An Empirical Evaluation of Structural Credit-Risk Models*

Nikola A. Tarashev
Bank for International Settlements

This paper evaluates the capacity of five structural credit-risk models to forecast default rates. In contrast to previous studies with similar objectives, the paper employs firm-level data and finds that model-based forecasts of default rates tend to be unbiased and to deliver point-in-time errors that are small in both statistical and economic terms. In addition, in- and out-of-sample regression analysis reveals that the models account for a significant portion of the variability of credit risk over time but fail to fully reflect its dependence on macroeconomic cycles.

JEL Codes: G33, E44, G28, C13.

1. Introduction

Accurate measures of credit risk are a key prerequisite for sound risk management and for meaningful supervisory evaluation of the vulnerability of lender institutions. In an appreciation of this, the new capital adequacy framework (known as Basel II) will require banks to validate the appropriateness of their credit-risk models, some of which are likely to incorporate—explicitly or implicitly—constructs developed in the academic literature.

*I am indebted to Claudio Borio and Kostas Tsatsaronis for their encouragement and for several insightful discussions on the topic. An anonymous referee, Jeff Amato, Ingo Fender, Michael Gordy, Hirotaka Hideshima, Ricardo Schechtman, Richard Vasicek, Haibin Zhu, and seminar participants at the Bank for International Settlements (BIS), Banque de France, the Vienna Workshop of the Basel Committee's Research Task Force, and Karlsruhe University provided useful comments on earlier drafts of this paper. I also thank Dimitrios Karampatos for valuable help with the data and Melanie Sykes for putting the Word document together. The views expressed in this paper are my own and do not necessarily reflect those of the BIS. Author contact: Bank for International Settlements, Centralbahnplatz 2, CH-4052 Basel, Switzerland; e-mail: nikola.tarashev@bis.org.

The academic literature has produced a variety of theoretical models of credit risk. This paper examines the family of *structural* models, which consider the stochastic process of a corporate borrower's assets and postulate that a default occurs when these assets fall below a threshold value. The structural models can be divided into an "endogenous-default" group and an "exogenous-default" group. Models in the former group let borrowers choose strategically the timing of default. In contrast, models in the latter group impose an ad hoc default threshold but develop richer stochastic environments in order to capture empirical regularities of credit markets that are unaccounted for by endogenous-default models.

The paper analyzes the extent to which model-based default-rate expectations—defined as the cross-sectional averages of model-based probabilities of default (PDs)[1]—account for the level and intertemporal evolution of actual default rates in three (BBB, BB, and B) rating classes. The analysis considers five structural models, which are calibrated to firm-specific borrower characteristics that are allowed to change over time. Two of the models incorporate endogenous defaults and are developed in Leland and Toft (1996) (henceforth, LT) and Anderson, Sundaresan, and Tychon (1996) (AST). The other three are exogenous-default models, developed in Longstaff and Schwartz (1995) (LS), Collin-Dufresne and Goldstein (2001) (CDG), and Huang and Huang (2003) (HH).[2]

One of the main conclusions of this paper is that the structural credit-risk models tend to account for the level of default rates. First, model-based expectations of default rates are generally unbiased at both the one- and five-year horizons. Second, the point-in-time errors in these expectations tend to be small in statistical terms. Third, these errors also appear to be of little economic importance, as they tend to translate into small discrepancies between (i) capital buffers derived under the internal-ratings-based (IRB) approach of Basel II on the basis of model-implied expected default rates and (ii) capital

[1]The paper abstracts from other components of credit risk, such as loss given default and exposure at default.

[2]It is of natural interest to compare the performance of these academic models with the performance of commercial models, such as the one underpinning the popular expected default frequencies of Moody's KMV. Data limitations, however, do not allow for including such comparisons in this paper.

buffers derived under the same approach but incorporating actual default rates as (ex post) measures of credit risk.[3]

These general findings mask important differences across models, which are due mainly to differences in the way that the models incorporate two borrower characteristics: the default threshold and the dead-weight cost at default. The endogenous-default LT and AST models allow the two characteristics to be extracted simultaneously from the data. Under the LT model, such an extraction leads to expected default rates that match the level of actual default rates closely and consistently across horizons and rating classes. Under the AST model, however, the default thresholds appear to be overly sensitive to changes in asset volatility, which leads to underpredictions of one-year default rates.

In turn, a calibration scheme that is applicable to the exogenous-default LS, HH, and CDG models needs to incorporate an ad hoc parameter value. In the specific application, such a value is assigned to the dead-weight cost, which is set equal to the average of its counterpart in the LT model and then is used for calculating exogenous-default thresholds. Under the LS and HH models, this leads to generally accurate default-rate expectations. By contrast, the CDG model, which forces the default threshold to gravitate too strongly toward a level that guarantees solvency, consistently underestimates default rates.

Owing to their calibration to borrower-specific parameters, all five models deliver expectation errors that are orders of magnitude smaller than those reported by Leland (2004). That paper calibrates the LT and LS models to the *representative* borrower and concludes that they substantially underestimate default rates at short—e.g., one-year—horizons. The difference between the conclusion of Leland (2004) and the finding of this paper is due to the strongly nonlinear relationship between inputs to the models and model-implied PDs. The nonlinearities give rise to the so-called Jensen inequality effect, whereby the average PD across borrowers—i.e., the model-based expectation of the default rate—is larger than the PD of the average—i.e., representative—borrower.

[3]See Basel Committee on Banking Supervision (2004).

The Jensen inequality effect is also the reason why the models underpredict credit risk when their calibration differentiates borrowers only on the basis of time averages of borrower-specific characteristics. Such calibration decreases the effective cross-sectional heterogeneity of the borrowers, which dampens the Jensen inequality effect and, thus, excessively depresses model-based expectations of default rates.

In addition, this paper conducts regression analysis in order to study the extent to which structural credit-risk models account for the evolution of default rates over time. All regressions include model-based expected default rates as explanatory variables. Furthermore, the analysis incorporates macroeconomic control variables pointed out by Estrella and Hardouvelis (1991) and Smets and Tsatsaronis (1997), who identify predictors of economic activity, and by Borio and Lowe (2002), who identify predictors of banking-system distress.

The results of the regression analysis, which are confirmed in out-of-sample exercises, indicate that structural credit-risk models account to an extent for the intertemporal evolution of credit risk but fail to reflect fully its dependence on the business and credit cycles. The models perform best when calibrated to time-varying, as opposed to average, borrower-specific parameters. In addition, simultaneously using the expected default rates implied by several models improves on the fit of single-model forecasts. Exploiting these regularities reveals "models-only" regressions that attain adjusted R^2 of 22 percent, 66 percent, and 63 percent in the BBB, BB, and B rating classes, respectively, which is in line with a finding of van Deventer, Li, and Wang (2006). At the same time, PDs implied by the structural models miss important information regarding the impact of real and financial activity on credit risk. This is seen in the finding that the macroeconomic controls add to and, for the BBB rating class, even substitute for the information content of the models.

The rest of the paper is organized as follows. After a review of the related literature in section 2, section 3 presents the five models and highlights their key assumptions. Then, section 4 describes the data, and section 5 outlines and discusses how the data are employed for the calibration of the models. Section 6 presents and discusses model-based predictions of one- and five-year default rates.

Considering mainly the one-year horizon, sections 7 and 8 evaluate, respectively, the statistical and economic significance of errors in the models' default-rate forecasts. Finally, section 9 analyzes the significance of one-year model-based PDs as in-sample explanatory variables and out-of-sample forecasters of default rates.

2. Related Literature

The analysis in this paper—which is in the spirit of the aforementioned Leland (2004) and van Deventer, Li, and Wang (2006)—contributes to a large and growing literature on the performance of structural credit-risk models.

A sizable portion of this literature consists of studies conducted by Moody's KMV staff. These studies perform two general exercises that analyze the capacity of the structural model underpinning Moody's KMV PD estimates to (i) account for the level of observed default rates and (ii) accurately differentiate credit risk across firms.[4] The first exercise is paralleled in this paper in the context of academic models. The second exercise rests on a large proprietary data set of default events and is thus beyond the scope of this paper. Importantly, these exercises reveal that the performance of the Moody's KMV model depends crucially on its flexible empirical implementation. Consequently, it is useful to keep in mind that the default-rate forecasts examined in this paper are based on a *direct* implementation of academic models.

In addition, a number of articles have examined the degree to which (variants of) the Black-Scholes-Merton (BSM) model can help predict credit risk. Hillegeist et al. (2004), e.g., incorporate BSM PDs in a hazard model and find that they contain significantly more information on credit risk than popular accounting-based measures. In related exercises, Bharath and Shumway (2004) and Duffie, Saita, and Wang (2007) reach the conclusion that a volatility-adjusted leverage—which is a key parameter in the BSM model and is also known as distance to default (DD)—plays an important role in, respectively, a sufficient statistic for the PD and forecasts of the PD term structure. Duffie, Saita, and Wang (2007) also find that the

[4]See, e.g., Stein (2000), Kealhofer and Kurbat (2001), and Bohn, Arora, and Korablev (2005).

information content of the DD variable is usefully complemented by macroeconomic covariates, a result echoed by the regression analysis below.

Furthermore, several articles have examined the models considered in this paper in order to evaluate their capacity to accurately price credit risk. Huang and Huang (2003), e.g., calibrate the models to observed default rates and then analyze the implied credit-risk premiums. For their part, Eom, Helwege, and Huang (2004) study individual bonds in order to compare their credit spreads to the predictions of structural credit-risk models. While these articles examine theoretical pricing implications, which are based on preference-weighted probabilities that should reflect bond holders' risk appetite, the analysis herein focuses on actual (or physical) PDs.

Finally, given that the considered models pertain to single-name credit risk, the analysis in this paper only touches on issues related to the interdependence of default events across borrowers. For recent advances in the treatment of such issues, which play a key role in estimates of portfolio credit risk, see, e.g., Basel Committee on Banking Supervision (2006) and the references therein, as well as Das et al. (2007).

3. The Models

Constructed in the spirit of the Merton (1974) framework, the five structural credit-risk models examined in this paper focus on an individual borrower that defaults when its assets, V, fall below a particular threshold, V^*. All models assume that the asset value evolves as follows:

$$dV_t/V_t = (r + \lambda - \delta)dt + \sigma dW, \tag{1}$$

where r denotes the risk-free rate, λ is the asset risk premium, δ is the asset payout ratio (reflecting, e.g., dividend and coupon payments), W is a standard Wiener process, and σ is the instantaneous asset volatility.

The PD over a particular horizon equals the probability that the first passage of V below V^* occurs over this horizon. The nature of the default threshold differs across models and makes it natural to divide them into two categories. Those in the first category—i.e., the

HH, LS, and CDG models—adopt an exogenous threshold, whose value has to be determined outside of the model. In contrast, the models in the second category—i.e., the LT and the AST models—derive the decision to default endogenously, as part of the borrower's optimization problem. Thus, the default threshold in these models is a function of borrower characteristics.

Given an initial value of assets, V_0, all five models imply the following:

$$\frac{\partial PD}{\partial V^*} > 0, \quad \frac{\partial PD}{\partial \sigma} > 0, \quad \frac{\partial PD}{\partial r} < 0, \quad \frac{\partial PD}{\partial \lambda} < 0, \quad \frac{\partial PD}{\partial \delta} > 0. \quad (2)$$

Intuitively, the PD increases in the default threshold, V^*, and in the level of risk, as captured by asset volatility σ. The implications of the remaining three parameters, which determine the drift in the value of assets, are best considered together. Tight credit conditions—caused by contractionary monetary policy that raises the risk-free rate, r, and/or by stronger aversion to risk that raises the risk premium, λ—seem to counterintuitively depress PDs. At the same time, however, tight credit conditions would tend to raise the payout ratio, δ, which increases PDs.

3.1 Exogenous-Default Models

Besides leaving the value of the default threshold to the discretion of the empirical researcher, the three exogenous-default models are similar in that they assume that all debt is of infinite maturity. This assumption delivers analytic tractability but makes it impossible to capture the empirical regularity that borrowers are less likely to default over a given horizon if they are to repay the debt principal further in the future.

What differs across exogenous-default models are their assumptions regarding the overall stochastic environment. The remainder of this subsection discusses these assumptions, paying particular attention to their implications for theoretical PDs.

3.1.1 The Model of Longstaff and Schwartz (1995)

In order to capture the cooling effect of rises in the interest rate on the macroeconomy, the LS model allows for a stochastic risk-free rate of return that correlates negatively with shocks to asset returns:

$$dr_t = k_r(\bar{r} - r_t)dt + \sigma_r dW_t^r, \qquad (3)$$

where $\bar{r}$ is the long-run risk-free rate of return, k_r reflects the speed of mean reversion, σ_r is the instantaneous volatility of the risk-free rate of return, and $corr(dW_t^r, dW_t) \equiv \sigma_{rV} < 0$.

In such a setting, a change of r has an ambiguous impact on the PD. By expression (1), a higher interest rate increases the deterministic drift in the value of assets and, ceteris paribus, lowers the PD. At the same time, the assumption $\sigma_{rV} < 0$ implies that a higher r tends to be associated with a negative shock to asset returns, which raises the PD. The relative importance of the latter impact increases when asset volatility, σ, is higher and/or the correlation $\frac{\sigma_{rV}}{\sigma_r \sigma}$ is closer to -1.

