maturity at the quotation date. 'Actual spread' is the average actual spread to the swap rate. 'Model spread' is the average Black-Cox model spreads of the bonds in a given maturity/rating bucket. The average spread is calculated by first calculating the average spread of bonds in a given month and then calculating the average of these spreads over months. 'Difference' is the difference between the model spread and the actual spread. In parenthesis are the t-statistics, Newey-West corrected with 12 lags, for the monthly differences; '*' implies significance at the 5% level and '**' at the 1% level. 'Observations' is the number of monthly observations. The bond yield spreads are from the period 1987-2012.

		3-30y	3-7y	7-13y	13-17y	17-23y	23-30y
Inv	β	0.93	0.86	0.85	0.78	0.56^{**}	0.89
		(0.06)	(0.08)	(0.09)	(0.14)	(0.08)	(0.08)
	R^2	0.88	0.85	0.70	0.75	0.44	0.81
Spec	β	0.86	0.81	0.82	1.39	2.30^{*}	4.42^{*}
		(0.40)	(0.42)	(0.20)	(0.23)	(0.58)	(1.37)
	R^2	0.14	0.13	0.25	0.64	0.51	0.52
AAA	β	-0.29^{**}		1.52^{**}	1.81^{**}	2.54^{**}	
		(0.22)		(0.15)	(0.21)	(0.29)	
	R^2	0.05		0.26	0.18	0.79	
AA	β	1.07	0.95	1.69^{**}		-0.02^{**}	0.16^{**}
		(0.60)	(0.40)	(0.24)		(0.25)	(0.04)
	R^2	0.19	0.03	0.62		0.00	0.06
A	β	0.70^{**}	0.56^{*}	0.57^{**}	0.37^{**}	0.49^{**}	0.62^{**}
		(0.05)	(0.20)	(0.11)	(0.04)	(0.08)	(0.10)
	R^2	0.74	0.48	0.58	0.70	0.48	0.67
BBB	β	0.88	0.75^{**}	0.92	0.82	0.58^{**}	0.84
		(0.07)	(0.09)	(0.10)	(0.14)	(0.11)	(0.14)
	R^2	0.87	0.81	0.74	0.71	0.50	0.68
BB	β	0.74	0.73	0.65^{*}	1.78^{*}	1.38^{**}	2.20^{**}
		(0.29)	(0.32)	(0.15)	(0.40)	(0.11)	(0.15)
	R^2	0.42	0.39	0.48	0.78	0.89	0.88
B	β	0.67	0.62	0.32^{**}	0.87	1.27	
		(0.27)	(0.30)	(0.20)	(0.20)	(0.22)	
	R^2	0.09	0.08	0.06	0.50	0.70	
C	β	-0.79^{**}	-0.36^{*}				
		(0.56)	(0.59)				
	R^2	0.04	0.01				

Table 8 *Commonality in time series variation of actual and model-implied yield spreads.* For a given rating and maturity group we calculate a monthly average spread by computing the average yield spread of bond observations with the corresponding rating and maturity in that month. We do this for both model-implied spreads and actual spreads (to the swap rate) resulting in a time series of monthly actual spreads $s_1, s_2, ..., s_T$ and implied spreads from the Black-Cox model $\hat{s}_1, \hat{s}_2, ..., \hat{s}_T$ for the period 1987-2012. The table shows the regression coefficient in the regression of the actual spread on the model-implied spread $s_t = \alpha + \beta \hat{s}_t + \epsilon_t$. In parenthesis is the standard error, Newey-West corrected with 12 lags and '*' implies that β is significantly different from one at the 5% level and '**' at the 1% level. In some months there may not be any observations and if there are less than 100 monthly observations we do not report regression coefficients.

		correlation of pricing error with			
	R^2	L_t	σ_t^e	σ_t^a	δ_t
Investment grade	0.44	-0.39	-0.11	0.11	-0.33
Speculative grade	0.13	0.02	-0.00	-0.13	0.02

Table 9 *Explaining individual pricing errors.* The first column shows the R^2 from running the regression of actual spreads on the implied spreads from the Black-Cox where we use all transactions in the data sample, separated into investment grade and speculative grade. The next columns show the correlation between the pricing error, defined as the difference between the actual spread and model-implied spread, and variables that may contribute to pricing errors. L_t is the leverage ratio on the day of the transaction, σ_t^e is the estimated equity volatility on the day of the transaction, σ_t^a is the issuing firm's asset volatility when estimated day-by-day, δ_t is the payout rate on the day of the transaction. The bond yield spreads are from the period 1987-2012.

