Practical Usage of Credit Risk Models in Loan Portfolio and Counterparty Exposure Management: An Update

Robert A. Jarrow; Donald R. van Deventer

Johnson Graduate School of Management, Cornell University; Kamakura Corporation

INTRODUCTION
Since the original version of this chapter was published in 1999, the state of the art in credit risk modelling has advanced even more rapidly than we had forecast. Three key developments have caused this acceleration. The first development has been the proliferation of credit-related derivatives such as credit default swaps, first-to-default swaps and collateralised debt instruments. The second related development was the refinement and enhancement of the reduced-form approach for modelling credit risk, which was introduced by Jarrow and Turnbull (1995). This modelling approach is called “reduced-form” because the model’s assumptions are imposed on the prices of a firm’s traded liabilities. This means that the model’s parameters can be deduced from observable securities prices – the reduced form. Duffie and Singleton (1999) and Jarrow (1999, 2001) have published significant extensions of the original Jarrow and Turnbull approach. The third development triggering an acceleration in credit risk modelling has been the New Capital Accords proposed by the Basel Committee on Banking Supervision (see http://www.bis.org for the most recent pronouncements of this committee). The Basel Accords’ contribution to credit risk modelling has been twofold. First, the Accords stress the need to
apply credit risk models to the entire balance sheet, not just to credit-related derivatives such as credit default swaps (CDS). Second, the Accords stress the need for quantitative proof that the models perform well. Performance measurement is critical both across credit risk categories and over time, i.e., through the peaks and valleys of credit cycles.\footnote{1}

The result of these three key developments has been a surge of interest in reduced-form models and increased implementation of these models around the world in financial institutions of all types, with applications to retail clients, private companies, listed companies and sovereigns. A prominent example is the 10 December 2003 announcement by the Federal Deposit Insurance Corporation (FDIC) that it has adopted a reduced-form-model-related Loss Distribution Model after many years of study (see Jarrow et al., 2003). The FDIC announcement is just another confirmation of a wholesale movement away from the older structural models of credit risk such as the original Merton (1974) risky-debt model and more recent extensions such as Shimko, Tejima and van Deventer (1993), among others.

Given the plethora of credit risk models now available, how and why did the FDIC conclude that the reduced-form model offered the most promise? More generally, how should a banker seeking to use credit risk models in credit risk management assess their relative performance and, then, implement them in practice? The answer to this difficult question is the subject of this chapter.

An outline for this chapter is as follows. The next section briefly reviews the various credit risk models available. The following section discusses how one should evaluate these credit risk models. Here, we argue that credit risk models need to be verified by multiple performance tests: their performance in ranking companies in the ordinal ranking of riskiness, the consistency between actual and expected defaults and their performance in hedging exposure to credit risk. The section following illustrates these tests using examples based on reduced-form models and Merton's risky-debt model. The final section concludes the chapter.

**THE CREDIT RISK MODELS**

Selection of the appropriate credit risk model is an important aspect of credit risk management. An inappropriate model contains
model error, and model error introduces risk into the credit risk management process. This model risk is as “real” in terms of profit/loss volatility, as is market, credit, liquidity or operational risk. Perhaps the biggest risk in credit risk modelling is the misspecification or underestimation of the amount of and drivers of correlation in defaults among different borrowers. This misspecification can lead to serious understatements of the amount of credit risk that an institution like the FDIC or a major financial institution might incur.

Historically, the first class of credit risk models formulated was the structural approach. Structural models include Merton (1974) and the extension of this model to random interest rates by Shimko, Tejima and van Deventer (1993). This class of models imposes assumptions on the evolution of the value of the firm’s underlying assets and the firm’s liability structure. The liability structure of the firm, in conjunction with the firm’s asset value fluctuations, determines the occurrence of bankruptcy and the payoffs (recovery rates) in the event of default. See Jones, Mason and Rosenfeld (1984), Risk magazine (September 1998), Jarrow and van Deventer (1998), Jarrow, van Deventer and Wang (2002) and van Deventer and Imai (2003) for a summary of the empirical results relating to this class of models.

Extensions of the structural approach, assuming exogenous recovery rates, include the chapters by Nielsen, Saa-Requejo and Santa-Clara (1993) and Longstaff and Schwartz (1995). The Longstaff and Schwartz chapter contains some empirical results.

The original Merton model, like the Black-Scholes model, assumed constant interest rates. Constant interest rates are inconsistent with market realities, and this assumption is one of the reasons why implementations of the original Merton model have not performed well in empirical tests.