3.1.2 The Model of Collin-Dufresne and Goldstein (2001)

Collin-Dufresne and Goldstein observe that debt issuance increases when the value of assets increases, as a result of which the leverage ratio (i.e., the ratio of debt to assets) exhibits mean reversion. The CDG model accommodates this empirical regularity, which implies that the default threshold, V_t^*, moves in step with the value of assets, V_t. Under the maintained assumption that V_t^* is a constant fraction of debt,

$$d \ln V_t^* = k_l \big(\ln V_t - \ln V_t^* - \nu \big) dt, \text{ where } k_l > 0, \nu > 0. \qquad (4)$$

The parameters ν and k_l affect theoretical PDs as follows. The restriction $\nu > 0$ implies that the borrower is inherently solvent, as its assets would stay above the default threshold in the absence of shocks. That said, the closer ν is to zero, the closer V_t^* and V_t tend to stay to each other and, thus, the higher the risk of V_t falling below V_t^*, which causes a default. In turn, a higher k_l implies that the ratio $\ln(V_t/V_t^*)$ is more likely to stay close to its long-run value ν. Since $\nu > 0$, an increase in k_l lowers the PD.

3.1.3 The Model of Huang and Huang (2003)

There is empirical evidence that equity risk premiums are negatively correlated with returns on broad equity indices. On the basis of such evidence, the HH model incorporates a stochastic risk premium that correlates negatively with shocks to the assets of the typical borrower:

$$d\lambda_t = k_\lambda(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda dW_t^\lambda, \quad corr\left(dW_t^\lambda, dW_t\right) \equiv \sigma_{\lambda V} < 0. \quad (5)$$

A higher $\bar{\lambda}$ implies a higher long-run drift in the value of assets, which, ceteris paribus, lowers the PD. The impact of $\bar{\lambda}$ is stronger the larger is the mean-reversion parameter k_λ. In addition, since $\sigma_{\lambda V} < 0$, a negative value of dW_t (which puts upward pressure on the PD) tends to be counteracted by an increase of the drift in the value of assets.

3.2 Endogenous-Default Models

The two endogenous-default models, which let the borrower decide when to default, differ in their assumptions regarding the debt contract. While the AST model allows the borrower to renege on and then alter the terms of its contract, no renegotiation is possible in the LT model, in which a defaulting firm is surrendered to creditors. In addition, while the AST framework focuses on perpetual bonds, the LT model assumes continuous issuance of debt that is of constant but finite time to maturity.

3.2.1 The Default Threshold in the Model of Anderson, Sundaresan, and Tychon (1996)

At the time of default, creditors in the AST model choose between two options. They can either liquidate the borrowing firm and seize its assets (net of a bankruptcy cost) or accept the terms of a new debt contract. Since a liquidation of the borrowing firm is the worst possible outcome for its equity holders, they propose a postdefault contract that is acceptable to creditors.

In this setup, the default threshold is determined by the borrower's capacity and willingness to renegotiate its debt contract without being liquidated. Thus, a higher (fixed) dead-weight cost, K—incurred by creditors only if they liquidate the borrower—boosts the borrower's bargaining power and allows it to set a higher default threshold V_{AST}^*. In addition, V_{AST}^* is also raised by an upward shift in the predefault value of debt, which—being equivalent to an upward shift in the debt-service burden—prompts borrowers to negotiate a more advantageous contract earlier. Consequently, excluding K, the impact of the remaining model parameters on V_{AST}^* is a result of their impact on the predefault value of debt. In concrete terms, this value is raised by a higher risk-neutral drift in the asset process (i.e., a higher r[5] or a lower δ); a higher debt principal, P; a higher coupon rate, c; a lower asset volatility, σ; and a lower monitoring cost, m.[6] In sum,

$$\frac{dV_{AST}^*}{dK} > 0, \frac{dV_{AST}^*}{dr} > 0, \frac{dV_{AST}^*}{d\delta} < 0, \frac{dV_{AST}^*}{dP} > 0,$$
$$\frac{dV_{AST}^*}{dc} > 0, \frac{dV_{AST}^*}{d\sigma} < 0, \frac{dV_{AST}^*}{dm} < 0. \tag{6}$$

3.2.2 The Default Threshold in the Model of Leland and Toft (1996)

In the LT model, a borrower forfeits its equity value as soon as it does not fulfill a contracted obligation. Thus, the willingness to service debt increases—i.e., the default threshold V_{LT}^* decreases—with increasing value of equity, which equals the value of the firm net of the value of its debt.[7] On the one hand, the value of the firm decreases in the default cost (which is a fraction α of assets) but, being assessed over the infinite horizon, is insensitive to the time to maturity, T, of the continuously issued debt contracts. On the other hand, the value of debt contracts decreases in α (but by less than the value of the firm) and in T, a rise in which heightens the risk that the borrower defaults before the contract matures. The upshot

[5]In the empirical exercises reported below, this positive effect of a rise in r on the predefault value of debt is weakened but never reversed by the fact that a rise in r also lowers the present value of coupon payments.

[6]The original AST model does not incorporate monitoring costs. The introduction of such costs generalizes the original formula for V_{AST}^* by replacing the coupon payments cP with $(1-m)cP$.

[7]In the LT model, the (market) value of the firm equals the asset value plus the value of tax benefits, less the value of bankruptcy costs, over the infinite horizon.

is that the default threshold V_{LT}^* increases in the default cost but decreases in the time to debt maturity.

The implications of the other model parameters are rationalized similarly. A higher coupon rate, c; principal, P; or asset payout rate, δ; decreases the value of equity, while a higher risk-free rate, r; asset volatility, σ; or tax benefits, τ; increases the value of equity. Thus,

$$\frac{dV_{LT}^*}{d\alpha} > 0, \frac{dV_{LT}^*}{dT} < 0, \frac{dV_{LT}^*}{dc} > 0, \frac{dV_{LT}^*}{dP} > 0, \frac{dV_{LT}^*}{d\delta} > 0,$$
$$\frac{dV_{LT}^*}{dr} < 0, \frac{dV_{LT}^*}{d\sigma} < 0, \frac{dV_{LT}^*}{d\tau} < 0. \quad (7)$$

4. Data: Sources and Descriptive Statistics

Data availability limits the analysis to corporate borrowers domiciled in the United States. Firm-specific borrower and debt characteristics are provided by Moody's (rating of senior unsecured debt,[8] coupon rate, and time to maturity of outstanding bond issues), Bloomberg (book value of total debt[9] and market capitalization), and DataStream (price of equity and dividend rate). In addition, Moody's provides data on the face value of defaulted debt and its price thirty days after default, which help estimate default recovery rates.

The paper also uses Moody's data on default rates over the one- and five-year horizons.[10,11] The cohorts of firms tracked by Moody's for the calculation of a particular default rate are chosen according to the rating of senior unsecured debt. Not surprisingly, only a few AAA, AA, and A firms fail on debt obligations, implying that the default history of these firms carries insufficient information with

[8]When a firm does not have senior unsecured debt, Moody's estimates the rating of such debt via interpolation.

[9]Total debt includes all interest-bearing obligations.

[10]Specifically, the paper uses the Default Risk Service and Credit Risk Calculator databases of Moody's Investors Service. These databases provide information about all bond issues in Moody's rating universe as well as ratings and default data.

[11]Moody's definition of default—i.e., the event in which a creditor incurs an economic loss—is in accordance with all the models considered in this paper. See Moody's Investors Service (1998) for further detail.

respect to changes in their creditworthiness. In addition, Moody's coverage of firms rated C or below is limited and prevents meaningful analysis. In light of this, the paper focuses on firms with a BBB, BB, or B rating.[12] Owing to rating migrations, the size of cohorts changes over time, averaging 517 (BBB-rated firms), 389 (BB-rated firms), and 482 (B-rated firms).

The macroeconomic variables used in this paper are provided by the International Monetary Fund, the Congressional Budget Office, and the Bank for International Settlements. These variables comprise an index of U.S. asset prices,[13] the U.S. GDP gap, the U.S. credit-to-GDP ratio, the term spread in the Treasury rate, and the one-year Treasury rate. The first two variables are deflated by the U.S. Consumer Price Index. The credit-to-GDP ratio and the asset-price index are used as gaps from their respective stochastic trends, which, following Borio and Lowe (2002), are calculated on the basis of data available in real time.[14] The term spread is set equal to the difference between the ten-year and three-month Treasury rates. In turn, the GDP gap is calculated as the difference between the log of real GDP and the log of potential real GDP. Finally, the one-year Treasury rate is used as a proxy for the risk-free rate of return.

The calculation of model-based PDs requires data from different sources, and it is the intersection of the Moody's and Bloomberg data sets that restricts the sample size. The upshot is that the available data allow for obtaining firm-specific PDs at a quarterly frequency, from 1990:Q1 to 2003:Q2. The smallest cross-sections of PDs are at the beginning of the sample: seventeen BBB-, fifteen BB-, and seven B-rated firms. The cross-sections expand over time and attain an average (maximum) size of 78 (140) for BBB-, 80 (127) for BB-, and 67 (166) for B-rated firms. The sample is dominated by nonfinancial

[12]In this paper, BBB refers to a Moody's rating between Baa1 and Baa3, BB to a rating between Ba1 and Ba3, and B to a rating between B1 and B3.

[13]The index is a weighted geometric mean of equity prices, and residential and nonresidential (commercial) property prices. The weights change through time and are based on households' annual net wealth. The data sources are the S&P Corporate 500 (equity prices), the National Council of Real Estate Investment Fiduciaries (commercial property prices), and the U.S. Office of Federal Housing Enterprise Oversight (residential property prices).

[14]Specifically, the date-t value of the trend is calculated via a Hodrick-Prescott (HP) filter, which uses data only up to time t. The parameter of the HP filter is set to 1600.

firms, which constitute 86 percent, 91 percent, and 92 percent of, respectively, the BBB-, BB-, and B-rated firms.[15,16]

5. Calibration Methodology

This section consists of five subsections. The first three subsections divide model parameters into three groups and describe their calibration in detail. The fourth subsection summarizes the calibration results. Finally, the fifth subsection provides justification of the adopted calibration scheme and, in the process, reports and discusses the implications of alternative calibrations.

5.1 Time-Invariant Parameters Common to All Firms

Two parameters of the asset process are fixed both across firms and over time. One of them is the initial value of assets, V_0, which is set to 100 without loss of generality because assets follow a geometric Brownian motion. The second parameter is the risk-free rate, which is set equal to the average one-year Treasury rate over the sample period: $r = 4.7$ percent.

In calibrating the LT and AST models, the applicable tax rate, τ, and the monitoring cost, m, are also held fixed. Following Leland (2004), $\tau = 0.15$. Since m and τ have similar implications for AST and LT PDs, respectively, the maintained assumption is that $m = 0.15$ as well.[17]

The final set of constant parameters are those specific to the exogenous-default models. Following Longstaff and Schwartz (1995),

[15] Financial firms enter Moody's calculation of default rates, and for consistency, such firms are considered for the derivation of theoretical PDs as well. That said, excluding financial firms from the sample leaves the results virtually unchanged.

[16] The reported sample sizes are obtained after cleaning the data in order to exclude leverage ratios, dividend rates, and asset volatilities that do not belong to the interval (0,1). (The calibration of leverage and asset volatility is described in section 5.2.) Such a filter removes a relatively small number of observations and is unlikely to influence the analysis. In addition, firm-quarter observations, which imply a default threshold larger than 90 percent of the assets' initial value, are also filtered out.

[17] When m is lowered to zero, the one-year PDs implied by the AST model increase on average by 0.3, 2, and 5 percentage points for BBB-, BB-, and B-rated firms, respectively.

Huang and Huang (2003), and Collin-Dufresne and Goldstein (2001), the parameters of the second stochastic process in the LS, HH, and CDG models are calibrated as follows: $k_r = 0.226$, $\bar{r} = 4.7$ percent, $\sigma_r = 0.0468$, and $\sigma_{rV} = -0.25$; $k_\lambda = 0.202$, $\sigma_\lambda = 0.031$, $\bar{\lambda} = 0.04165$, and $\sigma_{\lambda V} = -0.35$; $k_l = 0.2$ and $\nu = 0.6$.[18]

5.2 Time-Varying Firm-Specific Parameters

Most of the model parameters are calibrated at the firm level and change over time. This is the case for the coupon rate, c, and the time to maturity, T, which are obtained at the yearly frequency directly from data on bond issues.[19] The maintained assumption is that c and T are representative of all (bank and nonbank) debt. In addition, leverage, l, is calibrated quarterly as the ratio of the book value of total debt to the sum of total debt and market capitalization. Then, following Huang and Huang (2003), the payout ratio is determined as $\delta = l * c + (1 - l) * d$, where d is the dividend rate paid out to the firm's equity holders and is obtained at the quarterly frequency.[20] For its part, the debt principal P is just rescaled leverage: $P = l * V_0$.

Firm-specific values of the asset risk premium and volatility are calculated at the quarterly frequency on the basis of stock market data and the model of Leland (1994). This model, which is a special case of the LT model and incorporates only perpetuities, implies that equity value, E, equals

$$E_t = V_t - \frac{(1-\tau)cP}{r} + \left[\frac{(1-\tau)cP}{r} - V^*\right]\left(\frac{V_t}{V^*}\right)^{-x},$$

[18] Alternatively, $\bar{\lambda}$ and ν could be calibrated to firm-specific data; i.e., to the time average of a firm's asset risk premium and leverage ratio, respectively. Switching to these alternative parameterizations has virtually no effect on the implied PDs.

[19] For each firm-year pair in the sample, c and T equal, respectively, the average coupon rate and time to maturity of the firm's outstanding bond issues. The averages use weights proportional to the face values of the corresponding bonds. The average time to maturity typically declines over time on a firm-by-firm basis and, thus, a calibrated value of T is roughly twice as large as the average time to maturity over the remaining life of the firm's outstanding debt. This is consistent with the maturity structure assumed in the LT model.

