Kamakura Corporation
2222 Kalakaua Avenue
Suite 1400
Honolulu, Hawaii 96815, USA
phone 1 808 791 9888
facsimile 1 808 791 9898
www.kamakuraco.com

Advanced Financial Risk Management:

Tools and Techniques for Integrated Credit Risk, Market Risk, Liquidity Risk and Interest Rate Risk Management

April 2012

Donald R. van Deventer
Kenji Imai
Mark Mesler

Part 3: Risk Management Techniques for Credit Risk Analytics

Chapter 16
Reduced form credit models and credit model testing

For many decades, many researchers have been looking for a more comprehensive, all-encompassing framework for risk management than the legacy approaches to credit risk that we review in Chapter 18, our "credit risk museum." Many sophisticated financial practitioners were looking for a synthesis between models of default for major corporations and the credit scoring technology that has been used for decades to rate the riskiness of retail and small business borrowers. The framework long sought by bankers, the reduced form modeling framework, has blossomed dramatically in the years since the original Jarrow-Turnbull paper ("Pricing Derivatives on Financial Securities Subject to Credit Risk," *Journal of Finance*, 50 (1), March 1995) was published in 1995. On December 10, 2003, the Federal Deposit Insurance Corporation announced that it was adopting the reduced form modeling framework for its new Loss Distribution Model (see the website of the FDIC for details of the model) after many years of studying the Merton model, which we describe in Chapter 18, as an alternative.

In this chapter, we introduce the basic concepts of the reduced form modeling technology and the reasons that it has gained such strong popularity.[1]

The Jarrow-Turnbull Model, 1995

Many researchers trace the first reduced form model to Jarrow and Turnbull [1995], and we start our review of the reduced form model with that paper. Jarrow and Turnbull's objective was to outline a modeling approach that would allow valuation, pricing and hedging of derivatives where

- The asset underlying the derivative security might default or
- The writer of the derivative security might default

This kind of compound credit risk is essential to properly valuing the Jarrow-Merton put option on the assets of a financial institution, which we discussed as the best single integrated measure of the total risk of all types that a financial institution has. It is also essential for a more basic objective: measuring risk accurately regardless of the risk measure that one chooses to employ. This compound approach to default is also at the very core of the tsunami of regulatory initiatives (Basel II, Basel II.5, Basel III, Solvency II, and so on) over the last decade because of the prevalence of compound default risks in typical loan products, guarantees, and in derivatives markets. Analyzing compound credit risk is a substantial departure from the popular Black-Scholes approach that is all-pervasive in the market place. The Black-Scholes model assumes

- The company whose common stock underlies an option on the common stock will not default and
- The firm who has written the call option and has an obligation to pay if the option is "in the money" will not default

Ignoring this kind of compound default risk can lead to very large problems, like the potential loss of $2 billion that J.P. Morgan had as of this writing due to its large credit default swap position made public on May 8, 2012. The legacy approaches to credit risk that we discuss in Chapter 18 were not flexible enough nor accurate enough to provide the kind of framework that Jarrow and Turnbull were seeking[2].

The Jarrow-Turnbull Framework

Jarrow and Turnbull assume a frictionless economy where two classes of zero coupon bonds trade.[3] The first class of bonds is risk free zero coupon bonds $P(t,T)$, where t is the current time and T is the maturity date. They also introduce the idea of a "money market fund" $B(t)$ which represents the compounded proceeds of $1 invested in the shortest term risk-free asset starting at time zero. We used the money market fund concept in modeling the risk-free yield curve using the Heath, Jarrow and Morton approach in Chapters 6 through 9.

The second class of bonds trading is the risky zero coupon bonds of XYZ company, which potentially can default. These zero coupon bonds have a bond price $v(t,T)$. For expositional purposes, Jarrow and Turnbull assume that interest rates and the probability of default are statistically independent.[4] They assume that the time of bankruptcy T^* for firm XYZ is exponentially distributed with parameter λ. They assume that there is a Poisson bankruptcy process driven by the parameter μ, which they assume to be

[1] Interested readers who would like more technical details on reduced form models should see van Deventer and Imai [2003], Duffie and Singleton [2003], Lando[2004], and Schonbucher [2003]. See also the many technical papers in the research section of the Kamakura Corporation website www.kamakuraco.com .