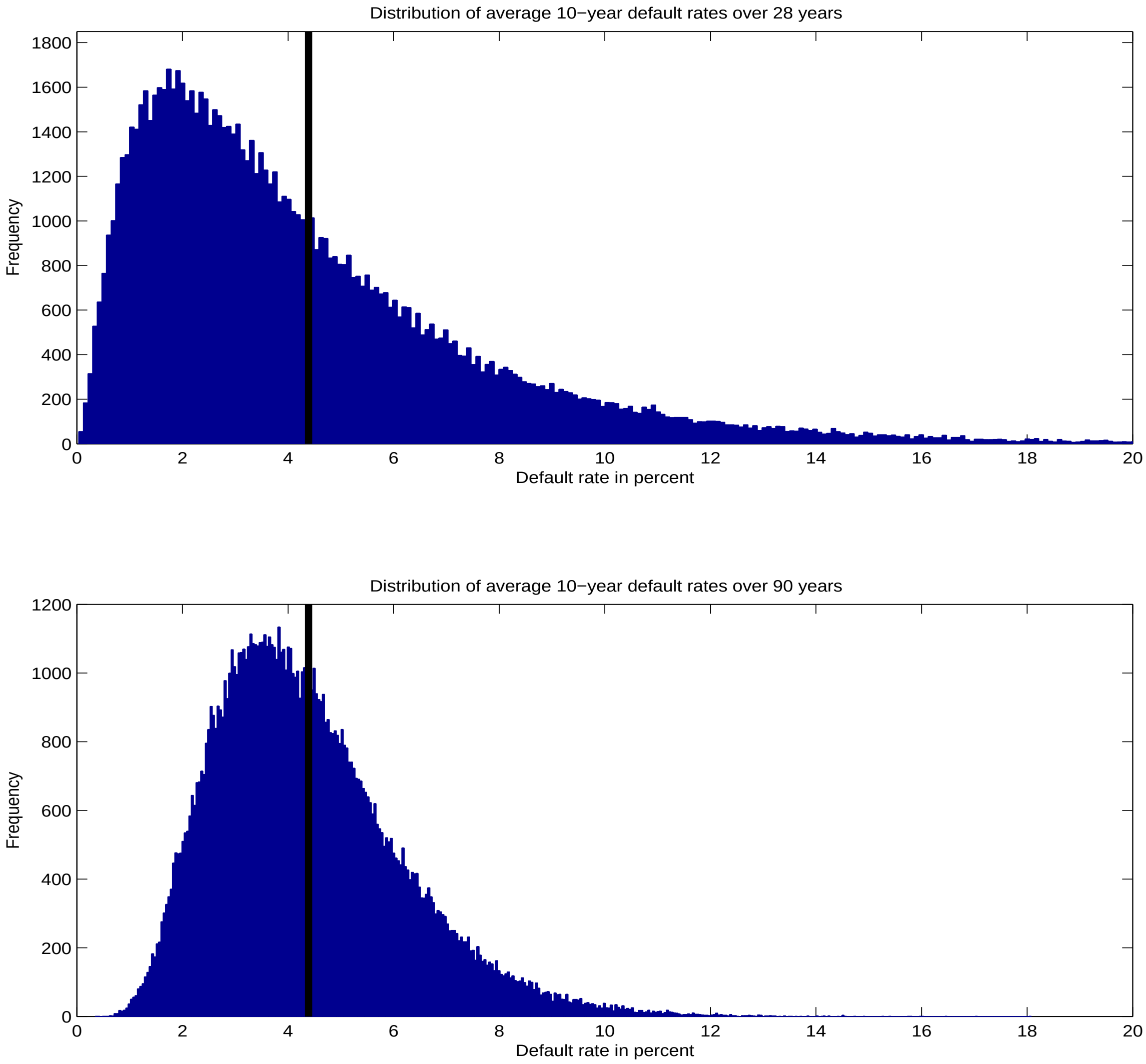

Fig. 2 *Distribution of realized default rate.* The figure shows the results of a simulation of the realized cumulative 10-year default rate when the mean cumulative 10-year default rate is equal to 4.39%, the rate for BBB-rated issuers published by Moody's using data for 1970-1998. We assume that in 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%.