[20] Ideally, d would also incorporate sales and repurchases of equity shares. The employed data sources do not, however, provide comprehensive coverage of these variables.

where $$x = \frac{r-\delta}{\sigma^2} - \frac{1}{2} + \left[\left(\frac{r-\delta}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}\right]^{1/2} \quad \text{and}$$

$$\mathrm{V}^* = \frac{(1-\tau)cP}{r}\frac{x}{1+x}.$$

The calibration of asset volatility, σ, relies on this equation and a numerical procedure described in Bharath and Shumway (2004). For each firm-quarter pair, this procedure requires an initial guess for σ, which is set equal to a leverage-adjusted volatility of equity returns over the previous year. Given a value for σ and a time series of E over the previous year, the above equation delivers a time series of asset returns. Then, this time series leads to an update of the value of σ and, thus, to an update of the time series of asset returns. The process continues until the absolute difference between two consecutive values of σ is less than 0.001.

The calibration of the asset risk premium λ incorporates a quarterly estimate of the corresponding equity risk premium, λ^E. This estimate is obtained in two steps. The first step derives a time series of marketwide equity premiums that (i) averages 8 percent over the sample period and (ii) exhibits the quarter-on-quarter growth rates estimated by Tarashev and Tsatsaronis (2006) for the S&P 500 stock market index. For each quarter, the second step employs results of Bhandari (1988) in order to estimate firm-specific equity premiums on the basis of the associated leverage ratios. Each cross-section of equity premiums obtained at the second step is rescaled so that its average matches the corresponding marketwide risk premium obtained at the first step.[21]

Given estimates of λ^E and σ, the estimate of λ relies on the Leland (1994) equation for the firm's equity value. Applying Ito's lemma to that equation delivers[22]

$$\lambda = \delta - r + \left[(\lambda^E + r - d)(1-l) - \frac{\sigma^2}{2}x(1+x)\Omega\right]\frac{1}{1-x\Omega},$$

[21]The transformation is necessitated by the fact that the results of Bhandari (1988) abstract from time variation in the equity market risk premium.

[22]Note that the variability of the equity drift over time is assumed to be absorbed by λ^E.

where $(\lambda^E + r - d)$ is the equity drift, $\Omega = \left((1-\tau)\frac{cP}{rV} - \frac{V^*}{V}\right)\left(\frac{V^*}{V}\right)^x$, and $(1-l)$ proxies for $\frac{E}{\mathrm{V}}$.

5.3 Default Threshold and Dead-Weight Cost of Default

The last two parameters that remain to be set are the default threshold, V^*, and the fraction of assets lost in default, α (or the fixed bankruptcy cost, K, in the AST model). The calibration of these parameters incorporates an estimate of the default recovery rate, ρ, which is defined as the price of debt thirty days after default divided by the associated face value. The estimate of ρ is set to 47 percent,[23] which is the average of the ex post recovery rates in Moody's default database.[24]

Under an endogenous-default model, the values of V^*_{LT} and α (LT model) or V^*_{AST} and K (AST model) are determined *simultaneously*. Specifically, each of these parameter pairs comprises the two unknowns in two equations:

$$\begin{aligned} V^*_{LT,it} &= V^*_{LT}(\alpha_t; T_{it}, c_{it}, P_{it}, \delta_{it}, r, \sigma_{it}, \tau) \text{ and} \\ \rho &= \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{(1-\alpha_t)V^*_{LT,it}}{P_{it}} \text{ for the LT model and} \\ V^*_{AST,it} &= V^*_{AST}(K_t; c_{it}, P_{it}, \delta_{it}, r, \sigma_{it}, m) \text{ and} \\ \rho &= \frac{1}{N_t}\sum_{i=1}^{N_t}\frac{V^*_{AST,it} - K_t}{P_{it}} \text{ for the AST model,} \end{aligned} \tag{8}$$

where t indexes the quarter and i the firm, N_t is the total number of firms in quarter t, and $V^*_{LT}(\bullet;\cdots)$ and $V^*_{AST}(\bullet;\cdots)$ summarize the endogenous nature of the two default thresholds.

---

[23] Alternatively, it is possible to estimate a time series of ρ so that each value in this series is based on default data from the previous several years. Unfortunately, the scarcity of the data that can be used to measure recovery rates leads to noisy time-varying estimates of ρ. Specifically, there is a total of 561 default events that can be used to measure recovery rates—or, on average, 40 defaults per year from 1990 to 2003.

[24] For consistency with the way Moody's calculates default rates, as well as with technical assumptions of the theoretical models under study, the calibrated recovery rate is based only on defaults on senior unsecured debt.

By contrast, the exogenous-default models require that the default cost, α, or the default threshold, V^*, be determined in an ad hoc fashion. In this paper, α is set to 45 percent, which is the average value of this parameter under the LT model. Then, the exogenous-default threshold is allowed to vary quarterly and across firms according to a simplified version of equations (8):

$$\rho = \frac{(1-\alpha)V^*_{M,it}}{P_{it}} = \frac{0.55V^*_{M,it}}{P_{it}}, \tag{9}$$

where M stands for the CDG, the HH, or the LS model. In the CDG model, where the default threshold is time varying, the calibrated value of $V^*_{CDG,it}$ is the *initial* value of the parameter.

5.4 Calibration Summary

Tables 1 and 2 provide alternative perspectives on the calibration of the models. Table 1 lists all the parameters and summarizes features of their calibration. For its part, table 2 reports parameter averages alongside their counterparts calibrated by Huang and Huang (2003) and/or Leland (2004).

The differences between the calibration results of this paper and those in the previous literature are generally small and are due primarily to three factors. First, this paper calibrates separately three rating classes, whereas the earlier papers extract certain parameters from data on the entire spectrum of ratings (table 2, fourth column). Second, previously used parameters are based on data that start in 1970, whereas the data used herein start in 1990. Third, the calibration in this paper incorporates total debt, whereas the literature has relied on total *liabilities*.[25] The effect of this surfaces most clearly in the values of leverage and asset volatility, which are, respectively, lower and higher than their analogues in earlier papers.

[25] In addition to total debt, total liabilities include obligations that do not involve interest payments (e.g., promises for physical deliveries). There are two reasons for choosing to work with total debt as opposed to total liabilities. First, a leverage ratio that is based on total liabilities would lead to an overestimation of coupon payments and payout rates. Second, total debt tracks more closely the bond instruments underlying the calculation of default recovery rates.

Table 1. Model Parameters and Their Calibration[a]

Parameter	Description	Firm Specific Y or N	Time Varying Y (y or q) or N
c	Coupon Rate	Y	Y(y)
T	Time to Maturity	Y	Y(y)
r	Risk-Free Rate of Return	N	N
l	Leverage Ratio	Y	Y(q)
δ	Asset Payout Rate	Y	Y(q)
λ	Asset Risk Premium	Y	Y(q)
σ	Asset Volatility	Y	Y(q)
ρ	Default Recovery Rate	N	N
α or K	Default Cost[b]	N	Y(q) in Endogenous-Default Models N in Exogenous-Default Models
V^*	Default Boundary[b]	Y	Y(q)
τ	Tax Rate	N	N
m	Monitoring Cost	N	N
k_r, σ_r, $\bar{r}$, σ_{rV}, k_λ, σ_λ, $\bar{\lambda}$, $\sigma_{\lambda V}$, k_l, ν	Parameters of the Second Stochastic Process (Exogenous-Default Models)	N	N

Note: Y = yes, N = no, y = yearly, q = quarterly.
[a]Unless stated otherwise, the parameter values are the same across models and are based on assumptions adopted by Huang and Huang (2003) and/or Leland (2004).
[b]The calibrated values of the default boundary and the (dead-weight) default cost depend on the type of the underlying model and on the values of other model parameters. In the exogenous-default models, α is fixed at 45 percent and, given an estimate of leverage, V^* is set to be consistent with an estimate of the default recovery rate. In the endogenous-default models, the values of α (or K) and V^* are determined simultaneously by estimates of the following debt characteristics: coupon rate, time to maturity (LT model only), risk-free rate of return, leverage, asset payout rate, asset volatility, default recovery rate, relevant tax rate (LT model only), and monitoring cost (AST model only).

5.5 Discussion of the Calibration Methodology

The calibration scheme described above incorporates several important choices that either need to be justified or compared with alternatives. This is done in the present subsection.

Table 2. Key Parameters of the Structural Credit-Risk Models

	BBB-Rated Firms[a] (Averages)	BB-Rated Firms[a] (Averages)	B-Rated Firms[a] (Averages)	Rep. Firm in the Literature[b]
Drift ($\lambda + r - \delta$)	7.6%	7.4%	5.6%	6%
Payout Rate (δ)	3.6%	3.8%	5.1%	6%
Volatility (σ)	26%	31%	34%	23%
Leverage (l)	30%	37%	50%	43.3% (BBB) 53.5% (BB) 66% (B)
Coupon Rate (c)	7%	8%	9%	8%
Time to Maturity (T)	10 Years	8 Years	7 Years	10 Years

[a]Based on a calibration to firm-specific time-varying parameters (see section 5).
[b]Parameter values adopted by Huang and Huang (2003) and/or Leland (2004).

One of the important choices lies in the decision to calibrate time-varying parameters even though the models assume that (virtually) all of their parameters stay constant over time. Since key parameters reflect risk premiums, debt-service payments, equity volatility, etc., this assumption of the models should be interpreted as referring to steady-state borrower characteristics that convey a long-term level of risk. Thus, the assumption should not be taken literally when the objective is to evaluate the time path of model forecasts with a short—e.g., one-year—horizon, over which credit risk depends significantly on transitory shocks. This motivates calibrating model parameters to their time-varying short-term estimates. Provided that debt characteristics can be expected to change little over the PD horizon, the adopted calibration procedure will be roughly consistent with the models' "constant parameter" assumption and will allow for intertemporal variability in each firm's credit risk.[26]

Nevertheless, it is natural to ask if the incorporation of time-varying parameters affects adversely the performance of the models.

[26]For the endogenous-default models, it is also necessary to assume that parameter variability is small enough to have a negligible impact on the default threshold. In general, the latter value would depend on both short- and long-term borrower characteristics.

In light of this, the paper also analyzes the implications of an alternative calibration, which sets parameter values to firm-specific time averages.

Another important choice is the calibration of a constant risk-free rate of return, r. The main reason for this choice is that it is inherently difficult to pin down the value of r used by the market at any specific point in time (see Feldhütter and Lando 2004). In addition, a particular calibration of a time-varying risk-free rate—which sets r to the average Treasury rate over the corresponding quarter—does not alter the overall level of model-based PDs but impairs, albeit slightly, the explanatory power of these PDs for intertemporal changes in default rates.[27]

The calibration of the asset risk premium, λ, and volatility, σ, also warrants an explanation. There are two main reasons for basing this calibration on the model of Leland (1994), as opposed to on models whose default-rate forecasts are analyzed in this paper. First, the Leland (1994) model delivers endogenous debt and equity valuation, which is key for relating λ and σ to observable stock market data but is beyond the scope of the exogenous-default HH, LS, and CDG models. Second, the default threshold in the Leland (1994) model is independent of default costs, which is key for operationalizing the entire calibration scheme but is not true in the AST and LT models. In addition to these reasons, keeping λ and σ constant across models allows for an evaluation of how the models process the same information.

As far as the asset risk premium λ is concerned, it may seem reasonable to estimate it—just like asset volatility—on the basis of model-implied asset returns over the previous year. Unfortunately, this alternative approach delivers extremely noisy estimates. To see this, one can start by calculating the mean and the standard deviation in the time series of asset risk premiums estimated for each firm under the alternative approach. The resulting cross-section of means averages 0.05, while the cross-section of standard deviations exhibits a significantly higher average of 0.3. For comparison, the corresponding statistics are 0.07 and 0.014 for the estimates of λ used in this paper and 0.3 and 0.065 for the estimates of σ.

[27]The results of this background exercise are available from the author upon request.

Finally, the calibration of the default threshold and the default dead-weight cost follows Leland (2004). In the context of the AST and LT models, this calibration incorporates the fact that (i) there is no consensus in the literature regarding the value of the default-cost parameter,[28] and (ii) this value can be derived simultaneously with the associated default threshold on the basis of data on recovery rates. An alternative scheme, which ignores information on recovery rates and incorporates an ad hoc value of default costs, is bound to introduce much arbitrariness in endogenous-default thresholds.

In contrast to the LT and AST models, the default dead-weight cost α—or, by equation (9), the default threshold—*must* be set in an ad hoc fashion under the CDG, HH, and LS models. The adopted calibration approach, explained in section 5.3, attains two objectives. First, by endowing these models with the average α obtained under the LT model, the approach avoids a predetermined wedge between the PD predictions of the exogenous-default and endogenous-default models. Second, by keeping the exogenous-default α constant over time, in contrast to its counterpart in the LT model, the calibration reflects the fact that the latter model provides a researcher with more information about the value of α. Importantly, it turns out that an alternative approach that endows the LT, CDG, HH, and LS models with the same time-varying value of α blurs in effect any distinction among the first three models.[29]

6. Model-Implied Expectations of Default Rates

The analysis in this paper draws parallels between model-based PDs and actual default rates. Implicitly, this analysis assumes that the parameter values used here for the calculation of model-based PDs accurately characterize the firms considered by Moody's for the calculation of default rates. Under such an assumption, a model's failure to account for default rates is attributed to the specification of the model.

[28] At one extreme, Andrade and Kaplan (1998) argue that α should not exceed 20 percent; at the other extreme, Leland and Toft (1996) set this parameter to 50 percent.

[29] The same point is made in a slightly different context by Leland (2004).

