Nikola Tarashev

+41 61 280 9213
nikola.tarashev@bis.org

Structural models of default: lessons from firm-level data[1]

Structural credit risk models account for the average level of default rates within rating categories only when calibrated on a firm by firm basis. Nevertheless, firm-specific information matters little when one is interested in forecasting the path of default rates over time. This is because economic factors common to all firms strongly influence the evolution of default predictions.

JEL classification: C520, G100, G300.

Financial industry practitioners as well as regulators are constantly searching for reliable measures of default (or credit) risk, ie the risk of a borrower not fulfilling its debt contract. Such measures are of fundamental importance for the sound management of lender institutions and for the supervisory assessment of their vulnerability. The family of "structural" credit risk models developed in the academic literature evaluates the likelihood of default on the basis of borrower characteristics. This special feature examines the performance of three representative models from that family by comparing the probabilities of default (PDs) they deliver with realised default rates.

While a number of related studies focus exclusively on the "representative" borrower endowed with average characteristics, the analysis here calibrates the models to individual firms. The new approach extracts theoretical predictions that account for the *average level* of default rates and stand in contrast to the PD of the representative borrower, which is biased downwards. However, the difference in the calibration approaches is largely inconsequential if one is interested only in explaining *changes* in default rates over time. In this respect, the explanatory power of the models is mixed under either approach.

The reason for this seemingly puzzling contrast in the relevance of data disaggregation is rather straightforward. Under a calibration at the firm level, the estimate of a default rate increases in the dispersion of borrower characteristics because of the non-linear structure of the models. If one focuses instead on the representative borrower, and thus on aggregate data,

[1] The views expressed in this article are those of the author and do not necessarily reflect those of the BIS.

one ignores any dispersion of characteristics across borrowers and hence underpredicts credit risk. By contrast, when tracking default rates over time, the models rely on the *evolution* of borrower characteristics, which are influenced by common economic factors, related to stock market developments and the cost of borrowing. Since the PD of the representative borrower reflects these common factors, the use of aggregate data does not impair the capacity of the models to forecast changes in credit risk.

The next two sections introduce the three structural credit risk models used in the empirical analysis and sketch their calibration. The subsequent sections evaluate in turn the ability of the models to (i) predict average default rates, (ii) identify future defaulters, and (iii) explain the time path of default rates.

Three representative structural credit risk models

Three structural models differ ...

We use three structural credit risk models developed in Leland and Toft (1996; henceforth, LT), Anderson, Sundaresan and Tychon (1996; AST) and Huang and Huang (2003; HH), respectively. All of these models extend the contingent claims framework of Merton (1974), in which a default occurs when the value of the borrower's assets falls below a particular threshold.

The determination of the threshold (default trigger) value of assets is what differentiates the three models from each other. In this respect, the HH model is closest to Merton's in adopting an exogenous default trigger, which does *not* incorporate the incentives of the borrower but is set to be consistent with an estimate of loss-given-default (LGD).[2, 3] In the AST and LT models the default trigger is endogenous, ie determined strategically by the borrower. The LT model, in which a defaulting firm is surrendered to its lenders, derives a trigger that maximises the equity value of the firm for any asset value. For its part, the AST model allows the borrower to restructure the debt contract and thus adopts a definition of default that comes closer to the one used by credit rating agencies. In such a setting, the borrower may default at a higher asset value than in the LT model in order to renegotiate its contract.

... in the implied default triggers

Theory-implied PDs are quite sensitive to the default trigger value of assets, which are set differently across models. The calibration of the HH model relies heavily on an estimate of the *average* LGD in each cross section in the sample. This limits the dispersion of exogenous default triggers across firms. In addition to matching the same estimates of average LGDs, the endogenous default thresholds in the LT and AST models depend on an array of *borrower-specific* characteristics, such as leverage, coupon rate and asset volatility, and, consequently, vary considerably more in each cross section.

---

2 LGD is the amount of assets lost at default, as a fraction of the face value of debt.

3 Tarashev (2005) finds that the PDs implied by the HH model are extremely similar to those obtained by Longstaff and Schwartz (1995), who also assume an exogenous default trigger.