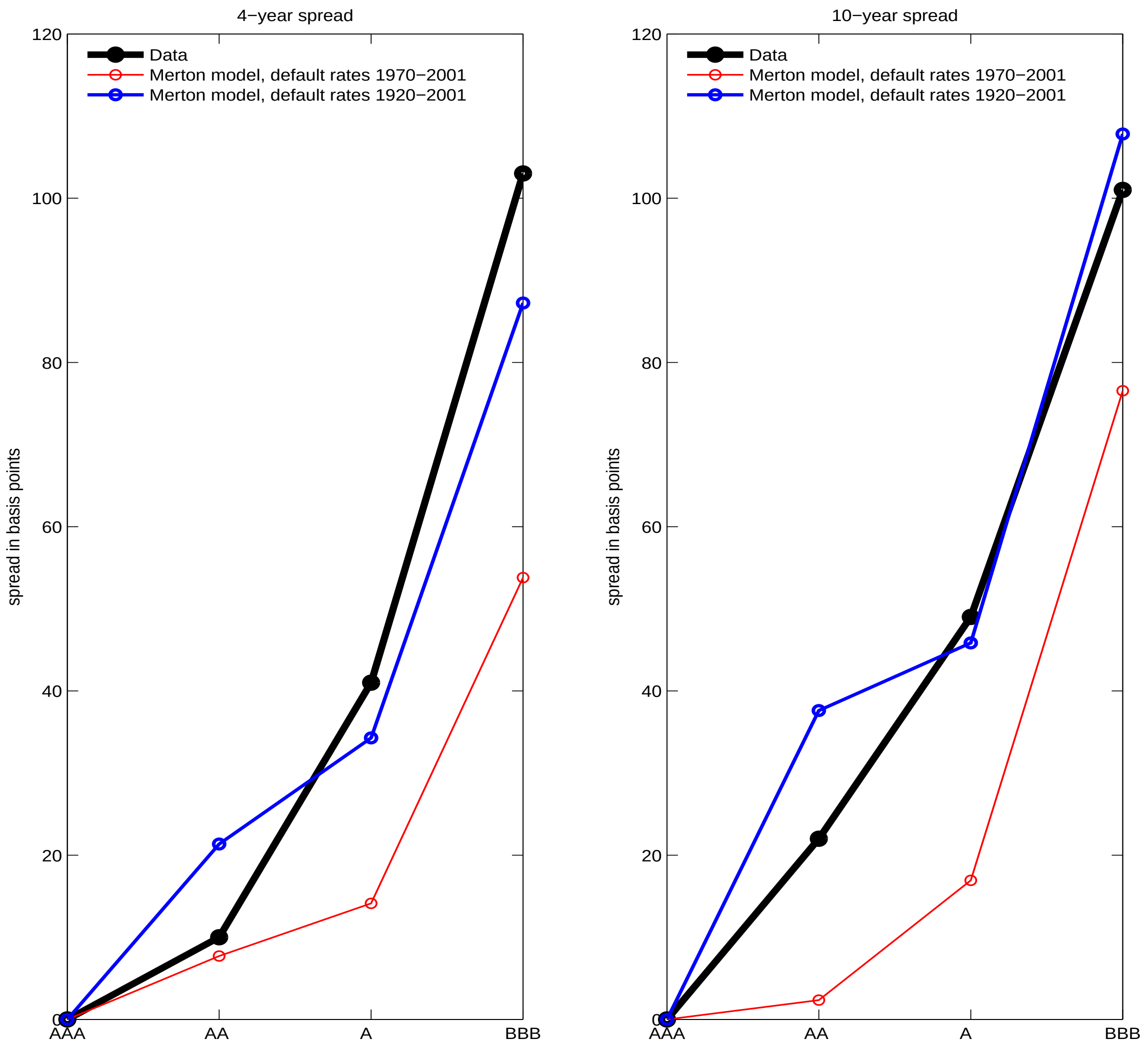

Fig. 3 *Corporate spreads to AAA-bond yields in the Merton model using default rates from different periods.* This figure shows actual and model-implied spreads to AAA yields. The red line shows spreads in the Merton model based on Moody's default rates from the period 1970-2001 and corresponds exactly to the calculations in Chen, Collin-Dufresne, and Goldstein(2009). The blue 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).