With this in mind, it is reassuring that model-based expectations of one-year default rates—i.e., cross-sectional averages of firm-specific PDs—are, in general, quite in line with actual one-year default rates. This is seen in figure 1, which reveals that the time series of such expectations tends to match the time series of default rates realized over the following year. In addition, table 3 (panel 1) reveals that—with the exception of the CDG and, to a lesser extent, the AST model—the structural credit-risk models exhibit little bias

Figure 1. Expected One-Year Default Rates (by Model) versus Actual Default Rates

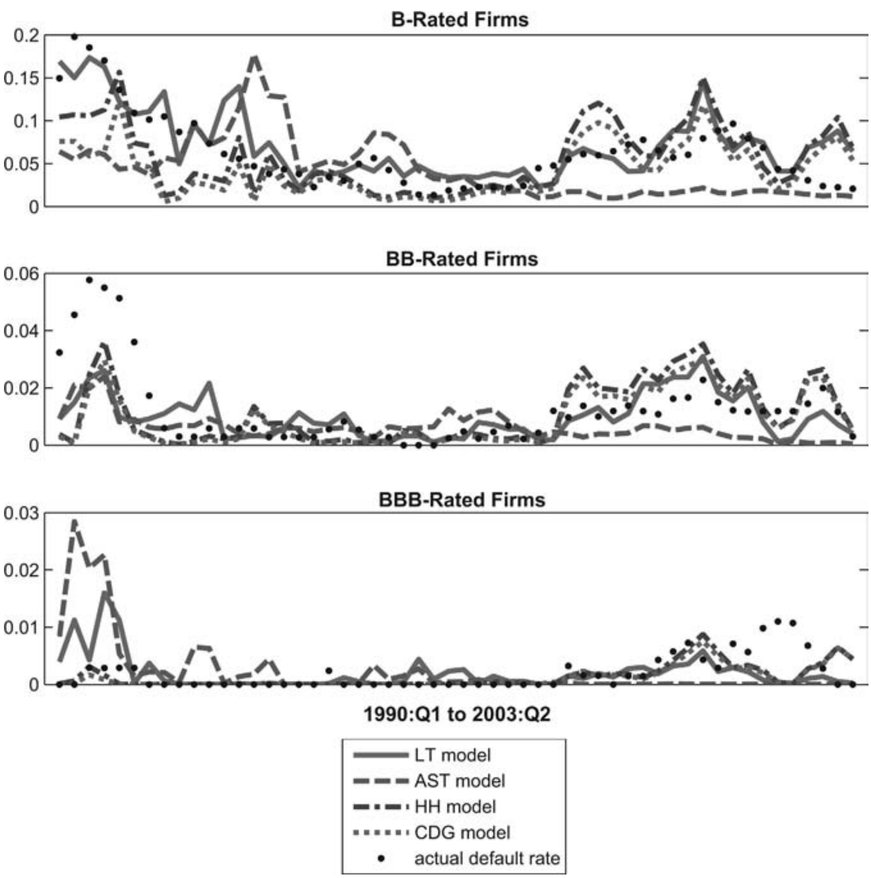


Table 3. Expected and Actual Default Rates (in percentage points)

One-Year Horizon

		B-Rated Firms					BB-Rated Firms					BBB-Rated Firms			
		Mean	**Median**	**Min.**	**Max.**		**Mean**	**Median**	**Min.**	**Max.**		**Mean**	**Median**	**Min.**	**Max.**
Actual Default Rate		6.3	5.5	1.3	19.8		1.20	0.90	0.00	5.80		0.19	0.00	0.00	1.10
***Panel 1:* Expected Default Rate**[a]	**LT**	7.2	5.9	2.2	17.4	**LT**	0.99	0.81	0.11	3.10	**LT**	0.19	0.09	0.00	1.61
	AST	4.3	3.1	1.0	17.8	**AST**	0.62	0.57	0.02	2.40	**AST**	0.23	0.02	0.00	2.85
	HH	5.8	4.4	1.1	15.7	**HH**	1.10	0.52	0.03	3.62	**HH**	0.13	0.01	0.00	0.88
	LS	5.8	4.4	1.0	16.0	**LS**	1.10	0.52	0.03	3.69	**LS**	0.13	0.01	0.00	0.88
	CDG	4.3	3.3	0.6	12.3	**CDG**	0.94	0.45	0.02	2.97	**CDG**	0.12	0.02	0.00	0.76
***Panel 2:* Expected Default Rate, Typical Firm Parameters**[b]	**LT**	6.7	5.8	3.8	15.9	**LT**	0.51	0.48	0.20	1.12	**LT**	0.03	0.01	0.00	0.12
	AST	3.9	3.7	1.0	9.7	**AST**	0.42	0.42	0.07	0.86	**AST**	0.09	0.03	0.00	0.56
	HH	3.4	4.0	0.5	5.8	**HH**	0.35	0.15	0.01	1.05	**HH**	0.03	0.01	0.00	0.31
	LS	3.4	4.0	0.5	5.8	**LS**	0.36	0.16	0.01	1.06	**LS**	0.03	0.01	0.00	0.32
	CDG	2.5	2.8	0.4	4.5	**CDG**	0.34	0.12	0.01	0.97	**CDG**	0.03	0.01	0.00	0.32
***Panel 3:* Expected Default Rate, Representative Firm**[c]	**LT**	1.1	0.9	0.1	4.7	**LT**	0.03	0.00	0.00	0.31	**LT**	0.0001	0.0000	0.0000	0.0017
	AST	0.8	0.2	0.0	8.2	**AST**	0.01	0.00	0.00	0.20	**AST**	0.0006	0.0000	0.0000	0.0097
	HH	2.0	1.0	0.0	9.9	**HH**	0.11	0.00	0.00	0.98	**HH**	0.0009	0.0000	0.0000	0.0141
	LS	2.0	1.0	0.0	10.0	**LS**	0.11	0.00	0.00	0.97	**LS**	0.0009	0.0000	0.0000	0.0137
	CDG	1.5	0.7	0.0	7.0	**CDG**	0.09	0.01	0.00	0.65	**CDG**	0.0012	0.0000	0.0000	0.0177
***Panel 4:* Expected Default Rate, Rep. Firm in the Literature**[d]	**AST**	0.250				**AST**	0.0289				**AST**	0.002910			
	HH	0.055				**HH**	0.0012				**HH**	0.000010			
	LS	0.052				**LS**	0.0011				**LS**	0.000010			
	CDG	0.032				**CDG**	0.0010				**CDG**	0.000015			

[a]Based on a calibration to firm-specific time-varying parameters (see section 5).
[b]Based on average parameters (over time) for each firm.
[c]Based on average parameters (across firms) in each quarter.
[d]Based on parameter values adopted by Huang and Huang (2003) and/or Leland (2004).

vis-à-vis actual default rates.[30] This is evidence in support of the models because the intertemporal mean of the default-rate expectations implied by an accurate model converges to the corresponding mean of the actual default rates as the length of the sample time series expands.

Next, it is important to examine whether the performance of the models is robust to changes in the forecast horizon. In light of this, table 4 (panel 1) reports summary statistics for model-based expectations of five-year default rates alongside the corresponding statistics for actual default rates. The expectations are based on typical firm-specific characteristics, captured by time averages of model parameters. Typical firm characteristics relate to the "steady-state" credit outlook, which is likely to dominate in the determination of long-horizon PDs.

Five-year expected default rates, which would be a key input into risk-management schemes with a long-term focus, tend to match closely actual default rates. The endogenous-default LT and AST models attain a close match across all rating classes considered. In comparison, the performance of the exogenous-default models is poorer: the CDG model consistently underpredicts default rates in the B rating class, while the HH and LS models consistently underpredict in the BBB rating class.

6.1 The Importance of Firm-Specific Calibration

In a recent paper, Leland (2004) finds that—when calibrated to the representative firm—the LT model and (a simplified version of) the LS model consistently and substantially underpredict actual one-year default rates. The data used in the present paper confirm this finding. Employing these data to calibrate the five structural credit-risk models to the representative firm—which is endowed with average borrower characteristics—delivers PDs that are *orders of magnitude* smaller than actual default rates. As revealed by table 3, this result, which is robust across models and rating classes, obtains

[30]Figure 1 and most of the subsequent analysis abstracts from the output of the LS model because, as testified by the results in table 1, this output is virtually indistinguishable from the output of the HH model (see also section 6.3 below).

Table 4. Expected and Actual Default Rates (in percentage points)

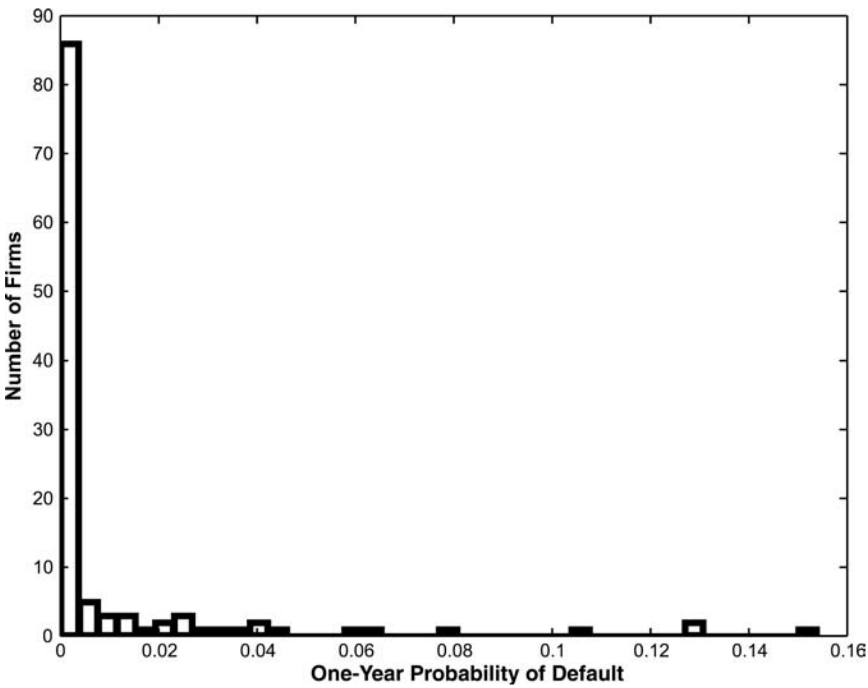
Five-Year Horizon		B-Rated Firms					BB-Rated Firms					BBB-Rated Firms			
		Mean	**Median**	**Min.**	**Max.**		**Mean**	**Median**	**Min.**	**Max.**		**Mean**	**Median**	**Min.**	**Max.**
Actual Default Rate		22.2	21.7	12.5	34.8		6.9	7.0	1.5	14.7		1.22	0.72	0.00	3.70
Panel 1: **Expected Default Rate**[a]	**LT**	26.7	24.3	20.2	34.2	**LT**	6.4	5.3	3.9	11.2	**LT**	0.93	0.61	0.29	1.63
	AST	22.6	23.3	17.4	29.9	**AST**	6.3	5.7	4.5	10.1	**AST**	1.29	0.81	0.50	4.05
	HH	18.6	16.5	11.5	34.5	**HH**	5.3	4.0	2.2	11.0	**HH**	0.65	0.47	0.07	1.82
	LS	17.6	14.8	10.7	33.1	**LS**	5.3	3.9	2.1	10.8	**LS**	0.69	0.52	0.09	1.83
	CDG	13.1	11.5	6.3	24.7	**CDG**	5.3	4.4	1.8	10.2	**CDG**	0.94	0.74	0.09	2.80
Panel 2: **Expected Default Rate, Representative Firm**[b]	**LT**	21.4	20.4	14.1	27.5	**LT**	3.7	2.6	1.8	8.0	**LT**	0.26	0.24	0.07	0.67
	AST	19.0	19.0	14.4	26.1	**AST**	4.3	3.8	2.4	7.7	**AST**	0.57	0.39	0.24	1.53
	HH	18.3	16.4	8.4	37.5	**HH**	3.3	1.9	0.9	8.6	**HH**	0.16	0.08	0.02	0.69
	LS	17.1	14.6	7.4	36.7	**LS**	3.2	1.9	0.9	8.1	**LS**	0.18	0.10	0.03	0.73
	CDG	9.1	6.0	2.5	25.7	**CDG**	2.7	1.8	0.7	6.5	**CDG**	0.23	0.10	0.04	1.01
Panel 3: **Expected Default Rate, Rep. Firm in the Literature**[c]	**LT**	8.7				**LT**	3.2				**LT**	1.00			
	AST	12.4				**AST**	6.9				**AST**	3.77			
	HH	9.7				**HH**	3.9				**HH**	1.30			
	LS	7.1				**LS**	2.7				**LS**	0.85			
	CDG	2.6				**CDG**	1.4				**CDG**	0.73			

[a] Based on average parameters (over time) for each firm.
[b] Based on average parameters (across firms) in each quarter.
[c] Based on parameter values adopted by Huang and Huang (2003) and/or Leland (2004).

irrespective of whether one allows the characteristics of the representative firm to change from quarter to quarter (panel 3) or considers a time-invariant representative firm, as done in the extant literature (panel 4).