[2] See Jarrow and Turnbull for a discussion of the complexities of using a Geske [1979]-style compound options approach to analyze compound credit risks.

[3] This assumption is necessary for the rigorous academic framework of the model but not for the model's practical implementation.

[4] Jarrow and Turnbull discuss how to relax this assumption, which we do in the next section.

a positive constant less than 1.[5] Jarrow and Turnbull assume that the risk-free interest rates are random and consistent with the extended Merton term structure model (the continuous time version of the Ho and Lee model) that we studied in Chapter 13, which presented special cases of the Heath, Jarrow and Morton framework. Jarrow and Turnbull also assume that in the event of bankruptcy $1 of XYZ zero coupon debt pays a recovery rate $\delta<1$ on the scheduled maturity date of the zero (not at the time of default). When bankruptcy occurs, the value of XYZ zero coupon debt drops to $\delta P(T^*,T)$ because the payment at maturity drops from $1 to δ, but the payment of δ is now certain and risk-free. The value of δ can be different for debt of different seniorities. For expositional purposes, Jarrow and Turnbull assume the recovery rate is constant. We relax that assumption in the next section.

Jarrow and Turnbull show that the value of a defaultable zero coupon bond in this model is

$$v(t,T) = \left[e^{-\lambda\mu(T-t)} + (1-e^{-\lambda\mu(T-t)})\delta\right]P(t,T)$$

We can interpret this formula in a practical way by noting a few points. The product ($\lambda\mu$) is the continuous probability of default at any instant in time. The quantity

$$1-e^{-\lambda\mu(T-t)}$$

is the (risk neutral) probability that XYZ company defaults between time t and maturity of the zero coupon bond T. The quantity

$$e^{-\lambda\mu(T-t)}$$

is the (risk neutral) probability that company XYZ does NOT default between time t and maturity of the zero coupon bond at time T. With this in mind, we can see that Jarrow and Turnbull have reached a very neat and, with hindsight, simple conclusion that the value of a zero coupon bond in the Jarrow-Turnbull model is the probability weighted present value of what the bond holder receives if we know there is no default (in which case we have the equivalent of a risk-free bond $P(t,T)$) and what we receive if we know for certain that default will occur, the recovery rate times the risk-free zero coupon bond price $\delta P(t,T)$.

This clever result depends on the assumption that default and interest rates are not correlated. We relax this assumption in the next section. Jarrow and Turnbull then go on to derive closed form solutions for a number of important securities prices:

- European call options on XYZ Company's defaultable zero coupon bond
- Defaultable European call option on XYZ's defaultable zero coupon bond (where the writer of the call option might default, assuming its default probability is independent of both interest rates and the bankruptcy of XYZ)
- European call option on XYZ Company's defaultable common stock

The Jarrow-Turnbull modeling framework is a very important extension of the risk management analytics we have discussed so far, because it provides an integrated and flexible framework for analyzing interest rate risk and credit risk in an integrated way, the central objective of this volume. We have one task left, and that is to discuss the extension of the approach above to interest rate-dependent and macro-economic factor-dependent default.

The Jarrow Model

The Jarrow [1999-2001] model is an important extension of the Jarrow-Turnbull [1995] model, which was perhaps the first reduced form model to experience wide-spread industry acceptance. Many other

[5] This again means that default risk and interest rates (for the time being) are not correlated. The parameter μ technically represents the Poisson bankruptcy process under the martingale process, the risk neutral bankruptcy process.

researchers, most notably Duffie and Singleton [1999], have done very important work on reduced form models but in the remainder of this chapter we will concentrate on Jarrow's work because of the continuity with the Jarrow-Turnbull model we discussed in the previous section.