Random interest rates, from both an intuitive and a theoretical perspective, should be an essential feature of any credit risk model. First, the credit-risky instruments most often modelled are fixed-income securities for which, by definition, random interest-rate modelling is essential. Perhaps even more important, interest rates are one of the key macroeconomic factors driving default, which is why random interest rates are central to the FDIC Loss Distribution Model. Indeed, the US taxpayers’ trillion-dollar experience with the
interest-rate-induced failures of many savings and loan institutions in the 1980s and 1990s provides anecdotal evidence for the validity of this point. More anecdotal evidence also comes from the 1997–98 Asian crisis, in which high interest rates (used to defend Asian currencies) triggered record bankruptcies. Random interest rates, in fact, has been a standard assumption underlying all recent models of the credit risk process (see Jarrow et al., 2003).

It was the desire to include random interest rates and the discovery of the Heath, Jarrow, Morton (HJM) (1992) term structure model that led to the reduced-form approach to modelling credit risk. The reason for this causality is that the reduced-form credit risk approach, after a clever transformation, reduces to an HJM-type model. This reduction enables existing computer code and empirical procedures to be directly employed in credit risk modelling.

As mentioned earlier, reduced-form models impose their assumptions directly on the prices of the firm's traded liabilities, primarily its debt, and on the default-free term structure of interest rates. Intuitively, the assumptions on the firm's debt prices relate to the credit spread, which is decomposable into the probability of bankruptcy (per unit time) multiplied by the loss (per promised dollar) in the event of bankruptcy. Exogenous assumptions are imposed on these quantities (the bankruptcy and recovery-rate process) directly. This procedure gives the reduced-form models added flexibility in calibrating the model to match market realities.

The structural approach of Merton (1974) and many of its more recent variants assume that corporate capital structure policy is static, with the liability structure fixed and unchanging. For example, the Merton model (and its extension by Shimko, Tejima and van Deventer) assumes that management puts a debt structure in place and leaves it unchanged, even if the value of corporate assets (and therefore equity) has doubled. This static structure is too simple to realistically capture management behaviour and bankruptcy dynamics. Our experience indicates that management attempts to maintain a more constant debt/equity ratio across time. For example, a real estate entrepreneur would readily confirm that, when a building purchased for US$100 million and financed with an US$80 million loan doubles in value, they would refinance for US$160 million.
The Merton model cannot capture this behaviour, while the reduced-form model can. This is reflected in the significant performance differentials between the two classes of models.

The most common implementation of the reduced-form models uses advanced hazard rate modelling to fit a reduced-form model to historical default data with financial ratios, equity market inputs and macroeconomic factors explaining the probability of default. These default probabilities can then be used seamlessly in enterprise risk-management software for credit risk modelling and to meet the requirements of the New Capital Accords. An example of such an implementation is the 1.1 million observation default database used by Kamakura Corporation in its Kamakura Risk Information Services default probability product and its related Kamakura Risk Manager enterprise-wide software product. Another approach for estimating reduced-form model default probabilities would be to imply these parameters from Treasury and corporate debt prices. This approach has also been implemented in the Kamakura Risk Manager system.

To accommodate the wider bid-offer spreads in thinly traded debt and a “liquidity” premium in credit spreads above and beyond the amount of expected loss, reduced-form models need to explicitly model liquidity risk. A method for handling liquidity risk in the context of reduced-form models can be found in Jarrow (1999, 2001). Jarrow sees liquidity risk as analogous to a convenience yield from holding thinly traded debt – sometimes positive and sometimes negative. As such, the standard methods for modelling commodity-related convenience yields can be applied. With this insight, properly constructed reduced-form models can now be used to extract implied default parameters from all subsets or any subset of the firm’s related market prices: equity, debt, subordinated debt and credit derivatives.

**EVALUATION OF CREDIT RISK MODEL PERFORMANCE**

The performance of models designed to predict a yes or no – zero or one – event, such as bankruptcy, has been the focus of mathematical statistics for more than 50 years. For an excellent summary of statistical procedures for evaluating model performance in this regard, see Hosmer and Lemeshow (2000). These performance tests, however, have only been applied to default probability
modelling in recent years. Van Deventer and Imai (2003) provide an overview of these standard statistical tests for default probability modelling. Related chapters are van Deventer and Outram (2002) and van Deventer and Wang (2003). Jarrow, van Deventer and Wang (2002) and van Deventer and Imai (2003) apply an alternative hedging approach to measuring credit model performance. We discuss each of these alternative testing procedures in turn.

The ROC accuracy ratio
The receiver operating characteristics (ROC) accuracy ratio has been used in electronics and medical statistics for more than 50 years, and its use is a standard part of establishing model performance in accordance with the New Capital Accords of the Basel Committee on Banking Supervision. The ROC accuracy ratio is a standard output of many third-party statistical packages, so any financial institution can perform this “non-proprietary” test as outlined in van Deventer and Imai (2003).

Formally, the ROC accuracy ratio is a measure of the trade-off between the relative accuracy in predicting default and the number of false positives. It can be best described by illustrating its computation using statistics provided by Kamakura Corporation from its Kamakura Risk Information Services (KRIS) default probability service.²

Step 1: Take the entire universe of default probability observations. In the case of the KRIS product, this universe has 1.1 million default probabilities on all listed companies in North America monthly from 1989 to the present.