As conjectured by Leland (2004), a calibration to the representative firm biases substantially downward model-based forecasts of one-year default rates. Figure 2 illustrates this point by focusing on a particular quarter—1998:Q4—and portraying the distribution of PDs implied by the LT model under the firm-specific calibration. This distribution has a pronounced right tail, implying that the mean (1.1 percent) is substantially larger than the median (0.0393 percent). The former statistic equals the expected default rate, implied by the LT model on the basis of the firm-specific

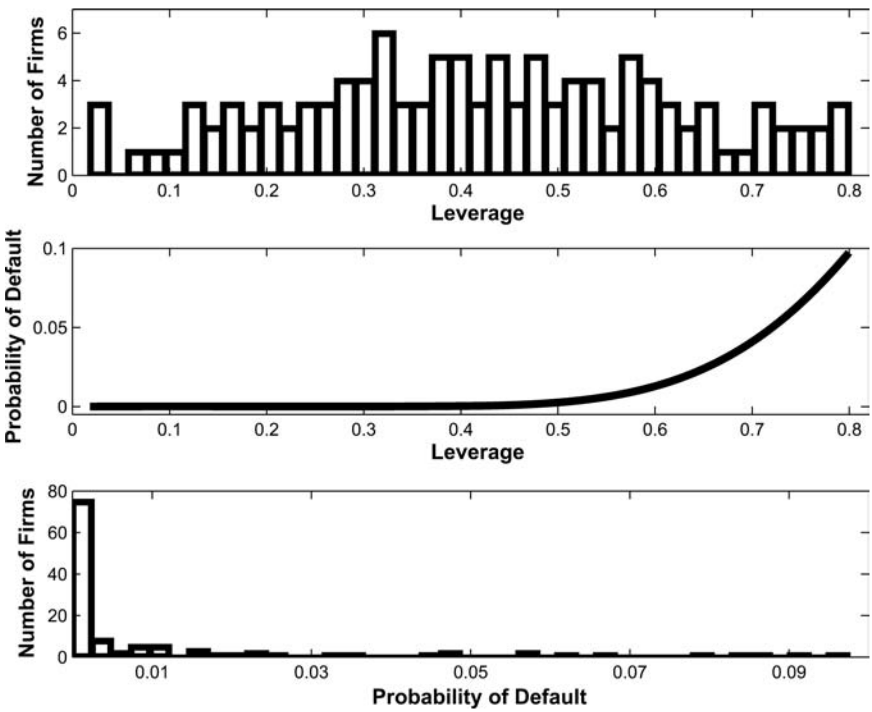
Figure 2. PDs Implied by the LT Model for Firms Rated BB in 1998:Q4



calibration, while the latter statistic is extremely close to the PD of the average (i.e., representative) firm (0.0395 percent).[31]

The strong dependence of expected default rates on whether the underlying model is calibrated to firm-specific parameters or to the representative firm is a particular manifestation of the so-called Jensen inequality effect. The Jensen inequality effect, which arises because of the convex mapping from firm characteristics into model-based PDs, can be best appreciated if such a PD is considered as a function of a single characteristic. Figure 3 illustrates such

Figure 3. The Jensen Inequality Effect (LT Model, BB-Rated Firms, 1998:Q4)



[31]The close match between the median of firm-specific PDs and the PD of the representative firm reveals also that firm-specific parameters (i) have an approximately symmetric distribution in the cross-section and (ii) exhibit a generally monotonic impact on model-based PDs.

a hypothetical scenario by focusing on the leverage ratios of the firms behind the plot in figure 2. In particular, figure 3 reveals that, ceteris paribus, the LT model translates a virtually symmetric distribution of leverage ratios (with a skewness of 0.015) into a PD distribution that has a long right tail (i.e., a skewness of 2.6). It is this tail that drives the large positive difference between the average PD and the PD of the average (representative) firm.

Likewise, the five-year PDs of the representative borrower are invariably lower than the expected default rates based on firm-specific PDs (see table 4, panels 2 and 3). However, again in line with the findings of Leland (2004), the resulting negative bias in representative-borrower PDs is smaller than at the one-year horizon.

6.2 The Impact of Time-Varying Parameters

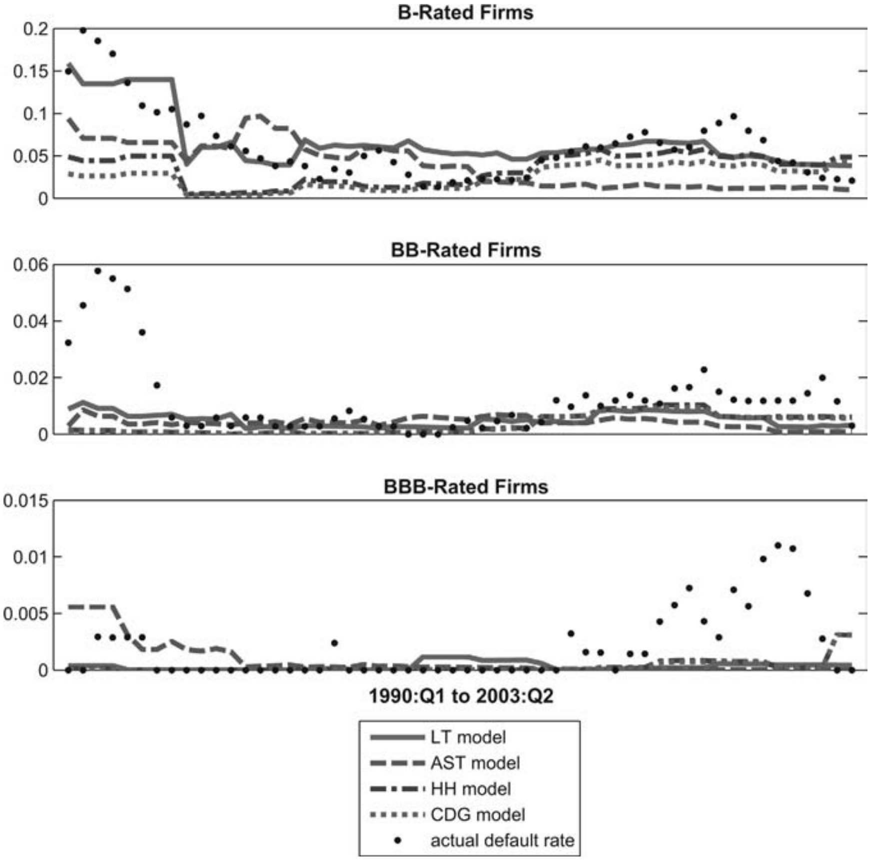
In order to see how the calibration of time-varying parameters affects the performance of the models, an alternative set of one-year expected default rates are calculated on the basis of typical (or average) firm-specific characteristics.[32] As revealed by table 3 (panel 2) and figure 4, these expected default rates are (i) generally lower and (ii) exhibit less variation over time than the expected default rates obtained under time-varying parameters. The first feature reveals that the switch from time-varying to typical parameters reduces the dispersion of borrower characteristics in the cross-section, which dampens the Jensen inequality effect discussed above and, as a result, depresses model-based PDs. The second feature is due to the fact that changes in the alternative expected default rates abstract from changes in credit risk at the firm level and are driven solely by exits from and entries into rating classes.

6.3 Comparison across Models

As illustrated by figure 1, default-rate expectations differ considerably across models. This section relates some of these differences to

[32] Recall that all model-based five-year PDs incorporate time-invariant parameters.

Figure 4. Expected One-Year Default Rates (by Model, Constant Parameters) versus Actual Default Rates



the specification of the models and the way in which specification differences affect calibration.

When considering the endogenous-default LT and AST models, it is necessary to note that differences in their default forecasts could stem solely from differences in the implied default thresholds, V^*_{LT} and V^*_{AST}. As outlined in section 5.3, V^*_{LT} and V^*_{AST} are calibrated together with the associated default-cost parameters, α and K, to match a time-invariant recovery rate. Background analysis (available upon request) unveils two stylized facts about this calibration. First, model-implied recovery rates decrease in α (or K).[33] Second, from the remaining parameters, only changes in asset volatility σ have a first-order impact on model-implied recovery rates. These two stylized facts imply that, given a target recovery rate, a rise in σ—which puts downward pressure on model-implied recovery rates by reducing V^*_{LT} (V^*_{AST})—necessitates a downward revision in α (K), leading to a further reduction of V^*_{LT} (V^*_{AST}).[34] Formally, denoting an endogenous-default model by M, $\frac{\partial V^*_M}{\partial \sigma} < 0$. In conjunction with (2), this implies that the impact of asset volatility on model-based PDs, $\frac{dPD_M}{d\sigma} = \frac{\partial PD_M}{\partial V^*_M}\frac{\partial V^*_M}{\partial \sigma} + \frac{\partial PD_M}{\partial \sigma}$, could be of either sign.

The impact of asset volatility on endogenous-default thresholds drives the wedge between PDs implied, respectively, by the LT and AST models. This impact is much stronger in the context of the latter model and leads to AST PDs that are *negatively* correlated with asset volatility. This negative correlation underpins, e.g., the AST model's underprediction of default rates in the second half of the sample (when σ is on the rise) and the overprediction of default rates in the BBB rating class at the beginning of the sample (when σ is relatively lower). By contrast, changes in σ have a sufficiently small (negative) impact on V^*_{LT} so that LT PDs correlate positively with asset volatility.

At the same time, by equation (9), the calibrated values of exogenous-default thresholds are insensitive to changes in asset volatility. This drives a wedge between the predictions of the endogenous-default and the exogenous-default (i.e., LS, HH, and CDG) models. In addition, in light of the discussion in the previous two paragraphs, it is not surprising to find that the differences between PDs based on an endogenous-default model and PDs based on an exogenous-default model are correlated negatively with asset volatility.

Finally, a comparison across the three exogenous-default models reveals remarkable similarities. As reported in table 3, the expected default rates delivered by the HH and LS models are virtually indistinguishable. By extension, this implies that the two models'

[33] Note that this stylized fact does not have to hold a priori because the endogenous V^*_{LT} and V^*_{AST} increase, respectively, in α and K, which puts upward pressure on model-implied recovery rates.

[34] To see this, recall expressions (6)–(8).

second stochastic processes—which are the only sources of potential differences—have a negligible impact on implied PDs. Thus, in what is to follow, analysis of the HH model's predictions applies equally to the LS model. Turning to the CDG model, the expected default rates it implies correlate strongly (with a correlation coefficient of 0.98 or higher) with their counterparts obtained under the HH or LS models. That said, the CDG model specifies strong mean reversion of the ratio V^*_{CDG}/V to a level above unity, which depresses CDG PDs relative to HH and LS PDs.[35]

7. Statistical Significance of Model-Based Forecast Errors[36]

Focusing on the one-year horizon, this section examines whether actual default rates support the probability distribution of default rates implied by model-based firm-specific PDs. Unfortunately, a key determinant of such a distribution, i.e., the interdependence of defaults across firms, cannot be estimated on the basis of the available data and has to be calibrated to some rule-of-thumb value.

Concretely, the probability distribution of default rates is constructed as follows. First, for a particular model and quarter in the sample, it is assumed that there are firm-specific standard normal variables that generate defaults—by falling below firm-specific thresholds—with *marginal* probabilities equal to the corresponding model-based PDs. Second, in line with Bohn, Arora, and Korablev (2005), the correlation among these random variables is assumed to be 20 percent.[37] This allows for generating

[35]As the mean-reversion parameter, k_l, decreases toward zero, CDG PDs converge to HH and LS PDs.

[36]This subsection builds on the analysis in Kurbat and Korablev (2002) and the related numerical procedure outlined in Bohn, Arora, and Korablev (2005).

[37]There are two reasons for implementing this correlation coefficient. First, Kurbat and Korablev (2002) report that a homogeneous median correlation works well as a proxy for a heterogeneous correlation structure. Motivated by this finding, Bohn, Arora, and Korablev (2005) examine proprietary data and settle on correlations of 18 percent and 19 percent in deriving the probability distribution of default rates. Second, robustness checks reveal that changing the homogeneous correlations of the default generators between 5 percent and 25 percent, which is the range invoked in the literature on portfolio credit risk, has little material impact on the main results.

interdependence across default events while circumventing the use of asset return correlations, which could be incorporated into the models but cannot be estimated from the available data. On the basis of this structure, a large number of draws (concretely, 100,000) from the joint distribution of the default generators delivers the probability distribution of default rates implied by the model in focus.

This exercise reveals that the model-implied probability distributions of default rates are generally consistent with actual default rates, even though there are noteworthy differences across models. Figures 5–8 display quarterly time series of the fifteenth and ninety-fifth percentiles (with the associated 80 percent confidence interval) and the medians of the model-implied probability

Figure 5. Probability Distribution of Default Rates (Based on LT Model) versus Actual Default Rates