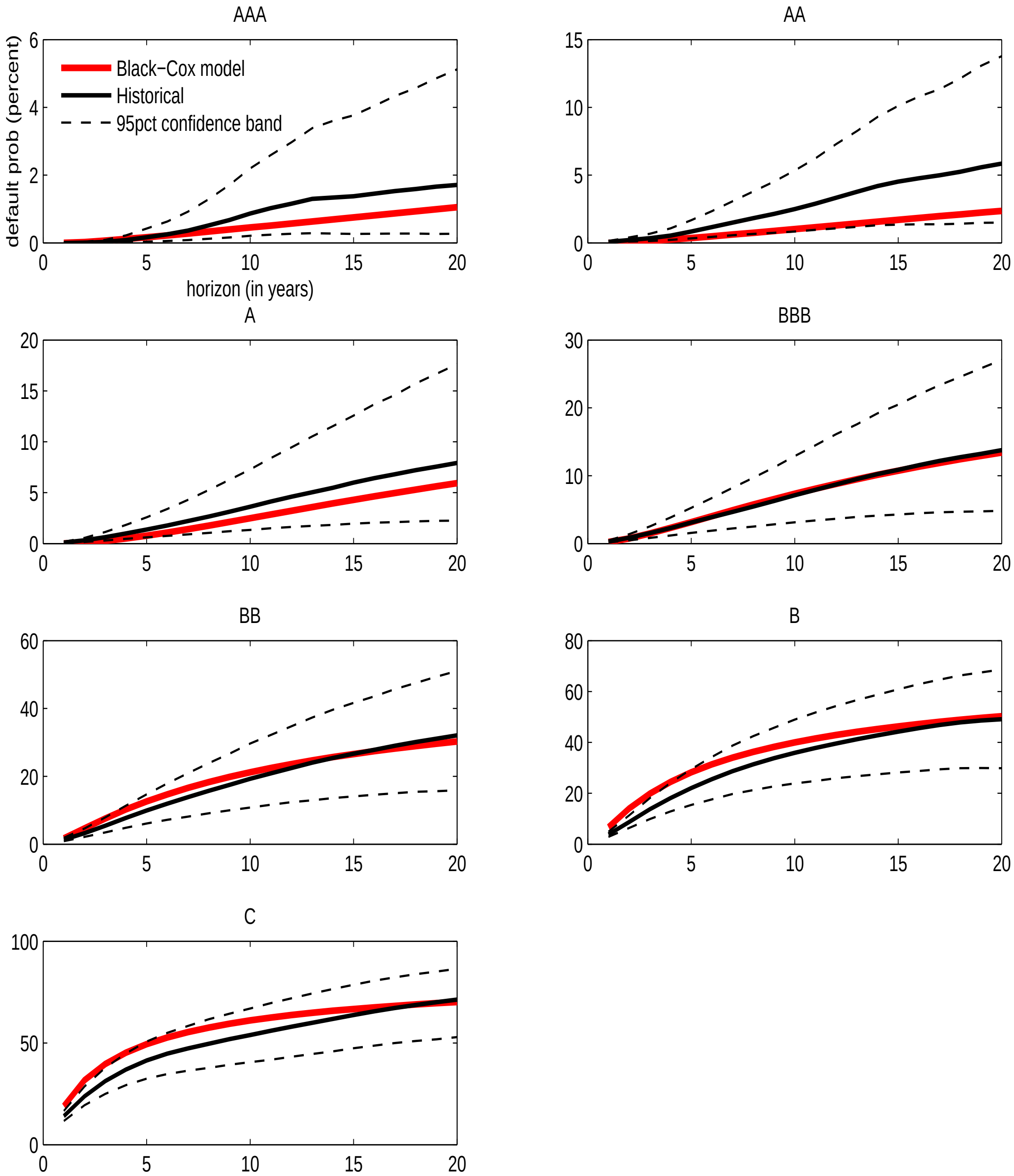

Fig. 4 *Average default probabilities in the Black-Cox model and historical default rates.* We merge firm data from CRSP/Compustat with ratings from Standard & Poors and for every firm and every year 1985-2012 we calculate a 1-, 2-,...,19-,20-year default probability in the Black-Cox model. The figure shows the average default probabilities along with the average historical default rate 1920-2012 calculated by Moody's. A 95% confidence band for the historical default rate is calculated following the approach in Section 2.2.