Step 2: Remove from this universe all of the observations for which “default” occurred in a period. In the case of the KRIS database, more than 1,600 companies “defaulted” over the 1989–2003 time period. So, these 1,600 observations are removed from the 1.1 million gross observations, leaving a database of slightly fewer than 1.1 million “non-defaulting” observations. For example, Enron defaulted in December 2001. This December observation of Enron is not included in the “non-defaulting” database, but all prior observations for Enron would be.
Step 3: Calculate the average percentile ranking of the 1,600 defaulters in the universe of non-defaulters. This number is the ROC accuracy ratio.

A perfect model would have an ROC accuracy ratio of 100, because this means all of the defaulting observations have a default probability greater than the default probabilities of the remaining non-defaulting observations. An ROC accuracy ratio of 50, however, indicates a worthless model because the defaulters are indistinguishable from the median non-defaulting company.

We can illustrate the evaluation of model performance with the ROC accuracy ratios from three models offered by the KRIS default probability service:

- KDP-jc, the Jarrow–Chava reduced-form Kamakura Default Probability based on historical estimation of default probabilities using macroeconomic factors, financial ratios and various inputs from the equity market;
- KDP-ms, the Kamakura implementation of the Merton structural model of risky debt; and
- KDP-jm, the Jarrow–Merton hybrid model, which uses KDP-jc and KDP-ms jointly to predict default.

The ROC accuracy ratios for the three models are as follows:

- 93.62%, KDP-jc Jarrow–Chava reduced-form model;
- 91.83%, KDP-jm Jarrow–Merton hybrid model; and
- 83.42%, KDP-ms Merton structural model.

These results show that the reduced-form KDP-jc model significantly outperforms the Merton model in an ordinal ranking of companies by credit riskiness. The ROC accuracy ratio for KDP-ms is similar to the Merton implementation of Sobehart, Keenan and Stein (2000) and Falkenstein and Boral (2000) when they were at Moody’s Investors Service. The results show that 6.48% of the non-defaulting observations for the KDP-jc model (100 – 93.62) are higher than the average default probability for defaulting observations. The same figure for the Merton model is 16.68 (100 – 83.42). This means that the Merton model has almost three times the number of false positives than the Jarrow–Chava reduced-form model.
Even more important, this is true for any mapping of the Merton default probabilities from their theoretical values to actual defaults that preserves the ordinal ranking of the companies by riskiness.\(^4\) This is because the ROC accuracy ratio does not change if the ordering of the observations is not changed.

In addition to the ROC accuracy ratio, there is another important evaluation method that can be applied in practical usage of credit models. We discuss this approach in the next section.

**Performance of credit models versus naïve models of risk**

Knowing the absolute level of accuracy is extremely useful, but it is just as important to know the relative accuracy of a credit model versus naïve models of credit risk. A typical "naïve" model would be one that uses only one financial ratio or equity market statistic to predict default.

It is well known among industry experts that many financial ratios outperform the Merton model in the ordinal ranking of riskiness. This result, however, seems unknown to many academics and financial services regulators so we present representative results here.\(^5\)

**ROC accuracy ratios for Merton model versus selected naïve models**

- 89.20\%, Relative ranking of the current stock price versus the stock price for the last \(N\) years;
- 88.10\%, Percentile ranking among all company stock prices on that day;
- 87.33\%, One-year excess return on the stock versus equity index;
- 87.37\%, One-month equity volatility;
- 85.33\%, Company size; and
- 83.42\%, KDP-ms, Merton structural model.

This observation that the Merton model underperforms common financial ratios has economic and political implications for the practical use of credit models. Can management of a financial institution or bank regulators approve a model whose performance is inferior to a financial ratio that management and bank regulators would not approve as a legitimate modelling approach on a stand-alone basis?
Of course, the answer is no. This finding is another factor supporting the use of reduced-form models that by construction outperform any na"ive model depending on a single financial ratio.6

**The predictive ROC accuracy ratio**
Another important test in the practical use of credit models is to compare long-term forecasting ability of the competing models. See van Deventer and Wang (2003) for an extensive discussion of this topic.

The standard ROC accuracy ratio implicitly studies the relevant period that the bankruptcy probability estimates. For example, if the study is done using annual default probabilities, then the forecasting period has a length of one year. The virtue of a monthly database is that the forecasting period can be more precisely formulated. In this regard, van Deventer and Wang (2003) define the predictive ROC accuracy ratio as an adjusted ROC accuracy ratio where the "defaulting observation" is the default probability for the defaulter N months prior to default. All observations on the defaulter after that date are removed from the calculation. The average percentile rank in the adjusted universe of non-defaulters is calculated based on the default probabilities for the defaulters N months before default.