Data

To compare model-implied PDs to realised default rates, this special feature relies on a data set covering corporate borrowers domiciled in the United States. The data set provides quarterly series of default rates and allows for the construction of a parallel series of firm-specific model-implied PDs, from the first quarter of 1990 to the second quarter of 2003. The horizon of default rates and PDs is one year. For the calculation of default rates, we follow standard practice and group potential defaulters according to their credit rating: BBB, BB or B.[4]

Calibrating the models at the firm level requires the use of several data sources.[5] The overlap of the alternative sources is not perfect, which restricts the size of the cross sections of theoretical PDs. The size increases continuously over time, with the average cross section consisting of 77 BBB-, 77 BB- and 59 B-rated firms. Non-financial firms comprise more than 90% of the sample.

Model-implied PDs and realised default rates

The models match average default rates ...

If a correct model is applied to a random selection of firms in a given credit rating class, the average one-year PD in the cross section is an unbiased estimate of the default rate realised in the same rating class over the following year. This estimate requires firm-level data, whereas an alternative estimate, the PD of the representative (average) borrower, necessitates only aggregate data for the rating class. To examine whether a model is unbiased and whether its bias depends on how disaggregated the data are, we average one-year default rates and their alternative estimates over time (Table 1).

Bias in alternative estimators of default rates[1]

Rating	Default rate[2]	LT model[3]		AST model[3]		HH model[3]	
		Average of firm-specific PDs	PD of representative firm	Average of firm-specific PDs	PD of representative firm	Average of firm-specific PDs	PD of representative firm
B	6.30	6.50	0.90	4.50	0.40	3.80	1.40
BB	1.20	1.40	0.05	1.20	0.01	0.90	0.20
BBB	0.20	0.20	$2*10^{-4}$	0.20	$5*10^{-4}$	0.09	$3*10^{-4}$

[1] In percentage points. The sample period is 1990 Q1–2003 Q2. [2] Fraction of firms that default within one year, by rating class; averages over time. [3] Theoretical one-year PDs; by rating class; averages across firms (when applicable) and time.

Table 1

[4] This data set is described in detail in Tarashev (2005). The paper also derives that the overall performance of the models changes little when the horizon is expanded to five years. However, for horizons longer than one year, the time span of the available data severely limits studies of the intertemporal changes in theoretical predictions.

[5] The data sources used here are Moody's KMV, Bloomberg and Datastream. For further information on the calibration of the structural models, see the box on page 102 and the sources cited therein.

Calibration of structural credit risk models

This box sketches the calibration of the parameters that play important roles in the models. The procedure is described more fully in Tarashev (2005) and closely follows Leland (2004) and Huang and Huang (2003).

Most of the borrower and debt characteristics can be set at the firm level. Specifically, the coupon rate and time to maturity of outstanding debt are obtained directly from the data and reflect averages across the debt instruments of the firm. Leverage is measured by the ratio of book value of total debt to the sum of book value of total debt and market capitalisation. The payout ratio, ie the fraction of assets paid out to debt and equity holders, is set equal to a weighted average of the coupon and dividend rates, with the weights determined by leverage. The asset risk premium and volatility are calibrated to be consistent with the equity risk premium and volatility of the corresponding firm. Except for the coupon rate and time to maturity, which change yearly, the other firm-level parameters are set quarterly.

The default trigger value of assets is different across models. In the "endogenous default" LT and AST models, the value is pinned down on the basis of firm-level characteristics (eg debt principal, coupon rate, leverage, asset payout rate and volatility) and an estimate of LGD, which is assumed constant within each cross section of firms but is allowed to vary from year to year. In the HH model, the exogenous default trigger is set to account for the same estimate of LGD and a value of the debt principal. Calibrated in this way, the LT, AST and HH default triggers change both quarterly and across firms but the variation across firms is considerably smaller for the HH trigger.

Finally, the theoretical PDs analysed here are based on a time-invariant estimate of the risk-free rate of return: namely, the average one-year Treasury rate over the entire sample. Tarashev (2005) finds that, if the risk-free rate is allowed to fluctuate through time, the general level of model-implied PDs changes little and their ability to explain the evolution of default rates worsens slightly. Since the risk-free rate is a macroeconomic variable, common to all firms, its calibration does not influence the models' capacity to differentiate borrowers according to their credit risk.