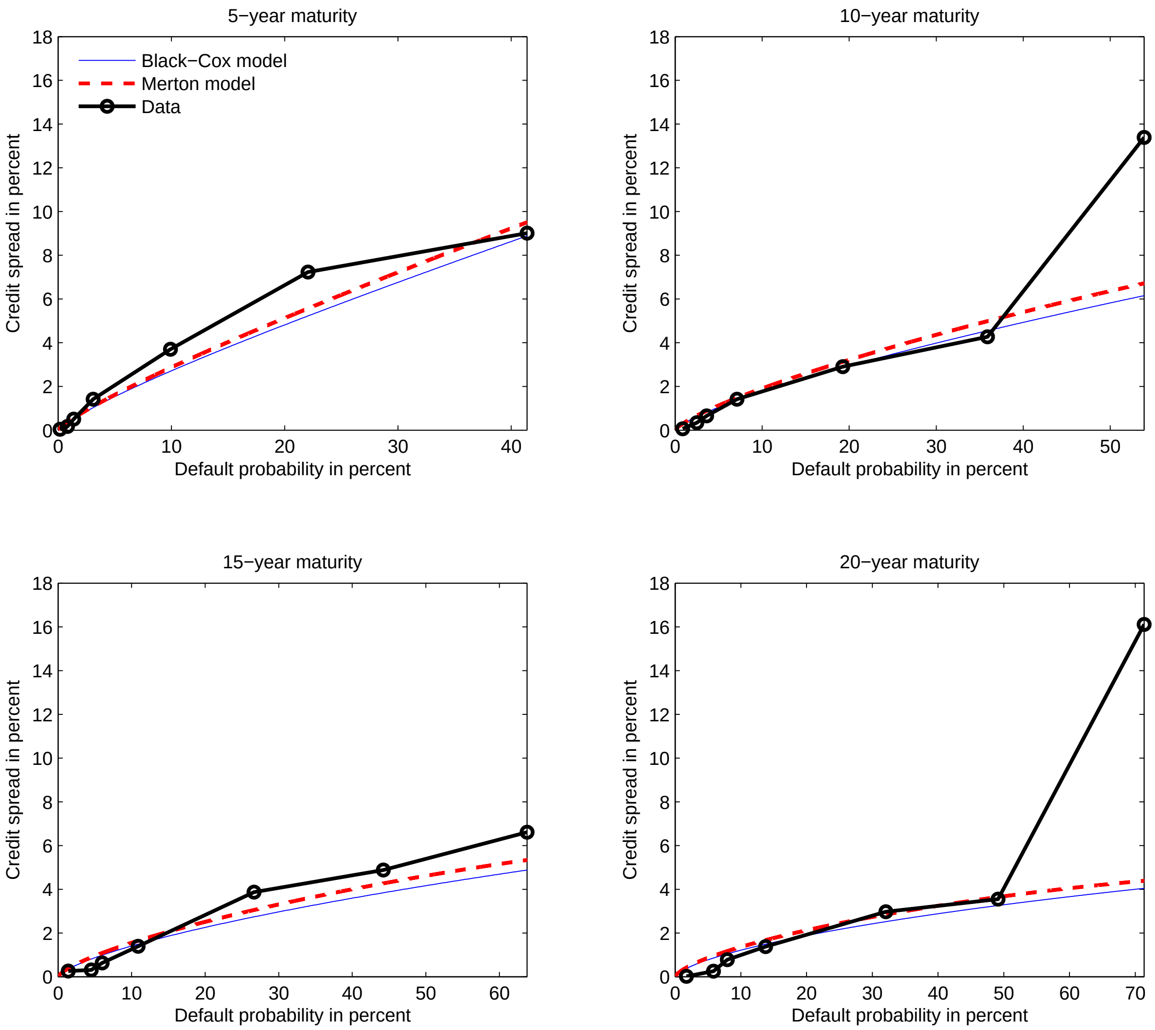

Fig. 5 *Credit spreads and default probabilities.* The panels show for different maturities a plot of the credit spread against the default probability in the data and in the Black-Cox and Merton models. To calculate model-implied spreads and default probabilities in the Merton model we use equation (2) with a Sharpe ratio of 0.22 and recovery rate of 37.8%. To calculate model-implied spreads and default probabilities in the Black-Cox model we use equations (9) and (10) and vary leverage and use an asset volatility of 22%, payout ratio of 4%, recovery rate of 37.8%, Sharpe ratio of 0.22, and a riskfree rate of 5%. To calculate actual spreads and default probabilities we use for different ratings the spreads in Table 6 and Moody's historical default rates 1920-2012 (each circle represents a rating category).