For the three previous models, van Deventer and Wang present the predictive ROC Accuracy ratio for time horizons ranging from the standard (i.e., in the month of default) to 36 months prior to default. Their results show that the performance differential for reduced-form models versus the Merton model widens 15–18 months prior to default. More recent research shows that reduced-form models have even better forecasting ability if the relative weighting of the explanatory variables is varied with the forecasting horizon, giving rise to a term structure of default probabilities.

**Consistency of actual and expected defaults**
Van Deventer and Wang (2003) also present a test to measure the performance of models over the credit cycle, an important issue as noted by Allen and Saunders (2003) in BIS working paper 126. Figure 1 shows the actual number of company failures in North America from 1990 to 2003 versus the number of defaults that would be expected under each of the three credit models listed above.
Van Deventer and Wang propose a quantitative measure of the consistency between actual and expected defaults – the adjusted $R^2$ of the regression equation

$$\text{Actual defaults} = a + b \ (\text{Expected defaults}).$$

Using annual time periods, the Jarrow–Chava model scores the best by this measure with an adjusted $R^2$ of 87%. The Merton model ranks second at 79%, and the hybrid model third with 70%. This also establishes that the Jarrow–Chava approach is better at modelling default level changes over the credit cycle.

**The Falkenstein and Boral test**

Falkenstein and Boral (2000), suggested another consistency test between actual and expected defaults. They suggest the following test:

- order the universe of all default probability observations from lowest to highest;
- create $N$ “buckets” of these observations with an equal number in each bucket;
- measure the default probability boundaries that define the low end and high end of each bucket;
measure the actual rate of default in each bucket; and
the actual default rate should lie between the lower and upper boundary of the bucket for most of the N buckets.

Falkenstein and Boral propose this test because the implied "early warning" of a credit model can be artificially "increased" by multiplying the default probabilities by some arbitrary ratio such as five. If the model was properly calibrated before this adjustment, the result of multiplication by five will result in an expected level of defaults that is five times higher than actual defaults, which would be unacceptable both to management and to financial institutions regulators under the New Capital Accords. Recent anecdotal evidence is consistent with reports that different vendors' default probabilities can differ by a factor of five or six.

If there is an accidental or intentional bias in default probabilities, the adjusted $R^2$ test in the prior section will detect date-related or credit-cycle-related biases. If the bias is related to the absolute level of the of the default probabilities, the Falkenstein and Boral test applies.

Figure 2 applies the Falkenstein and Boral test using 100 time buckets for version 3.0 of the KDP-jc Jarrow–Chava reduced-form model.

The grey line defines the upper default probability in each of the 100 buckets. Each bucket has 11,583 observations. The lighter line
defines the lower default probability bound of each bucket. The darker line is the actual frequency of defaults. The graph shows a very high degree of consistency in the KDP-jc modelling of actual versus expected defaults.

Figure 3 displays the same results on a logarithmic scale.

The Falkenstein and Boral test does not work well for the lower default probability buckets due to small sample size issues having to do with the discreteness of a default event. As indicated in Figure 3, it is impossible for the actual default frequency to fall between the upper and lower bounds of the bucket because zero defaults out of 11,583 is below the lower bound and one default out of 11,583 is above the upper bound. Only one of the buckets has an actual number of defaults (two defaults out of 11,583) outside of the 99% confidence interval in the lower default probability ranges. Consequently, this test is most useful for the intermediate buckets.

Figure 4 shows the same results for the Kamakura implementation of the Merton model.

Again, the consistency of actual and expected defaults is very high.

Another test suggested by Falkenstein is to plot the distribution of actual defaults by the percentile of the default probabilities.
The first percentile (the lowest 11,583 of the 1,158,268 default probabilities in the sample) should have the lowest number of defaults and the 100th percentile should have the highest.

Figure 5 shows these test results for the Merton model and KDP-jc. KDP-jc shows a better performance because the actual default experience is more heavily concentrated in the highest default probability percentile buckets. In contrast, the default rate actually declines in the highest percentile buckets for the Merton model.

**Hedging based tests of a credit model**


All credit risk models are based on the option pricing technology underlying the famous Black–Scholes (1973) formula, sometimes called the “risk-neutral” valuation technology. In this framework, derivatives are priced by synthetic replication in complete markets that are arbitrage-free. Pricing by “synthetic replication” means that a derivative’s price is determined by the cost of synthetically constructing the derivative using other traded securities. For example, a call option on a stock is priced in the Black–Scholes model by
determining the cost of constructing the option using a dynamic portfolio of the underlying stock and riskless borrowing.

This means that a model is valid if and only if its implied "hedge" works. The implied "hedge" is the synthetic replication portfolio used in reverse. This insight implies that another valid way to test the outputs of a credit risk model is to test its hedging performance. Hence, a model that fails a hedging test is misspecified. A misspecified model, if used to infer default probabilities, will generate only misspecified estimates.