In the early days of the credit derivatives markets, dealers logically turned to the legacy Merton model (1974) for pricing purposes. Dealers soon realized that the credit default swap market prices were consistently different from those indicated by the Merton model for many of the reasons we discuss in Chapter 18. Dealers were under considerable pressure from both management, risk managers, and external auditors to adopt a modeling technology with better pricing accuracy. Dealers' instinct for self-preservation was an added incentive. The Jarrow-Turnbull model was the first model which allowed the matching of market prices and provided a rational economic basis for the evolution of market prices of everything from corporate debt to credit derivatives.

The original Jarrow-Turnbull model [1995] assumes that default is random but that default probabilities are non-random time dependent functions. The Jarrow [1999-2001] model extends the Jarrow-Turnbull model in a large number of significant respects. First, default probabilities are assumed to be random, with explicit dependence on random interest rates and an arbitrary number of lognormally distributed risk factors. These lognormally distributed risk factors are extremely important in modeling correlated default, a critical factor in the credit crisis of 2007-2009 (driven by home price declines) and the two "lost decades" in Japan (1990-201) in which commercial real estate prices, home prices, and stock prices all suffered very large declines.

Jarrow goes beyond the Jarrow-Turnbull framework by explicitly incorporating a liquidity factor that affects the prices of bonds, but not the prices of common stock. This liquidity factor can be random and is different for each bond issuer. In addition, the liquidity parameter can be a function of the same macro risk drivers that determine the default intensity. We discuss this issue in detail in Chapter 17. Although we use the term "liquidity factor," we intend the term to describe everything affecting bond prices above and beyond potential losses due to default, the intersection of supply and demand for the credit of each borrower. This can include things like bid-offered price spreads, the impact on market prices of the sale of large holdings of a specific bond issue (the Long Term Capital Management phenomenon), general risk aversion in the fixed income market, size of the firm, location of the firm, and so on. We give specific examples in Chapter 17.
The Jarrow model also allows for the calculation of the implied recovery given default δ_i. This parameter, which can be random and driven by other risk factors, is defined by Jarrow as the fractional recovery of δ where ν is the value of risky debt a fraction of an instant before bankruptcy at time τ and the subscript *i* denotes the seniority level of the debt.

$$recovery = \delta_i(\tau)v(\tau-, T:i)$$

This is a different recovery structure than Jarrow-Turnbull, where the recovery rate was specified as a fraction of the principal value to be received at maturity T.

Duffie and Singleton [1999] were the first to use this specification for the recovery rate. They recognized that the traditional bankers' thinking on the recovery rate, expressed as percentage of "principal," made the mathematical derivation of bond and credit derivatives prices more difficult. They also recognized that thinking of recovery as a percentage of principal was too narrow—what is the principal on an "in the money" interest rate cap? On a foreign exchange option? On an undrawn loan commitment? Expressing recovery as a percentage of value one instant before bankruptcy is both more powerful and more general. We can easily convert it back to the traditional "recovery rate as a percentage of principal" to maximize user-friendliness of the concept.

The hazard rate, or "default intensity", in the Jarrow model is given by a linear combination of three

terms:

$$\lambda(t) = \lambda_0 + \lambda_1 r(t) + \lambda_2 Z(t)$$

This is a much simpler expression than the Merton credit model discussed in Chapter 18, and it is much richer than the structure of Jarrow-Turnbull. The first term, λ_0, can be made time dependent, extending the credit model exactly like we extended term structure models in Chapters 6 through 9 and Chapter 13. The term *Z(t)* is the "shock" term with mean zero and standard deviation of one which creates random movements in the macro factor(s) (like home prices, in the case of the 2007 to 2010 credit crisis), which drives default for that particular company.[6] Movements in this macro factor are generally written in this form:

$$dM(t) = M(t)[r(t)dt + \sigma_m dZ(t)]$$

The change in the macro factor, say home prices, is proportional to its value and drives upward at the random risk free interest rate r(t), subject to random shocks from changes in Z(t), multiplied by the volatility of the macro factor, σ_m. While the default intensity in the Jarrow model as we have written it describes interest rates and one macro factor as drivers of default, it is easy to extend the model to an arbitrary number of macro factors. This is because the sum of a linear combination of normally distributed variables like Z(t) is still normally distributed. In this volume, we will use the one factor notation for expositional purposes. Please note that the incorporation of these macro-economic drivers of default mean that the default probabilities (and default intensities) are correlated due to their dependence on common factors. This is a very important step forward from Jarrow-Turnbull where the default probabilities of the two companies in that example were uncorrelated with each other and with interest rates.