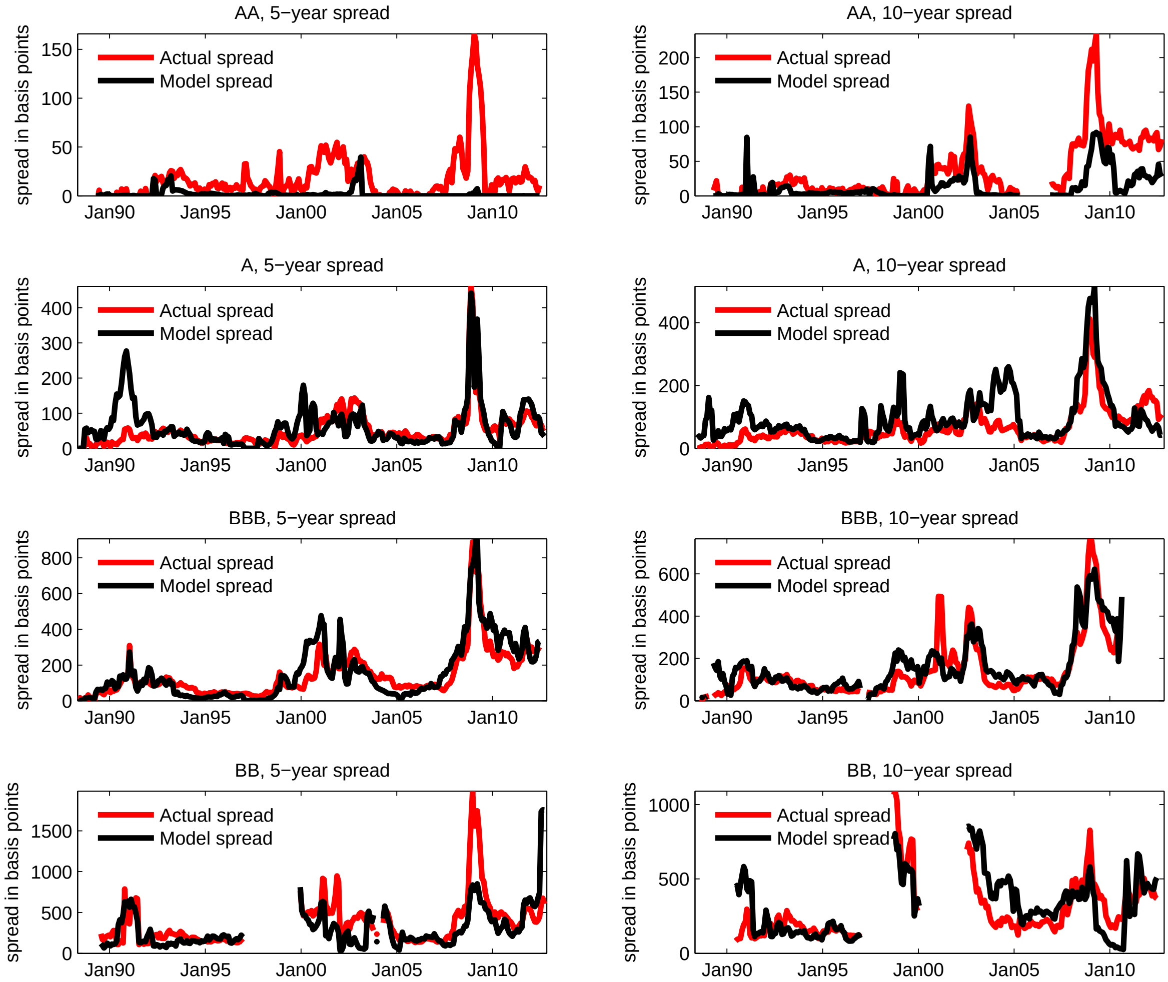

Fig. 6 *Time series variation in corporate bond yield spreads.* This graph shows the time series of actual and model-implied 5-year and 10-year corporate bond spreads, where the model-implied spreads are from the Black-Cox model. Each month all daily yield observations in bonds in a given rating category and with a maturity between 3-7 and 7-13 years are collected and the average actual spread (to the swap rate) and the average model-implied spread are computed. The graphs the time series of monthly spreads for rating and maturities for which there are more than 210 monthly observations.