... only when calibrated at the firm level ...

The results reveal that the bias of a model does depend on the level of data disaggregation. Under all the models considered, the theoretical PDs of the representative firms severely underpredict realised default rates in all the rating classes. This underprediction was first observed by Leland (2004). In contrast, when calibrated at the firm level, the two "endogenous default" models exhibit virtually no bias,[6] whereas the bias of the "exogenous default" HH model is reduced substantially but not eliminated.

The non-linear structure of the models explains the different bias across estimators. A deterioration in a borrower's characteristics has a substantially larger (positive) impact on the theoretical PD than a commensurate improvement in these characteristics (which lowers the PD). As a result, the average of firm-level PDs is raised by any dispersion of borrower characteristics, while the PD of the representative firm abstracts from such dispersion. Likewise, the sustained negative bias of the HH model can be traced to the limited dispersion of the exogenous default trigger across borrowers (see above), which depresses the average PD in each cross section.

Does any single borrower characteristic drive the models' capacity to match the general level of default rates? We calculate borrower-specific PDs using firm-level values for only *one* parameter at a time (leverage – ie the ratio

[6] The only exception to this general conclusion is the AST model's underprediction of the average default rate in the B rating class.

Impact of three borrower-level characteristics on theoretical PDs[1]

Rating	Default rate[2]	LT model[3] Leverage	LT model[3] Equity volatility	LT model[3] Coupon rate	AST model[3] Leverage	AST model[3] Equity volatility	AST model[3] Coupon rate	HH model[3] Leverage	HH model[3] Equity volatility	HH model[3] Coupon rate
B	6.30	5.00	0.80	1.11	3.10	1.10	0.60	2.30	2.40	1.50
BB	1.20	0.90	0.08	0.07	0.40	0.07	0.05	0.50	0.50	0.20
BBB	0.20	0.20	$3*10^{-3}$	$3.7*10^{-4}$	0.10	$7*10^{-3}$	$2*10^{-3}$	0.02	0.04	$4*10^{-4}$

[1] In percentage points. The sample period is 1990 Q1–2003 Q2. [2] Fraction of firms that default within one year, by rating class; averages over time. [3] Theoretical one-year PDs, under the assumption that only the parameter identified in the column heading varies across firms; by rating class; averages across firms and time. Table 2

of debt to assets – equity volatility, *or* the coupon rate) and setting the remaining borrower characteristics equal to their averages in each quarter/rating class pair. The averages of these PDs for different specifications are reported in Table 2. A comparison across parameters indicates that the dispersion of leverage ratios does the most to raise the average firm-specific PDs implied by the AST and LT models. By contrast, no single borrower characteristic dominates the predictions of the HH model.

Model-implied PDs of actual defaulters

As argued above, structural credit risk models account for sample averages of default rates only when calibrated on a firm by firm basis. This prompts the question of whether the models can also identify *specific* future defaulters. The available sample contains too few defaults and, as a result, does not allow one to reach a definitive answer. Nonetheless, some noteworthy empirical regularities stand out.

... and can identify future defaulters

To evaluate the models' success in flagging future defaulters, we calculate quarterly cross sections of firm-specific PDs. For a given quarter, we identify the defaulters in all subsequent periods and record what fraction of these firms are being assigned PDs higher than the median model-based PD in the current cross section.[7] The average of these fractions over time represents the "success" ratio of a given model and is reported in Table 3.

Ability of models to flag future defaulters[1]

Averages over time

Number of future defaulters[2]	9.6		
	LT model	AST model	HH model
Success ratio[3]	0.75	0.69	0.76

[1] The sample period is January 1990–December 2004 for defaults and 1990 Q1–2003 Q4 for model calibration. [2] Number of borrowers that are observed in a particular quarter, from 1990 Q1–2003 Q4, and default in any one of the subsequent quarters. [3] Fraction of future defaulters flagged by the model. Table 3

[7] We consider defaults occurring up to December 2004 but derive theoretical PDs up to the second quarter of 2003. Thus, potential defaulters are tracked for at least 18 months.