It could be argued that the implicit default probabilities obtained from a credit risk model could be empirically verified by comparing them to historical default frequencies, as we have done in previous sections. In theory this is correct, but in practice it can be difficult to perform the tests of the previous sections due to small sample size problems (as in the case of the Falkenstein and Boral test). This is because default is a rare event and not enough observations may be available to obtain reliable estimates. Indeed, if one believes that existing firms are samples from a population of similar firms, some of which have defaulted in the past, then statistical methods do apply. However, defaults are so rare that the standard errors of the default likelihoods are often too large and the power of standard statistical procedures too small. The bottom line is that
for most estimation procedures a wide range of reasonable default probability estimates are consistent with the data. Therefore, comparisons of estimates with historic default probabilities may be less informative than conclusions drawn using daily bond price series of the entire universe of defaulters and non-defaulters.\textsuperscript{10}

To illustrate this last point, we perform an empirical investigation of the Merton model to demonstrate the wide range of default probabilities that are consistent with a reasonable specification of the model’s parameters, thereby casting doubt on the reliability of the estimates obtained.

AN EMPIRICAL INVESTIGATION OF THE MERTON MODEL
This section provides an empirical investigation of the Merton model using a unique dataset to determine the implied default probabilities. The unique dataset employed is the First Interstate Bancorp dataset previously used in Jarrow and van Deventer (1998) and reported in its entirety in van Deventer and Imai (2003). This dataset contains weekly quotes on potential new issues of various maturity bonds of First Interstate Bancorp from 3 January 1986 to 20 August 1993. For this illustration, the two-year debt issue is employed. For these data, we show that the range of the implied default probabilities obtained for reasonable specifications of the input parameters is too wide and too time-varying to be verifiable using historical default data. This evidence shows the difficulties of using historical default data to test a credit risk model.

Default probabilities from Merton’s model
Merton’s model of risky debt views equity as a call option on the firm’s value. In this simple model, debt is a discount bond (no coupons) with a fixed maturity. Firm value is represented by a single quantity interpreted as the value of the underlying assets of the firm. The firm defaults at the maturity of the debt if the asset value lies below the promised payment. Using the Black–Scholes technology, an analytic formula for the debt’s value can be easily obtained. In this solution, it is well known that the expected return on the value of the firm’s assets does not appear. In fact, it is this aspect of the Black–Scholes formula that has made it so usable in practice.

From Merton’s risky-debt model, one can infer the implied \textit{pseudo}-default probabilities. These probabilities are not those
revealed by actual default experience, but those needed to do valuation. They are sometimes called martingale or risk-adjusted probabilities because they are the empirical probabilities, after an adjustment for risk, used for valuation purposes. If we are to compare implied default probabilities from Merton’s model with historic default experience, we need to remove this adjustment.

Removal of this adjustment is akin to inserting expected returns back into the valuation procedure. To do this, we need a continuous-time equilibrium model of asset returns consistent with the Merton risky-debt structure. Merton’s (1973) intertemporal capital asset-pricing model provides such a structure. We now show how to make this adjustment.

Let the value of the $i^{th}$ firm’s assets at time $t$ be denoted by $V_i(t)$ with its expected return-per-unit time denoted by $a_i$ and its volatility-per-unit time denoted by $\sigma_i$.

Under the Merton (1974) structure, we have that

$$V_i(t) = V_i(0)e^{\mu t + \sigma_i Z(t)}$$

where $\mu_i = a_i - \sigma_i^2/2$ and $Z(t)$ is a normally distributed random variable with mean 0 and variance $t$.

The evolution of $V_i(t)$ above is under the empirical probabilities. In Merton’s (1973) equilibrium asset-pricing model, when interest rates and the investment opportunity set are constant, the expected return on the $i^{th}$ asset is equal to

$$a_i = r + \frac{\sigma_i \rho_{iM}}{\sigma_M} (a_M - r)$$

where $r$ is the risk-free rate, the subscript $M$ refers to the “market” portfolio or equivalently, the portfolio consisting of all assets of all companies in the economy, and $\rho_{iM}$ denotes the correlation between the return on firm $i$'s asset value and the market portfolio.

Using this equilibrium relationship, the drift term on the $i^{th}$ company's assets can be written as

$$\mu_i = -\frac{1}{2} \sigma_i^2 + (1 - b_i) r + b_i a_M$$

where $b_i = \sigma_i \rho_{iM}/\sigma_M$ is the $i^{th}$ firm's beta.
For expository purposes, we parameterise the expected return on the market \( a_M \) as equal to a constant \( k \) times the risk-free interest rate \( r \), i.e.,

\[
a_M = kr.
\]

This is without loss of generality. Using this relation, we have that

\[
\mu_i = -\frac{1}{2} \sigma_i^2 + (1 - b_i + b_k) r.
\]

In Merton's risky-debt model, the firm defaults at the maturity of the debt if the firm's asset value is below the face value of the debt. We now compute the probability that this event occurs.