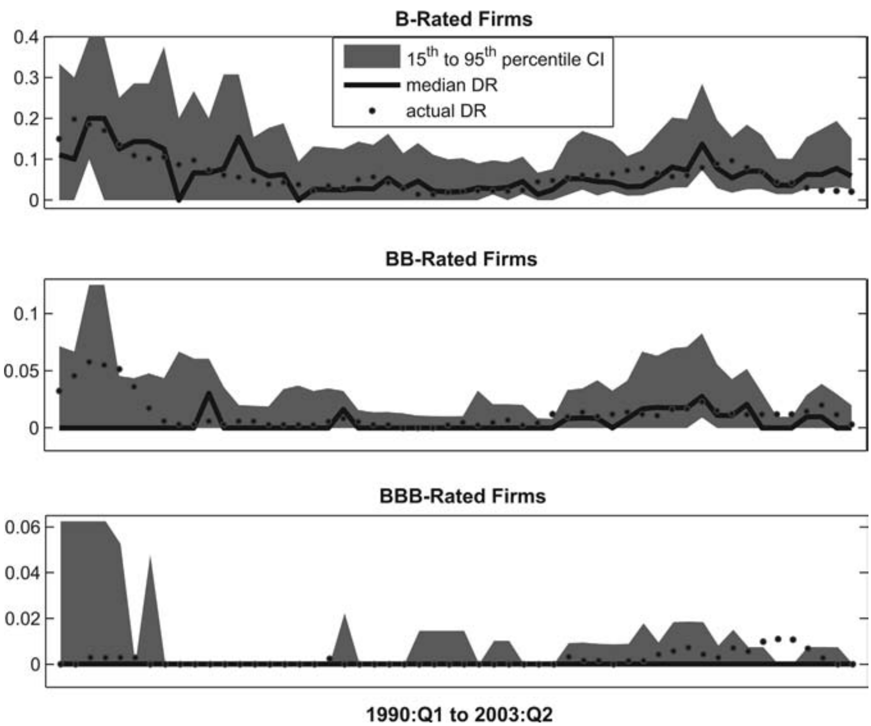
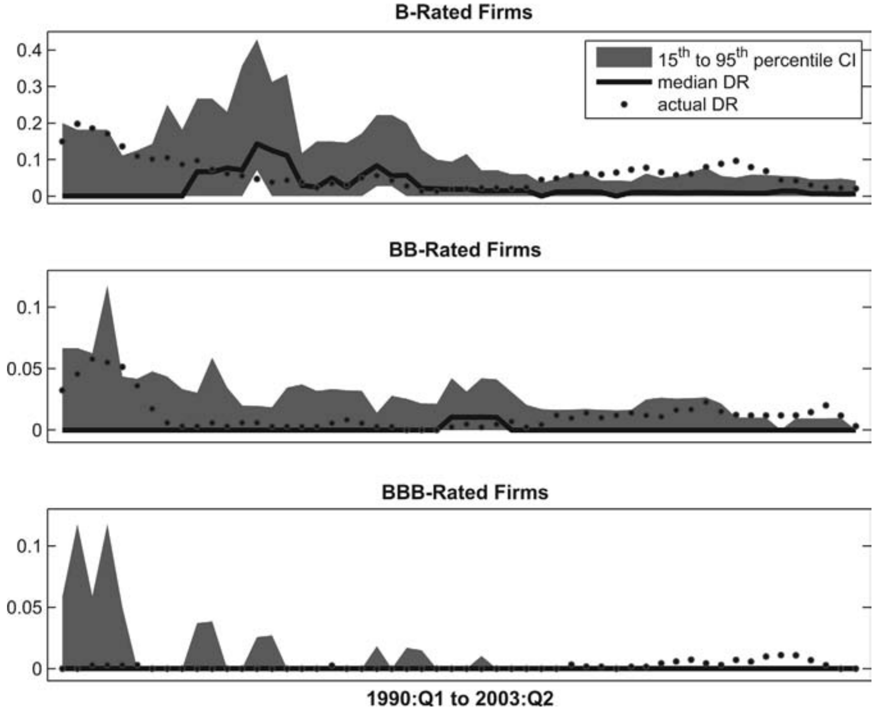


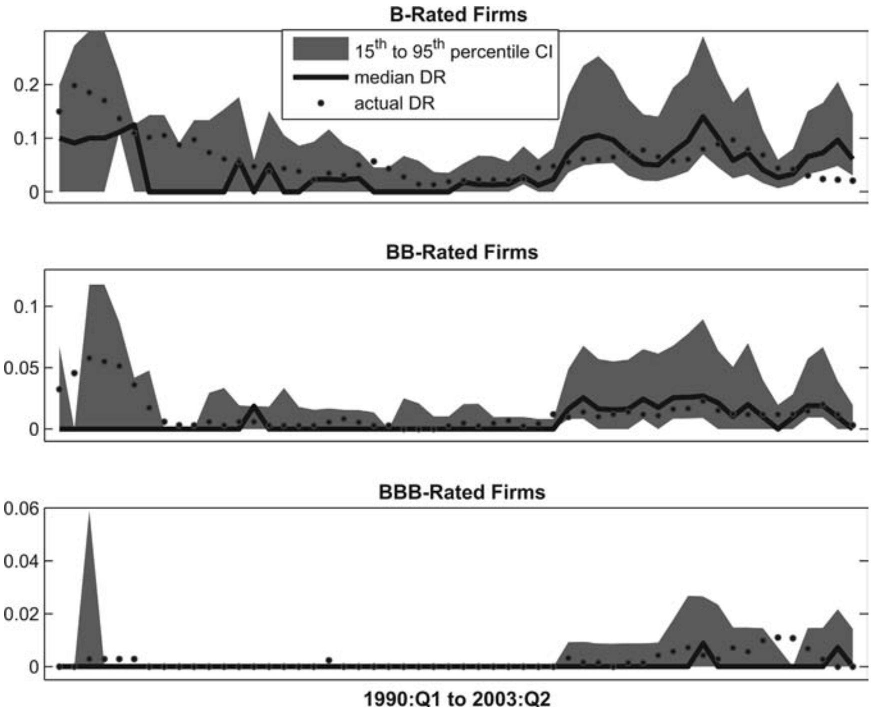
Figure 6. Probability Distribution of Default Rates (Based on AST Model) versus Actual Default Rates



distributions,[38] alongside the corresponding actual default rates. The vast majority of these default rates fall within the 80 percent confidence intervals implied by the LT and HH models. The former model misses, at most, five (out of fifty-four) default rates in a rating class, and the misses of the latter model do not exceed six. By contrast, the performance of the AST and CDG models is much poorer, with the former missing as many as nineteen default rates in the BBB rating class and the latter following behind with ten misses in the BB rating class. One would note that most of the misses are violations of the upper bound of the confidence

[38]Note that the means of these distributions equal the cross-sectional averages of PDs, which are plotted in figure 1.

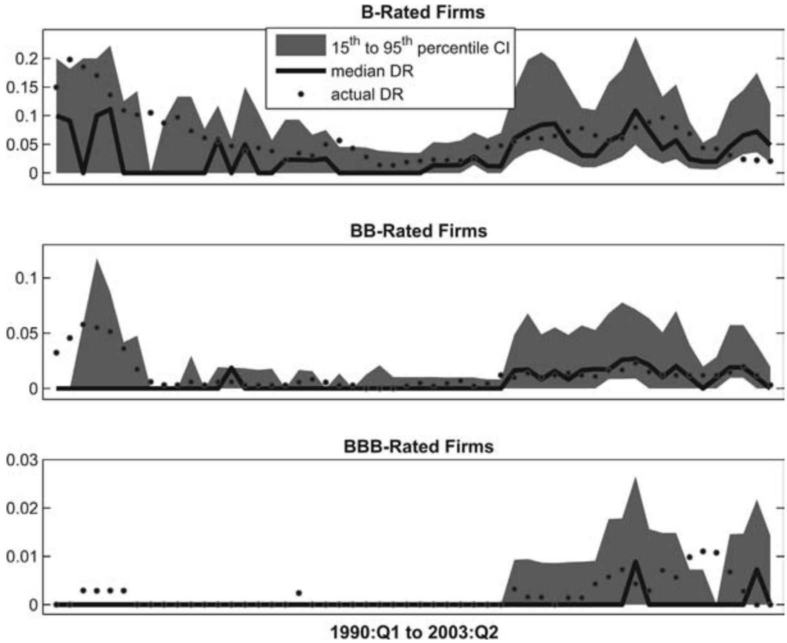
Figure 7. Probability Distribution of Default Rates (Based on HH Model) versus Actual Default Rates



intervals—driven by too-low model-based PDs—whereas even the unconventionally high lower bound (set to the fifteenth percentile) is violated seldom and only in the context of the B rating class. Even in this class, however, the more conventional fifth percentile of the distributions of default rates is invariably zero and is, thus, never violated.[39]

[39] At the five-year horizon, actual default rates almost always belong to the 80 percent confidence intervals implied by model-based PDs. The exceptions, which emerge only in the BBB rating class, occur in five, eight, and four quarters under the AST, HH, and CDG models, respectively, and are always associated with default rates that are higher than the corresponding model-based ninety-fifth percentile.

Figure 8. Probability Distribution of Default Rates (Based on CDG Model) versus Actual Default Rates



8. Economic Significance of Model-Based Forecast Errors

Having discussed the statistical significance of forecast errors, it is also of interest to evaluate their economic significance. This section conducts such an evaluation in a stylized setting in which lenders determine the amount of capital to set aside on the basis of the foundation IRB approach of Basel II and default-rate expectations implied by the firm-specific calibration of a particular credit-risk model. These lenders' capital is compared to two benchmarks: (i) capital based on ex post information about credit risk and (ii) capital based on the credit rating of the exposures.

Consider six hypothetical banks that invest in a bond at time t. The true one-year PD of that bond is assumed to equal the default rate in a particular rating class (BBB, BB, or B), which is realized over the year starting at t. Two of the banks are used as benchmarks. One of them, the "perfect-foresight" bank, calculates its regulatory capital in a particular quarter by setting the PD equal to the one-year default rate realized over the following year and using the foundation IRB approach.[40] The other benchmark bank adopts the standardized approach (SA) of Basel II. The capital requirements of the SA bank are thus based on the rating of the bond issuer and do not change over time. Each one of the remaining four banks relies respectively on the LT, AST, HH, or CDG model. In each quarter, such a bank determines its regulatory capital on the basis of the expected one-year default rate (i.e., the average PD) implied by the adopted model and the foundation IRB approach.

In order to be able to draw sharp conclusions, it is assumed that the optimal capital levels are those calculated by the perfect-foresight bank.[41] At the other extreme is the SA bank. This bank uses an approach that relies on publicly available credit ratings that, in principle, reflect borrowers' *relative* creditworthiness as opposed to probabilities of default.[42] In this context, a credit-risk model provides value added only if the bank relying on it matches the optimal capital level better than the SA bank.[43]

Table 5 reports results of this exercise. The first row of the table contains descriptive statistics of the optimal capital levels, set by the perfect-foresight bank. The remaining rows contain the means of the capital levels calculated by the other five banks, as well as the average and maximum absolute discrepancies between these capital levels and the optimal capital.

[40]Under the foundation IRB approach of Basel II, the capital charge for a particular exposure is determined exclusively by an estimate of the associated one-year PD.

[41]This stylized exercise makes no distinction between regulatory (or required) capital and economic (or optimal) capital.

[42]The officially announced objective of credit-rating agencies is to rate firms according to their long-term financial characteristics. This means that, conditional on the latter characteristics, credit ratings should not vary across the business cycle, even though PDs are largely countercyclical. In addition, credit ratings are supposed to distinguish riskier firms from safer firms and, thus, need to provide an ordinal ranking of firms but need not reflect an absolute measure of credit risk. For further discussion on this topic, refer to Amato and Furfine (2004).

[43]Note that the optimal capital is based on ex post default rates and thus may incorporate more information about PDs than what could possibly be known ex ante. In this sense, model-implied capital cannot be expected to match exactly the optimal one.

Table 5. Capital Measures (averages over time, in percentage points)

	BBB-Rated Firms			**BB-Rated Firms**			**B-Rated Firms**		
	Mean	**Min.**	**Max.**	**Mean**	**Min.**	**Max.**	**Mean**	**Min.**	**Max.**
Optimal Capital	2.7	1.2	7.6	6.7	1.2	12.6	12.5	8.1	19.0
	Mean Cap.	**Mean AE**	**Max. AE**	**Mean Cap.**	**Mean AE**	**Max. AE**	**Mean Cap.**	**Mean AE**	**Max. AE**
LT Bank	2.8	1.6	6.5	6.6	1.9	5.1	13.3	1.5	5.3
AST Bank	2.5	2.5	9.0	5.6	2.8	7.1	10.9	3.3	6.8
HH Bank	2.4	1.0	6.4	6.2	2.2	9.4	12.1	2.2	7.7
CDG Bank	2.3	1.0	6.4	5.8	2.2	9.9	10.9	2.6	9.8
SA Bank	8.0	5.3	6.8	8.0	2.3	6.8	12	2.2	7.0

Note: AE stands for absolute error vis-à-vis the optimal capital level. An SA bank is a bank that adopts the standardized approach of Basel II.

Since the models were seen above to differ in their bias vis-à-vis actual default rates, it is not surprising to find that the models also differ in the success with which they match optimal capital. On average over time, the LT model is quite successful across all of the three rating classes considered. In the BB rating class, e.g., the LT bank experiences a mean discrepancy vis-à-vis the optimal capital of 0.1 percentage points and a mean absolute discrepancy of 1.9 percentage points. These discrepancies amount to 0.9 percent and 17 percent, respectively, of the optimal capital's range, which equals 11.4 percentage points. In comparison, the AST and CDG models perform more poorly, with the banks that adopt these models experiencing mean absolute discrepancies equal to 25 percent and 19 percent, respectively, of the range of optimal capital in the BB rating class. Finally, the HH bank exhibits intermediate performance.

At the same time, again in line with differences between model-based expectations and actual default rates (recall figure 1), all models lead to substantial discrepancies vis-à-vis the level of optimal capital at specific points in time. The LT model, e.g., generates maximum absolute errors that equal 45 percent to 102 percent of the optimal capital ranges, depending on the rating class. Under the other models, these errors tend to be even larger.

Finally, the performance of the SA bank is mixed. The SA bank matches quite closely the optimal capital in the B rating class, where it performs as well as or better than the banks that adopted credit-risk models. However, the performance of the SA bank deteriorates significantly as the credit rating improves to BB and then to BBB. This result reflects the fact that the standardized approach assigns the same regulatory capital (8 percent per unit of exposure) to both BBB- and BB-rated exposures. This results in conservative requirements that overshoot somewhat the optimal regulatory capital even for the riskier category. In the context of a BBB rating, however, the average overshooting is substantial: at 5.3 percentage points, it amounts to 83 percent of the optimal capital's range.

9. Model-Based PDs and Turning Points in the Outlook of Credit Risk

Since the analysis in sections 6–8 examined forecast errors through the prism of intertemporal aggregates, it shed little light on the models' explanatory power with respect to turning points in the outlook of credit risk. Such explanatory power is of natural interest to both market participants and supervisors, not least because, in the context of structural credit-risk models, it can be easily traced to borrower characteristics that drive the evolution of PD forecasts over time.[44]

The evaluation of the explanatory power of credit-risk models relies on time-series regressions (see tables 6–8). Unless stated otherwise, one-year actual default rates are regressed on corresponding model-implied expected default rates that incorporate firm-specific time-varying parameters whose calibration is described in sections 5.1–5.3.[45] In all regressions, period-t expectations forecast the default rate realized over the year *starting* in t. In addition, several control variables allow for dissecting the models' explanatory power: the lagged default rate (realized over the year ending in t) and macroeconomic indicators (realized *prior* to the year ending in t).