Based on this rather crude criterion, the models perform reasonably well: on average, they flag up to three out of four future defaulters. "Misses" are due to low leverage ratios, which is in line with the strong impact of this borrower characteristic on the level of theoretical PDs. All of the defaulters missed by the LT and AST models feature a leverage ratio that is smaller than the median leverage in the corresponding cross section. Similarly, the leverage ratio is low for 90% of the firms that are not flagged by the HH model but default later.

Model-implied PDs and the time path of default rates

When one predicts the time profile of default rates ...

Policymakers are interested not only in the average level of default rates but also in their time profile. In this section, we consider the correlation between predicted and realised default rates in a time series context. More specifically, we regress the default rate realised over a particular year on its one-year lag and on a default prediction delivered by a particular model at the end of the previous year. If a model provides useful information for explaining changes in default rates over time, the PDs it implies should enter the regressions with statistically significant coefficients. Furthermore, the PDs of a truly successful model would incorporate *all* currently available information that is useful for forecasting default rates. Thus, if a model is truly successful, past default rates should not be statistically significant in the regressions.

The predictive power of average firm-specific PDs[1]

Dependent variable: realised default rate

	B-rated firms			BB-rated firms			BBB-rated firms		
Constant	0.02 (0.12)	**0.03** (0.00)	0.02 (0.23)	**0.004** (0.01)	0.003 (0.21)	**0.007** (0.00)	–0.002 (0.11)	–0.001 (0.44)	–0.001 (0.34)
One-year lag of realised default rate	**0.37** (0.06)	**0.45** (0.04)	**0.46** (0.01)	.	**0.40** (0.08)	.	**0.66** (0.02)	**0.63** (0.03)	**0.62** (0.03)
LT PD[2]	0.13 (0.61)	.	.	**0.42** (0.01)	.	.	**0.62** (0.08)	.	.
AST PD[2]	.	–0.26 (0.21)	.	.	0.26 (0.23)	.	.	–0.39 (0.47)	.
HH PD[2]	.	.	0.13 (0.69)	.	.	**0.62** (0.00)	.	.	–0.14 (0.87)
Adjusted R-squared	0.27	0.29	0.26	0.31	0.24	0.23	0.13	0.07	0.07

Note: The regressions in Tables 4–7 are weighted, with the weight increasing with the size of the cross section in the corresponding quarter/rating class pair. In Tables 4 and 5 the lagged dependent variable is included only when its coefficient is statistically significant at the 10% level. In Tables 6 and 7 the lagged dependent variable is included if and only if it appears in the corresponding regression in Table 5. 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). In the BBB rating class, 22 of the 54 realised default rates equal zero. To account for this, the BBB regressions are based on the Tobit model and the default rates are assumed to be "censored" at a low positive value (ie 0.03%). The adjusted R-squared of the BBB regressions reflects the goodness of fit vis-à-vis an estimated uncensored version of the dependent variable, which is a linear function of the regressors.

[1] Regressions based on one-year theoretical PDs and default rates. Fifty-four observations from 1990 Q1–2003 Q2. P-values are in parentheses. Entries in bold indicate coefficients that are statistically significant at the 10% level. [2] Cross-sectional averages of firm-specific PDs.

Table 4

The predictive power of representative firm PDs[1] Dependent variable: realised default rate									
	B-rated firms			BB-rated firms			BBB-rated firms		
Constant	**0.02** (0.01)	**0.03** (0.01)	0.01 (0.31)	**0.008** (0.00)	**0.008** (0.00)	**0.009** (0.00)	–0.003 (0.10)	–0.001 (0.41)	–0.002 (0.28)
One-year lag of realised default rate	**0.34** (0.07)	**0.43** (0.03)	**0.51** (0.00)	.	.	.	**0.73** (0.01)	**0.60** (0.03)	**0.64** (0.03)
LT PD	0.73 (0.15)	.	.	**4.20** (0.03)	.	.	**435.00** (0.03)	.	.
AST PD	.	–0.62 (0.57)	.	.	**45.40** (0.00)	.	.	–202.40 (0.18)	.
HH PD	.	.	**1.02** (0.07)	.	.	**1.73** (0.00)	.	.	52.60 (0.75)
Adjusted R-squared	0.29	0.26	0.36	0.11	0.41	0.31	0.16	0.10	0.07

Note: See note in Table 4.