Let \( t \) be the maturity of the discount bond and let \( B \) be its face value. Bankruptcy occurs when \( V_i(t) < B \), formally

\[
\text{Probability(default)} = \text{Probability}(V_i(t) < B) = N\left( \frac{\ln(B / V_i(0)) - \mu_i t}{\sigma_i \sqrt{t}} \right)
\]

where \( N(\cdot) \) represents the cumulative normal distribution function.

We see from this expression that default probabilities are determined given the values of the parameters \( (B, V_i(0), \sigma_i, r, \sigma_M, \rho_{IM}, k) \). We discuss the estimation of these parameters in the next section.

**Empirical estimation of the default probabilities**

As mentioned previously, this estimation is based on the First Interstate Bancorp bond data. The time period covered is from 3 January 1986 to 20 August 1993. Over this time period, the bank solicited weekly quotes from various investment banks regarding new issue rates on debt issues of various maturities. The two-year issue is employed in the subsequent analysis (see Jarrow and van Deventer, 1998, for additional details).

To estimate the default probabilities, we need to estimate the seven parameters \( (B, V_i(0), \sigma_i, r, \sigma_M, \rho_{IM}, k) \).

The face value of the firm's debt \( B \) can be estimated using balance sheet data. To do this, we choose \( B \) to be equal to the total value of all of the bank's liabilities, compounded for two years at the average liability cost for First Interstate.
The firm value at time 0, \( V_t(0) \), and the firm's volatility parameter \( \sigma_t \) are both unobservable. This is one of the primary difficulties with using Merton's model and with the structural approach, in general. For this reason, both the market value of First Interstate's assets and the firm's asset volatility were implied from the observable values of First Interstate common stock and First Interstate's credit spread for a two-year straight bond issue. We choose those values that minimised the sum of squared errors of the market price from the theoretical price (see Jarrow and van Deventer, 1998, for details).

The spot rate \( r \) is observable and the volatility of the market portfolio \( \sigma_M \) can be easily estimated from market data. As a proxy for the market portfolio, we used the S&P 500 index. The volatility of the S&P 500 index over the sample period was 15.56%.

Lastly, this leaves the two parameters: the expected return on the market and the correlation between the firm's assets and the market portfolio. As both of these quantities are unobservable and arguably difficult to estimate, we choose them as "free" parameters. That is, we leave these values unspecified and estimate default probabilities for a range of their values. The range of values for the correlations employed is from \(-1\) to 1, and the range of ratios for the market return to the risk-free rate is from \( k = 1 \) to \( k = 5 \).

Finally, in order to compare the derived two-year default probabilities with the observable credit spread, which is quoted on an annual basis, we convert the actual two-year default probability to a discrete annual basis using the following conversion formula

\[
\text{Probability[annual]} = 1 - \sqrt[1]{1 - \text{Probability[two year]}}.
\]

Table 1 provides these annualised default probabilities on 17 November 1989.\(^{14}\)

Surprisingly, one can see the wide range of default probabilities possible as we vary the free parameters. Across the entire table, the annualised default probability varies between 0.06% and 92.68%. For a correlation of 1 (the typical asset), the range in default probabilities for a market return to risk-free rate ratio between \( k = 2 \) and \( k = 3 \) is 9.09% and 2.57%.\(^{15}\) These ranges in default probabilities are quite large and they cast doubts upon our abilities to credibly validate the model using historical default frequencies.
More insight into the imprecision of the Merton's default probability estimates can be obtained by looking at the variability in these estimates across time. For illustrative purposes, we set the correlation between First Interstate assets and the market portfolio to be 1. We look at the two extreme market return to risk-free rate ratios, \( k = 5 \) and \( k = 1 \).

The time series graphs of the annualised default probabilities are contained in the Figures 6 and 7.\(^{16} \)