The regressions focus on one rating class at a time and include fifty-four quarterly observations (from 1990:Q1 to 2003:Q2). For the B and BB rating classes, the analysis relies on OLS regression results. In the BBB rating class, a significant number of the observed default rates equal 0, which prompts the use of a censored-regression specification based on the Tobit model.[46] In each case, the estimated standard errors account for serial correlation in the residual.[47] Finally,

[44]Background analysis (the results of which are available from the author upon request) reveals that intertemporal movements of model-based expectations of default rates are driven primarily by changes in leverage, asset volatility, and coupon rates and to a much smaller extent by the drift in the asset process or the time to maturity (in the LT model only).

[45]Given that the available data spans less than thirteen years, the regression analysis carried out in this section abstracts from PDs and observed default rates at the five-year horizon.

[46]Twenty-two of the fifty-four default rates in the BBB rating class equal 0. To account for this, observed default rates are assumed to be "censored" at a low positive value (specifically, 0.03 percent). Under the censored-regression model, the reported adjusted R^2 reflects the goodness of fit vis-à-vis the latent dependent variable, which is a linear function of the regressors.

[47]Since the time series are at the quarterly frequency and the credit-risk horizon is one year, the regression errors should be expected to be serially correlated. To account for this, the p-values are based on Newey-West robust covariance matrices (for the regressions pertaining to BB- and B-rated firms) or on Huber-White robust covariance matrices (for the regressions pertaining to BBB-rated firms).

Table 6. BBB-Rated Firms

Dependent Variable: Default Rate								
	Models Only				Adding Macro Explanatory Variables			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Constant	**−0.00**	−0.002	**−0.003**	**−0.003**	−0.002	**−0.002**	0.00	**0.004**
	(0.06)	(0.17)	**(0.04)**	**(0.03)**	(0.19)	**(0.05)**	(0.95)	**(0.01)**
Default Rate (4-Quarter Lag)	**0.52**	**0.53**	0.16	0.32	−0.07	−0.01	**−0.37**	0.29
	(0.07)	**(0.06)**	(0.57)	(0.28)	(0.78)	(0.96)	**(0.09)**	(0.17)
LT PDs	0.37			**0.65**	0.48	−0.023	−0.13	0.03
	(0.16)			**(0.09)**	(0.15)	(0.94)	(0.58)	(0.90)
AST PDs		−0.09		−0.31	−0.20	−0.08	0.03	−0.20
		(0.55)		(0.14)	(0.26)	(0.65)	(0.83)	(0.19)
HH PDs			**1.19**	**0.86**	0.29	0.22	0.19	**0.74**
			(0.01)	**(0.06)**	(0.44)	(0.52)	(0.48)	**(0.02)**
Credit/GDP Gap (12-Quarter Lag)					**0.16**			
					(0.00)			
Asset-Price Gap (18-Quarter lag)						**0.09**		
						(0.00)		
GDP Gap (8-Quarter Lag)							**0.34**	
							(0.00)	
Term Spread (5-Quarter Lag)								**−0.39**
								(0.00)
Adjusted $R^{2(a)}$	0.08	0.04	0.19	0.22	0.50	0.49	0.74	0.66

Note: Tobit regressions, based on one-year actual default rates and cross-sectional averages of one-year theoretical PDs. The sample period is from 1990:Q1 to 2003:Q2 (fifty-four quarterly observations). There are thirty-two censored and twenty-two uncensored observations of the dependent variable. The p-values (in parentheses) are based on Huber-White robust covariance matrices. Entries in bold indicate coefficients that are statistically significant at the 10 percent level.

[a] Reflects the goodness of fit vis-à-vis the latent dependent variable.

Table 7. BB-Rated Firms

Dependent Variable: Default Rate								
	Models Only				**Adding Macro Explanatory Variables**			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Constant	0.00	−0.00	−0.00	**−0.006**	0.00	**−0.006**	**−0.005**	**−0.008**
	(0.75)	(0.81)	(0.81)	**(0.00)**	(0.71)	**(0.00)**	**(0.02)**	**(0.05)**
Default Rate (4-Quarter Lag)	0.41	0.37	**0.49**	**0.34**	0.03	**0.30**	0.30	**0.37**
	(0.13)	(0.11)	**(0.05)**	**(0.04)**	(0.82)	**(0.06)**	(0.10)	**(0.02)**
LT PDs	**0.59**							
	(0.00)							
AST PDs		**1.24**		**1.28**	**1.63**	**1.36**	**1.35**	**1.21**
		(0.03)		**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**	**(0.00)**
HH PDs			**0.53**	**0.54**	0.10	**0.38**	**0.48**	**0.58**
			(0.00)	**(0.00)**	(0.24)	**(0.00)**	**(0.08)**	**(0.00)**
Credit/GDP Gap (17-Quarter Lag)					**0.32**			
					(0.00)			
Asset-Price Gap (16-Quarter Lag)						0.07		
						(0.18)		
GDP Gap (8-Quarter Lag)							0.08	
							(0.32)	
Term Spread (8-Quarter Lag)								0.09
								(0.51)
Adjusted R^2	40%	48%	49%	66%	79%	68%	66%	66%

Note: Based on one-year actual default rates and cross-sectional averages of one-year theoretical PDs. The sample period is from 1990:Q1 to 2003:Q2 (fifty-four quarterly observations). The p-values (in parentheses) are based on Newey-West robust covariance matrices (three lags). Entries in bold indicate coefficients that are statistically significant at the 10 percent level.

Table 8. B-Rated Firms

Dependent Variable: Default Rate								
	Models Only				Adding Macro Explanatory Variables			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Constant	−0.00	**0.02**	0.00	−0.00	0.00	−0.01	0.00	0.02
	(0.87)	**(0.03)**	(0.86)	(0.76)	(0.59)	(0.39)	(0.96)	(0.15)
Default Rate (4-Quarter Lag)	**0.18**	**0.61**	**0.50**	**0.21**	0.14	**0.30**	**0.27**	**0.33**
	(0.06)	**(0.00)**	**(0.00)**	**(0.07)**	(0.30)	**(0.01)**	**(0.02)**	**(0.00)**
LT PDs	**0.72**			**0.62**	**0.61**	**0.61**	**0.60**	**0.51**
	(0.00)			**(0.03)**	**(0.01)**	**(0.03)**	**(0.02)**	**(0.01)**
AST PDs		−0.03		−0.03	0.07	0.09	0.17	**0.21**
		(0.83)		(0.81)	(0.54)	(0.58)	(0.21)	**(0.03)**
HH PDs			**0.47**	0.13	−0.03	0.03	−0.06	0.03
			(0.01)	(0.24)	(0.84)	(0.84)	(0.64)	(0.85)
Credit/GDP Gap (16-Quarter Lag)					0.45			
					(0.16)			
Asset-Price Gap (12-Quarter Lag)						**0.27**		
						(0.02)		
GDP Gap (4-Quarter Lag)							**1.04**	
							(0.00)	
Term Spread (2-Quarter Lag)								**−1.69**
								(0.00)
Adjusted R^2	63%	34%	52%	63%	65%	66%	71%	75%

Note: Based on one-year actual default rates and cross-sectional averages of one-year theoretical PDs. The sample period is from 1990:Q1 to 2003:Q2 (fifty-four quarterly observations). The p-values (in parentheses) are based on Newey-West robust covariance matrices (three lags). Entries in bold indicate coefficients that are statistically significant at the 10 percent level.

in light of section 6.3, strong multicolinearity across regressors is avoided by focusing on the default-rate expectations implied by only one of the exogenous-default models: the HH model.

9.1 "Models-Only" Regressions

The first four columns in tables 6–8 report the estimated coefficients from regressions that incorporate default-rate expectations implied by the LT, AST, and/or HH models but abstract from macroeconomic indicators. The first three of these regressions focus on one of the models at a time, whereas the fourth one represents a "horse race" among the three models.

The results of the single-model regressions reveal that the structural credit-risk models do exhibit explanatory power vis-à-vis default rates. That said, the size of this explanatory power changes substantially both across models and rating classes. The HH model is the only one that attains statistically significant positive coefficients in all three rating classes. This model accounts for roughly half of the intertemporal variability of default rates in the B and BB rating classes but for only one-fifth of this variability in the BBB rating class. In comparison, the LT model performs somewhat better for B-rated firms, slightly worse for BB-rated firms, and has no explanatory power for BBB-rated firms. The worst performer in the single-model regressions is the AST model, which exhibits significant explanatory power only in the BB-rating class.

Importantly, the different models appear to complement each other in explaining the evolution of default rates over time. This is seen in the context of BBB- and BB-rated firms, where using the expected default rates of several models improves the fit of the regressions.[48] The improvement is greater in the context of the BB rating class, where the AST and HH PDs complement each other in capturing the three phases of credit risk in the sample: (i) the spike in default rates in 1990–91, (ii) their subsequent drop until 1998,

[48]In the BB rating class, the predictions of the LT model are not combined with those of the other two models in order to avoid multicolinearity issues. In particular, the expected default rates delivered by the LT and HH models have a correlation coefficient of 77 percent, which renders parameter estimates quite noisy. Including the LT instead of the HH model in column 4 of table 7 reduces the goodness-of-fit measure to 51 percent.

and (iii) their moderate pickup thereafter. On the one hand, the HH PDs are consistently close to observed default rates but miss their relative levels in phases (i) and (iii); on the other hand, the AST PDs underpredict over most of the sample but exhibit a global peak in the early 1990s, just like observed default rates (recall figure 1).

By contrast, using expected default rates from several models does not improve the fit of the regression in the B rating class. In this rating class, the LT model exhibits strong explanatory power on its own—leading to an adjusted R^2 of 63 percent—and dominates the explanatory power of the other models.

Finally, the above findings are put into perspective by examining how the regression results change when the models are calibrated to the typical (i.e., average) parameters of each firm. As revealed by figure 4, such a calibration dampens the time variability of default-rate expectations, which should be expected to impair the capacity of the models to account for turning points in credit risk. This is confirmed by table 9, which reports goodness-of-fit measures that are lower than their counterparts in tables 6–8 (columns 1–4). The only exception is a slight improvement of the already good fit delivered by the LT model in the B rating class.

9.2 Macroeconomic Control Variables

As seen in figure 1, the model-based expected default rates often miss or erroneously predict important changes in actual default rates. In light of the firm-level focus of the models, this finding suggests that they might be ignoring important marketwide determinants of credit risk. In order to investigate the validity of this hypothesis, this subsection incorporates macroeconomic variables in the regression analysis.

The choice of macro variables is motivated by the extant literature that has identified predictors of turning points in the credit and business cycles. In particular, Borio and Lowe (2002) discover that positive deviations of the credit-to-GDP ratio and real asset prices from their respective trends reflect the buildup of financial market imbalances and forecast well banking-system distress years down the road. To the extent that such distress is associated with a deterioration of banks' lending portfolios, the two financial-side variables would be useful in the context of the present paper, as they would

Table 9. Dependent Variable: Default Rate

	Typical Firm Parameters[a]											
	BBB-Rated Firms[b]				BB-Rated Firms[c]				B-Rated Firms[c]			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Constant	−0.00	−0.00	−0.00	−0.00	−0.00	0.00	0.00	−0.00	−0.01	0.02	−0.00	−0.03
	(0.22)	(0.24)	(0.12)	(0.41)	(0.26)	(0.90)	(0.16)	(0.80)	(0.51)	(0.15)	(0.94)	(0.16)
Default Rate	**0.55**	**0.53**	0.49	0.58	0.32	**0.52**	**0.54**	0.23	0.17	**0.54**	**0.57**	0.16
(4-Quarter Lag)	**(0.06)**	**(0.06)**	(0.17)	(0.11)	(0.23)	**(0.08)**	**(0.07)**	(0.41)	(0.22)	**(0.00)**	**(0.00)**	(0.32)
LT PDs	−1.16			−1.31	**2.38**			**3.13**	**0.91**			**0.57**
	(0.68)			(0.63)	**(0.01)**			**(0.05)**	**(0.00)**			**(0.04)**
AST PDs		−0.43		−0.48		0.99		−0.84		0.22		0.48
		(0.42)		(0.40)		(0.42)		(0.50)		(0.57)		(0.10)
HH PDs			0.27	−0.20			**0.67**	−0.53			**0.71**	0.88
			(0.87)	(0.91)			**(0.07)**	(0.53)			**(0.00)**	(0.11)
Adjusted $R^{2(d)}$	6%	5%	4%	2%	45%	33%	34%	44%	65%	35%	43%	66%

Note: Based on one-year actual default rates and cross-sectional averages of one-year theoretical PDs. The sample period is from 1990:Q1 to 2003:Q2 (fifty-four quarterly observations). Entries in bold indicate coefficients that are statistically significant at the 10 percent level.

[a]Model-based PDs incorporate average parameters (over time) for each firm.

[b]Tobit regressions, in which thirty-two of the dependent-variable observations are censored and twenty-two are uncensored. The p-values (in parentheses) are based on Huber-White robust covariance matrices.

[c]The p-values (in parentheses) are based on Newey-West robust covariance matrices (three lags).

[d]For BBB-rated firms, reflects the goodness of fit vis-á-vis the latent dependent variable.

help predict changes in observed default rates. In addition, in order to account for real-side imbalances that might impact lending criteria, the analysis herein considers the deviation of real GDP from its potential level as a harbinger of future changes in credit risk. Finally, the Treasury term spread, which helps predict changes in real economic activity according to Estrella and Hardouvelis (1991) and Smets and Tsatsaronis (1997), is also incorporated in the regressions. The term spread should help predict default rates insofar as the business cycle affects firms' willingness or capacity to service their debt.