[1] Regressions based on one-year theoretical PDs and default rates. Fifty-four observations from 1990 Q1–2003 Q2. P-values are in parentheses. Entries in bold indicate coefficients that are statistically significant at the 10% level. Table 5

We first consider the explanatory power of the models on the basis of disaggregated information. Hence, as a predictor of default rates, we use the average of the *firm-specific* PDs in each quarter/rating class pair. For a given rating class, we report three regressions in Table 4: one for each of the three structural models.

The estimates of these regressions reveal that none of the models delivers fully successful forecasts. The information in lagged default rates tends to be clearly superior to that contained in theoretical predictors. More concretely, in all but three cases lagged default rates enter the regressions with statistically significant coefficients while the coefficients of average model-implied PDs are insignificant. Two of the three exceptions are due to the LT model, which contributes to the forecasts of BBB default rates and even renders the lagged dependent variable insignificant within the BB rating class. The third exception is due to the HH model, which exhibits strong explanatory power for BB default rates.[8]

... firm-level data have limited value added ...

Next, we examine whether calibrating the models to the representative borrower would affect their capacity to explain the time path of default rates. To this effect, Table 5 reports the same regressions as above with representative borrower PDs substituting for average PDs across borrowers. This substitution does not affect the goodness-of-fit measures in any systematic way and, in several cases, leads to improved significance of theoretical forecasts.[9]

[8] Tarashev (2005) finds that one model may contain information about default rates that complements the information in another model. As a result, using PDs from different models in the same regression of default rates may substantially improve the goodness-of-fit measure.

[9] The magnitude of the regression coefficients increases substantially when one uses representative borrower PDs instead of average firm-specific PDs. This is so because, as reported in Table 1, the former estimate is orders of magnitude smaller than the latter one.

How do we reconcile the similar forecasting power of average borrower-specific PDs and representative borrower PDs with the earlier finding that only predictors of the former type account for the average level of default rates? One possibility is the existence of market factors that induce individual borrower characteristics to change in the same direction over time (which makes the two alternative theoretical forecasts move in tandem) but affect these characteristics differently at each point in time (which helps to differentiate across firms, with an impact only on average borrower-specific PDs).[10] As candidates, we consider two sets of common factors. One is associated with stock market developments, as reflected in leverage, equity risk premiums and volatility; the other is related to the cost of borrowing, as reflected in coupon rates.[11]

… because of the importance of common factors

To examine the role of the stock market and cost of borrowing factors, respectively, we fix the corresponding parameters at their sample averages and use these for calculating modified PDs of the representative firm (Tables 6 and 7). If a particular factor drives model-implied predictions, then suppressing its time variability would lead to weak explanatory power of the modified PDs for realised default rates.

The effect of calibrating constant stock market variables[1]

Dependent variable: realised default rate

	B-rated firms			BB-rated firms			BBB-rated firms		
Constant	**0.03** (0.03)	**0.03** (0.06)	0.03 (0.11)	**0.004** (0.06)	**0.006** (0.00)	**0.006** (0.07)	**–0.004** (0.03)	–0.001 (0.51)	0.001 (0.35)
One-year lag of realised default rate	**0.49** (0.00)	**0.44** (0.01)	**0.37** (0.08)	.	.	.	**0.60** (0.02)	**0.60** (0.03)	0.30 (0.27)
LT PD[2]	–0.59 (0.57)	.	.	**20.15** (0.00)	.	.	**9,540.60** (0.01)	.	.
AST PD[2]	.	–1.17 (0.82)	.	.	**182.40** (0.05)	.	.	***–7,968.10*** (0.07)	.
HH PD[2]	.		–0.11 (0.69)	.	.	5.90 (0.18)	.	.	***–2,611.20*** (0.00)
Adjusted R-squared	0.27	0.26	0.26	0.11	0.15	0.17	0.25	0.14	0.29

Note: See note in Table 4.

[1] Regressions based on one-year theoretical PDs and default rates. Fifty-four observations from 1990 Q1–2003 Q2. P-values are in parentheses. Entries in bold indicate coefficients that are statistically significant at the 10% level. Italicised entries mark statistically significant coefficients that are of the "wrong" sign. [2] Theoretical PDs of the representative firm when stock market variables (ie leverage, equity premium and volatility) are held constant over time.