**Table 1** Merton model default probability

<table>
<thead>
<tr>
<th>Date</th>
<th>17/11/98</th>
<th>Actual credit spread</th>
<th>1.220%</th>
<th>Ratio of expected return on market to risk-free rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation of asset returns with market returns</td>
<td></td>
<td></td>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td>1.00</td>
<td>0.06%</td>
<td>0.49%</td>
<td>2.57%</td>
<td>9.09%</td>
</tr>
<tr>
<td>0.75</td>
<td>0.49%</td>
<td>1.76%</td>
<td>5.06%</td>
<td>11.79%</td>
</tr>
<tr>
<td>0.50</td>
<td>2.57%</td>
<td>5.06%</td>
<td>9.09%</td>
<td>14.99%</td>
</tr>
<tr>
<td>0.25</td>
<td>9.09%</td>
<td>11.79%</td>
<td>14.99%</td>
<td>18.67%</td>
</tr>
<tr>
<td>0.00</td>
<td>22.83%</td>
<td>22.83%</td>
<td>22.83%</td>
<td>22.83%</td>
</tr>
<tr>
<td>-0.25</td>
<td>43.08%</td>
<td>37.64%</td>
<td>32.38%</td>
<td>27.42%</td>
</tr>
<tr>
<td>-0.50</td>
<td>64.78%</td>
<td>54.16%</td>
<td>43.08%</td>
<td>32.38%</td>
</tr>
<tr>
<td>-0.75</td>
<td>82.14%</td>
<td>69.70%</td>
<td>54.16%</td>
<td>37.64%</td>
</tr>
<tr>
<td>-1.00</td>
<td>92.68%</td>
<td>82.14%</td>
<td>64.78%</td>
<td>43.08%</td>
</tr>
</tbody>
</table>

**Figure 6** Annualised Merton model default probabilities for First Interstate data, with the expected return on the market = 5.0 \times\text{ treasury rate}
The line connecting the data is drawn from point to point chronologically. The highest end of the line represents the newest data points.

The primary observation from these graphs is the high degree of instability that these estimates exhibit across time. This is especially true for the larger market return to risk-free rate ratio \( k = 5 \). The more stable estimates for \( k = 1 \), however, appear implausibly high.

To obtain a better sense for the magnitude of the variability in these estimates, we computed the standard deviation in these default probability estimates across time. These numbers are contained in Table 2.

These standard deviations are quite large. To place these numbers in perspective, the standard deviation of the First Interstate Bancorp's two-year credit spread over the sample period is only 0.566%.

In Table 3, we take the ratio of the annualised default probability standard deviations to the standard deviation of First Interstate's two-year credit spread.

These numbers add some perspective to the instability of the estimated default probabilities. Except for the case where the correlation coefficient is 1 and the market-return-to-risk-free-rate ratio \( k = 5 \), the variation in default probabilities is far in excess of the volatility of the underlying credit spread.
Table 2  Standard deviation of Merton default probabilities 1986–93

<table>
<thead>
<tr>
<th>Correlation of asset returns with market returns</th>
<th>5.00</th>
<th>4.00</th>
<th>3.00</th>
<th>2.00</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.48%</td>
<td>1.13%</td>
<td>2.34%</td>
<td>4.10%</td>
<td>6.41%</td>
</tr>
<tr>
<td>0.75</td>
<td>1.13%</td>
<td>1.98%</td>
<td>3.16%</td>
<td>4.61%</td>
<td>6.41%</td>
</tr>
<tr>
<td>0.50</td>
<td>2.34%</td>
<td>3.16%</td>
<td>4.10%</td>
<td>5.17%</td>
<td>6.41%</td>
</tr>
<tr>
<td>0.25</td>
<td>4.10%</td>
<td>4.61%</td>
<td>5.17%</td>
<td>5.76%</td>
<td>6.41%</td>
</tr>
<tr>
<td>0.00</td>
<td>6.41%</td>
<td>6.41%</td>
<td>6.41%</td>
<td>6.41%</td>
<td>6.41%</td>
</tr>
<tr>
<td>−0.25</td>
<td>9.52%</td>
<td>8.68%</td>
<td>7.87%</td>
<td>7.11%</td>
<td>6.41%</td>
</tr>
<tr>
<td>−0.50</td>
<td>12.56%</td>
<td>11.16%</td>
<td>9.52%</td>
<td>7.87%</td>
<td>6.41%</td>
</tr>
<tr>
<td>−0.75</td>
<td>13.84%</td>
<td>13.10%</td>
<td>11.16%</td>
<td>8.68%</td>
<td>6.41%</td>
</tr>
<tr>
<td>−1.00</td>
<td>12.88%</td>
<td>13.84%</td>
<td>12.56%</td>
<td>9.52%</td>
<td>6.41%</td>
</tr>
</tbody>
</table>

Table 3  Ratio of standard deviation of Merton default probabilities to standard deviation of credit spread 1986–93