With the above motivation, the asset-price, credit-to-GDP, and GDP gaps and the Treasury term spread are used together with model-based PDs as explanatory variables of default rates.[49] The results, reported in columns 5–8 in tables 6–8, shed light on two related questions: (i) Could the forecast errors of theoretical PDs be attributed to factors related to the business or credit cycles? (ii) Is the explanatory power of theoretical PDs robust to controlling for variables that capture marketwide phenomena?

Focusing on the two financial-side variables, the results suggest an affirmative answer to the first question. The credit-to-GDP gap and/or the real asset-price gap enter the regressions with statistically significant coefficients and improve the fit, especially for the BBB and BB rating classes.[50] In the context of the BBB rating class, e.g., the adjusted R^2 reaches 50 percent, which is 28 percentage points higher than the highest R^2 in the models-only regressions. Moreover, in line with the conclusions of Borio and Lowe (2002), positive shocks to the financial-side variables indicate a buildup of financial vulnerabilities and tend to increase default rates three to five years in the future.[51]

[49] In order to avoid overfitting, each macro explanatory variable is used separately.

[50] The lag of a macro explanatory variable is chosen to minimize the Akaike information criterion. The results are robust to changing the adopted lags by one or two quarters in either direction.

[51] The similarities between the results of the present exercise and the findings of Borio and Lowe (2002) notwithstanding, one should keep in mind that the latter paper builds a nonlinear indicator of banking-system distress. The length of the sample size employed herein limits the analysis to the linear case.

Examination of the real-side macro variables also suggests that the forecast errors of the structural credit-risk models can be attributed, at least partially, to factors related to the business cycle. The GDP gap and the term spread improve substantially the goodness of fit of the regressions in the BBB and B rating classes. For example, the introduction of these variables in the context of BBB-rated firms raises the adjusted R^2 by 52 and 44 percentage points, respectively. Positive shocks to the GDP gap indicate an overheating of the economy, which tends to be associated with an increase of default rates one to two years into the future. In turn, negative shocks to the term spread are associated with higher default rates of BBB- and B-rated firms within the following two years. This is consistent with findings of Estrella and Hardouvelis (1991) and Smets and Tsatsaronis (1997) that a tightening of the term spread heralds recessions up to two years in the future.

As regards the second question posed above, the explanatory power of the structural credit-risk models is largely robust to controlling for macro variables. This is seen both in the B rating class, where the magnitude and significance of the LT coefficient is maintained across all regressions, and in the BB rating class, where the same is true for the AST and, to a lesser extent, HH coefficients. In comparison, the findings are quite different in the BBB rating class. Since the goodness-of-fit measures of the models-only regressions are quite low in this rating class, it is not surprising that the explanatory power of macro variables clearly dominates that of the models.

9.3 Out-of-Sample Analysis

The above regression results convey two messages regarding the in-sample performance of structural credit-risk models. First, the models help account for the intertemporal evolution of default rates. Second, models-only accounts of default rates are substantially improved upon on the basis of macroeconomic variables. In order to investigate the robustness of these two messages, this section considers out-of-sample forecasts that rely on model-based expected default rates and macroeconomic variables available in real time.[52]

Specifically, there are four exercises that (i) perform out-of-sample forecasts of the default rates realized over the last eight quarters of the time period under study (i.e., from 2001:Q3 to 2003:Q2) and (ii) differ in the amount of information incorporated in these forecasts. For the first two exercises, models-only specifications include expected default rates from several models (see tables 6–8, regression 4) and forecast the default rate in quarter t on the basis of regression coefficients estimated from data covering 1990:Q1 to 2001:Q2 (i.e., without updates) or 1990:Q1 to $t-1$ (i.e., with updates). The regression coefficients underpinning the other two exercises rely on data spanning the same time periods and incorporating, in addition to model-based expected default rates, the macro explanatory variable that has led to the best in-sample fit of actual default rates for the rating class in focus.[53] Table 10 reports results of the four out-of-sample exercises alongside corresponding results obtained in sample.

The out-of-sample performance of the models-only regressions is consistent with the degree to which the structural credit-risk models account for default rates in sample. In the context of BB- and B-rated classes, the *with*-updates exercise reveals out-of-sample mean absolute errors (MAEs) that are, respectively, 8 and 25 basis points higher than their in-sample counterparts. This worsening of the fit is small relative to the mean of the actual default rates between 2001:Q3 and 2003:Q2 and, thus, translates into only slight increases of mean average percentage errors (MAPEs). In comparison, the worsening of the fit is larger in the BBB rating class, where the in-sample performance of the models is poorest.[54]

[52] The interpretation of this exercise abstracts from the fact that some of the parameters underpinning model-based expected default rates (i.e., the risk-free rate of return and the recovery rate) are averaged over the entire sample period.

[53] In light of tables 6–8, this macroeconomic variable is the GDP gap (for BBB-rated firms), the gap in the credit-to-GDP ratio (BB-rated firms), and the term spread (B-rated firms).

[54] The MAEs and MAPEs for the BBB rating class incorporate the censoring specification outlined in footnote 46.

Table 10. Out-of-Sample Performance

BBB-Rated Firms						
	Models Only			**Adding a Macro Variable[d]**		
	Without Updates[a]	**With Updates[b]**	***In-Sample*[c]**	**Without Updates[a]**	**With Updates[b]**	***In-Sample*[c]**
MAE	0.0062	0.0074	0.0055	0.0046	0.0035	0.0016
MAPE	105%	125%	92%	78%	59%	26%
BB-Rated Firms						
	Models Only			**Adding a Macro Variable[d]**		
	Without Updates[a]	**With Updates[b]**	***In-Sample*[c]**	**Without Updates[a]**	**With Updates[b]**	***In-Sample*[c]**
MAE	0.0065	0.0057	0.0049	0.0032	0.0033	0.003
MAPE	54%	47%	41%	26%	27%	25%
B-Rated Firms						
	Models Only			**Adding a Macro Variable[d]**		
	Without Updates[a]	**With Updates[b]**	***In-Sample*[c]**	**Without Updates[a]**	**With Updates[b]**	***In-Sample*[c]**
MAE	0.027	0.0255	0.023	0.0108	0.0096	0.0072
MAPE	65%	61%	56%	26%	23%	17%

Note: Summary statistics of the difference between the actual default rates and the forecast or fitted default rates from 2001:Q3 to 2003:Q2. MAE = mean absolute error; MAPE = mean absolute percentage error. The percentage errors are calculated with respect to the mean default rate realized from 2001:Q3 to 2003:Q2.

[a]Based on regression coefficients (regression 4, tables 6–8) that incorporate data from 1990:Q1 to 2001:Q2.

[b]Based on regression coefficients (regression 4, tables 6–8) that incorporate data from 1990:Q1 to t, where t is a quarter from 2001:Q2 to 2003:Q3 and occurs one quarter before the forecast default rate.

[c]Based on regression coefficients (regression 4, tables 6–8) that incorporate the full sample period (from 1990:Q1 to 2003:Q2).

[d]The added macro variable is the best performer according to regressions 5 to 8 in tables 6–8: GDP gap (BBB-rated firms), credit-GDP gap (BB-rated firms), and term spread (B-rated firms).

Again in line with the in-sample results reported in section 9.2, adding a macroeconomic explanatory variable improves substantially the out-of-sample forecasts of default rates. This is seen most clearly in the fact that, irrespective of the rating class, the MAEs and MAPEs of the *without*-updates forecasts incorporating a macro variable are lower than those of the best (i.e., in-sample) forecasts based on models-only regressions. Of course, out-of-sample forecasts that incorporate a macro variable and quarterly updates of the regression parameters tend to perform even better. These with-updates forecasts fit actual default rates of B- and, especially, BB-rated firms virtually as well as the corresponding in-sample forecasts. In comparison, out-of-sample forecasts incorporating a macro variable perform more poorly in the BBB rating class.

10. Conclusion

This paper has used firm-level data in order to evaluate the degree to which five structural credit-risk models account for the level and intertemporal evolution of actual default rates. In contrast to previous analyses, which have addressed similar objectives on the basis of aggregate data, the paper finds that PDs implied by the models tend to match the level of actual default rates. In addition, the models explain a substantial portion of the variability of default rates over time. That said, the model-based forecasts of default rates are substantially improved upon by the introduction of macroeconomic variables, which are not modeled explicitly but reflect the business and credit cycles.

Data limitations have imposed several restrictions on the analysis, which are to be kept in mind when interpreting its conclusions. The mitigation of such limitations would, by itself, lead to useful future contributions to the subject matter of this paper. On the one hand, longer data series, incorporating several credit cycles, would shed further light on the robustness of the models' success in accounting for upturns and downturns in credit risk. On the other hand, larger cross-sections would increase significantly one's confidence in the theoretical forecasts of default rates at specific points in time. For example, larger cross-sections in the historical default data would allow for a reliable estimate of a time-varying recovery rate, which would improve considerably the calibration of the models.

References

Amato, J. D., and C. H. Furfine. 2004. "Are Credit Ratings Procyclical?" *Journal of Banking and Finance* 28 (11): 2641–77.

Anderson, R. W., S. Sundaresan, and P. Tychon. 1996. "Strategic Analysis of Contingent Claims." *European Economic Review* 40 (3–5): 871–81.

Andrade, G., and S. Kaplan. 1998. "How Costly Is Financial (Not Economic) Distress? Evidence from Highly Leveraged Transactions that Become Distressed." *Journal of Finance* 5 (5): 1443–93.

Basel Committee on Banking Supervision. 2004. "International Convergence of Capital Measurement and Capital Standards: A Revised Framework." Bank for International Settlements.

———. 2006. "Studies on Credit Risk Concentration: An Overview of the Issues and a Synopsis of the Results from the Research Task Force Project." BCBS Working Paper No. 15.

Bhandari, L. C. 1988. "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence." *Journal of Finance* 43 (2): 507–28.

Bharath, S. T., and T. Shumway. 2004. "Forecasting Default with the KMV-Merton Model." Working Paper.

Bohn, J., N. Arora, and I. Korablev. 2005. "Power and Level Validation of the EDFTM Credit Measure in the U.S. Market." Moody's KMV White Paper.

Borio, C., and P. Lowe. 2002. "Assessing the Risk of Banking Crises." *Quarterly Review* (Bank for International Settlements) (December): 43–54.

Collin-Dufresne, P., and R. Goldstein. 2001. "Do Credit Spreads Reflect Stationary Leverage Ratios?" *Journal of Finance* 56 (5): 1929–57.

Das, S. R., D. Duffie, N. Kapadia, and L. Saita. 2007. "Common Failings: How Corporate Defaults Are Correlated." *Journal of Finance* 62 (1): 93–117.

Duffie, D., L. Saita, and K. Wang. 2007. "Multi-Period Corporate Default Prediction with Stochastic Covariates." *Journal of Financial Economics* 83 (3): 635–65.

Eom, Y., J. Helwege, and J.-Z. Huang. 2004. "Structural Models of Corporate Bond Pricing: An Empirical Analysis." *The Review of Financial Studies* 17 (2): 499–544.

Estrella, A., and G. A. Hardouvelis. 1991. "The Term Structure as a Predictor of Real Economic Activity." *Journal of Finance* 46 (2): 555–76.

Feldhütter, P., and D. Lando. 2004. "The Credit Risk Component in Corporate Spreads and Swap Spreads." Working Paper.

Hillegeist, S. A., E. K. Keating, D. P. Cram, and K. G. Lundstedt. 2004. "Assessing the Probability of Bankruptcy." *Review of Accounting Studies* 9 (1): 5–34.

Huang, J.-Z., and M. Huang. 2003. "How Much of the Corporate-Treasury Yield Spread Is Due to Credit Risk?" Working Paper.

Kealhofer, S., and M. Kurbat. 2001. "The Default Prediction Power of the Merton Approach, Relative to Debt Ratings and Accounting Variables." Moody's KMV White Paper.

Kurbat, M., and I. Korablev. 2002. "Methodology for Testing the Level of the EDF^{TM} Credit Measure." Moody's KMV Technical Report No. 020729.

Leland, H. E. 1994. "Corporate Debt Value, Bond Covenants, and Optimal Capital Structure." *Journal of Finance* 49 (4): 1213–52.

———. 2004. "Predictions of Default Probabilities in Structural Models of Debt." *Journal of Investment Management* 2 (2): 1–16.

Leland, H. E., and K. B. Toft. 1996. "Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads." *Journal of Finance* 51 (3): 987–1019.

Longstaff, F. A., and E. S. Schwartz. 1995. "A Simple Approach to Valuing Risky Fixed and Floating Rate Debt." *Journal of Finance* 50 (3): 789–820.

Merton, R. C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (2): 449–70.

Moody's Investors Service. 1998. "Historical Default Rates of Corporate Bond Issuers, 1920–1998." Special Comment.

Smets, F., and K. Tsatsaronis. 1997. "Why Does the Yield Curve Predict Economic Activity? Dissecting the Evidence for Germany and the United States." BIS Working Paper No. 49.

Stein, R. M. 2000. "Evidence on the Incompleteness of Merton-Type Structural Models for Default Prediction." Moody's KMV White Paper.

Tarashev, N., and K. Tsatsaronis. 2006. "Risk Premia across Asset Markets: Information from Option Prices." *Quarterly Review* (Bank for International Settlements) (March): 93–103.

van Deventer, D. R., L. Li, and X. Wang. 2006. "Advanced Credit Model Performance Testing to Meet Basel Requirements: How Times Have Changed." In *The Basel Handbook: A Guide for Financial Practitioners*, ed. M. Ong. Risk Books.