Table 6

[10] Tarashev (2005) relates the performance of the models to a variety of directly observable macroeconomic indicators: the Treasury term spread and the deviations from trend of the credit/GDP ratio, an asset price index and real GDP. The paper reaches the conclusion that these variables cannot fully account for the explanatory power of the models.

[11] Admittedly, leverage ratios could respond to credit market conditions as well. The calculation of these ratios, however, uses book value of debt, which is typically stable over time, and market capitalisation, which is a volatile variable.

The effect of calibrating a constant coupon rate[1]

Dependent variable: realised default rate

	B-rated firms			BB-rated firms			BBB-rated firms		
Constant	**0.02** (0.01)	**0.03** (0.00)	0.01 (0.40)	**0.008** (0.00)	**0.008** (0.00)	**0.008** (0.00)	**–0.003** (0.09)	–0.001 (0.41)	–0.002 (0.23)
One-year lag of realised default rate	**0.36** (0.05)	**0.44** (0.04)	**0.55** (0.00)	.	.	.	**0.75** (0.01)	**0.65** (0.03)	**0.66** (0.03)
LT PD[2]	0.74 (0.30)	.	.	**3.38** (0.02)	.	.	**281.10** (0.02)	.	.
AST PD[2]	.	–1.80 (0.25)	.	.	**59.20** (0.02)	.	.	–1,139.20 (0.53)	.
HH PD[2]	.	.	**1.65** (0.03)	.	.	**3.25** (0.00)	.	.	110.20 (0.49)
Adjusted R-squared	0.28	0.30	0.38	0.00	0.01	0.28	0.18	0.07	0.08

Note: See note in Table 4.

[1] Regressions based on one-year theoretical PDs and default rates. Fifty-four observations from 1990 Q1–2003 Q2. P-values are in parentheses. Entries in bold indicate coefficients that are statistically significant at the 10% level. [2] Theoretical PDs of the representative firm when the coupon rate is held constant over time. Table 7

Taken together, the findings reported in Tables 5–7 reveal that market-wide factors do indeed contain useful information about future default rates. When measures of borrower features linked to stock market developments are assumed to be constant, all three models effectively cease to explain the time path of default rates in the B and BB rating classes. Namely, the associated slope coefficients become statistically insignificant or negative and the goodness-of-fit measures often plummet. The picture is similar for BBB-rated firms, where the LT PDs provide the only exception. In general, holding the coupon rate constant through time affects the performance of the models only slightly. It worsens materially, however, the goodness of fit of the regressions that rely on the "endogenous default" models for predicting default rates of BB-rated firms.

Conclusion

This special feature has analysed the capacity of three structural credit risk models to predict default rates. To account for average default rates, these models need to be calibrated at the firm level. However, common factors, reflected in aggregated data, influence strongly the evolution of individual borrower characteristics over time. As a result, the use of firm-level data does not improve the (limited) explanatory power of the models for the time profile of default rates.

The above results are an encouraging step towards understanding the empirical performance of structural credit risk models. The results, however, should be considered with caution because they are based on a small sample of borrowing firms that covers a short time period. Longer data series, incorporating several credit cycles, would put the analysis on firmer foundations and help one to better assess the extent to which the models account for upturns and downturns in economy-wide credit risk. Similarly, larger cross sections would significantly increase confidence in the forecasts of individual defaults and of default rates at different points in time.

References

Anderson, R W, S Sundaresan and P Tychon (1996): "Strategic analysis of contingent claims", *European Economic Review*, vol 40, pp 871–81.

Huang, J and M Huang (2003): *How much of the corporate-treasury yield spread is due to credit risk?*, working paper.

Leland, H (2004): "Predictions of expected default frequencies in structural models of debt", *Journal of Investment Management*, vol 2, no 2, pp 1–16.

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

Longstaff, F and E Schwartz (1995): "Valuing risky debt: a new approach", *Journal of Finance*, vol 50, pp 789–820.

Merton, R C (1974): "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, vol 29, pp 449–70.

Tarashev, N (2005): "An empirical evaluation of structural credit risk models", *BIS Working Papers*, no 179, July.