<table>
<thead>
<tr>
<th>Correlation of asset returns with market returns</th>
<th>5.00</th>
<th>4.00</th>
<th>3.00</th>
<th>2.00</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>84%</td>
<td>200%</td>
<td>413%</td>
<td>723%</td>
<td>1131%</td>
</tr>
<tr>
<td>0.75</td>
<td>200%</td>
<td>350%</td>
<td>557%</td>
<td>815%</td>
<td>1131%</td>
</tr>
<tr>
<td>0.50</td>
<td>413%</td>
<td>557%</td>
<td>723%</td>
<td>912%</td>
<td>1131%</td>
</tr>
<tr>
<td>0.25</td>
<td>723%</td>
<td>815%</td>
<td>912%</td>
<td>1017%</td>
<td>1131%</td>
</tr>
<tr>
<td>0.00</td>
<td>1131%</td>
<td>1131%</td>
<td>1131%</td>
<td>1131%</td>
<td>1131%</td>
</tr>
<tr>
<td>−0.25</td>
<td>1680%</td>
<td>1533%</td>
<td>1390%</td>
<td>1256%</td>
<td>1131%</td>
</tr>
<tr>
<td>−0.50</td>
<td>2218%</td>
<td>1971%</td>
<td>1680%</td>
<td>1390%</td>
<td>1131%</td>
</tr>
<tr>
<td>−0.75</td>
<td>2443%</td>
<td>2312%</td>
<td>1971%</td>
<td>1533%</td>
<td>1131%</td>
</tr>
<tr>
<td>−1.00</td>
<td>2274%</td>
<td>2443%</td>
<td>2218%</td>
<td>1680%</td>
<td>1131%</td>
</tr>
</tbody>
</table>

In summary, given the wide variation in default probabilities for small changes in the input parameters, this illustration supports the statement that comparing estimated default probabilities with historical default frequencies may not provide a very powerful test of a credit risk model.

CONCLUDING COMMENTS

The successful implementation and practical use of a credit model involves critical choices. In these choices, bank management and bank regulators have a fiduciary responsibility to shareholders and depositors to completely "vet" or audit all models used. A blackbox approach to modelling fails this test. All credit models used
should be subjected to critical analysis and review, either publicly or privately.

All models should be tested “out of sample” by the user (or an auditor employed by the user). The richest tests involve historical periods with substantially different market conditions (say the 1979–85 high-interest period in the United States) or tests in other countries with far different market conditions. Japan and the rest of Asia over the last decade, for example, provide a much better testing ground than just using the nearly 20 years of prosperity in the United States.

All models have weaknesses, and it is better to aggressively seek them out than to identify them after a problem has occurred. Perhaps an ideal credit risk management system should utilise multiple credit models to help diversify this model risk. Such a system would allow the user full control of the model audit and performance testing process. This kind of system ends the debate about relative model performance, because it is “agnostic”. Through its usage the user will obtain definitive proof of “best model performance”.

Despite their inherent difficulties, models provide the key ingredient to a successful credit risk management system. This is because they can be used to estimate true credit-adjusted valuations that correctly reflect the risk-adjusted value of a borrower’s promise to repay. Such credit-adjusted valuations can and should be used for:

- all major derivative exposures;
- callable bonds;
- standby letters of credit;
- other contingent credit lines;
- all value-at-risk calculations;
- all risk-adjusted capital calculations;
- middle-office exposure management;
- mark-to-market real-time trade authorisations; and
- net-income simulation with default adjustment.

Finally, a good credit risk model should be rich enough to allow extension to retail credit scoring as well as small-business credit scoring.

Credit model risk management has a lot in common with loan portfolio management. Diversification of model risk is essential,
and so is the transparency and comprehensiveness of the analytics used. The pursuit of perfection in credit risk modelling will continue for decades, and practical bankers should plan for and implement smooth transitions from one model to the next as the state of the art improves.

2 Version 3.0.
3 Complete results are described in the Kamakura Risk Information Services Technical Guide, version 3.0, February 2004. This confidential document is distributed to KRIS clients and to financial services regulatory agencies upon request.
4 Mapping in such a way that the ranking of companies is not changed is very common in the industry.
5 Based on 1.1 million monthly observations consisting of all listed companies in North America for which data were available from 1989 to 2003, Kamakura Risk Information Services database, version 3.0.
6 Recall that the hazard rate estimation procedure determines the "best" set of explanatory variables from a given set. In the estimation procedure previously discussed, a single financial ratio was a possible outcome, and it was rejected in favour of the multivariable models presented.
7 Private conversation with one of the authors, autumn 2003.
8 It is sometimes believed that, if a derivatives model correctly matches market prices, the model is "proven". This is not a sufficient test. Any reasonable model (with enough time-varying parameters) can be calibrated to match the relevant market prices. Given that this is true for many models, matching market prices cannot be used to differentiate them.
9 It can be argued that another way to test a model is to test its inputs, i.e., its assumptions. For example, if the model assumes constant interest rates, then one can test to see if this assumption is empirically valid.
10 Or using the prices of any other credit risky securities.
11 Under the pseudo-probabilities, \( \tilde{r} = r \).
12 The investment opportunity set is the means and covariances of all the assets' returns. This implies that the mean and covariances with the market return are also deterministic.
13 In this structure, this is the spot rate of interest on default-free debt.
14 This date was chosen because it was the mid-point of the First Interstate dataset.
15 The typical asset has a beta of 1. If the volatility of the market and the firm's asset are equal, the correlation is one as well.

BIBLIOGRAPHY