In the Jarrow model, interest rates are random, but the term structure model chosen is a more realistic model than the extended Merton model. The term structure model assumed for the riskless rate of interest is a special case of the Heath Jarrow Morton [1992] framework, the Hull-White or Extended Vasicek model which we reviewed in Chapter 13. For practical use, though, it is critical that the risk free interest rate assumptions be completely consistent with the current observable yield curve (and any other interest rate derivatives on that yield curve) as discussed in Chapter 14. The Extended Vasicek

[6] In more mathematical terms, Z(t) is standard Brownian motion under a risk neutral probability distribution Q with initial value 0 that drives the movements[6] of the market index *M(t)*

model contains an "extension" from the original Vasicek model which allows the theoretical yield curve to exactly match the actual yield curve. In this model, as we discussed in Chapter 13, the short term rate of interest r(t) drifts over time in a way consistent with interest rate cycles, subject to random shocks from the Brownian motion W(t), which like Z(t) has mean zero and standard deviation of 1.[7]

$$dr(t) = a[\bar{r}(t) - r(t)]dt + \sigma_r dW(t)$$

Zero Coupon Bond Prices in the Jarrow Model

Jarrow shows that the price of a zero coupon bond can be derived that explicitly incorporates the interactions between the default intensity, interest rates, and the macro factor. The results are simple enough to be modeled in spreadsheet software, especially the case where the coefficients of short term interest rates and the macro factors are zero. We discuss practical implementation of this version of the Jarrow model below.

The Jarrow Model and the Issue of Liquidity in the Bond Market

One of the many virtues of reduced form models is the ability to fit parameters to the model from a wide variety of securities prices. Of course the type of data most directly related to the interests of lenders is bond prices or credit derivatives prices. Since a large company like IBM has only one type of common stock outstanding but could have 10, 20 or 30 bond issues outstanding, it is logical to expect that there is less liquidity in the bond market than in the market for common stock. Jarrow makes an explicit adjustment for this in his model by introducing a very general formulation for the impact of liquidity on bond market prices. This is a very powerful feature of the model that reduces the question of bond market liquidity to a scientific question (how best to fit a parameter which captures the liquidity impact) rather than a philosophical or religious question of whether bond prices can be used to calculate the parameters of a credit model. Note that there can be either a liquidity discount or liquidity premium on the bonds of a given issuer, and there is no implicit assumption by Jarrow that the liquidity discount is constant—it can be random and is flexible in its specifications, consistent with the research we discuss in Chapter 17.

Jarrow's original [1999-2001] model shows how the Jarrow credit model can be fit to bond prices and equity prices simultaneously. It can also be fit to bond prices alone, using the techniques of Chapter 17. Jarrow and Yildirum [2002] show how the model can be fit to credit derivatives prices and Chava and Jarrow [2002a, 2002b] fit the model to historical data on defaults.

The Jarrow-Merton Put Option as a Risk Index and a Practical Hedge

The Jarrow model is much better suited to hedging credit risk on a portfolio level than the Merton model we discuss in Chapter 18 because the link between the (N) macro factor(s) M and the default intensity is explicitly incorporated in the model. Take the example of Exxon, whose probability of default is driven by interest rates and oil prices, among other things. If *M(t)* is the macro factor oil prices,

[7] The expression for random movements in the short rate is again written under a risk neutral probability distribution.

it can be shown that the size of the hedge that needs to be bought or sold to hedge one dollar of risky debt zero coupon debt with market value v under the Jarrow model is given by

$$\partial v_l(t,T:i)/\partial M(t) = \\ -[\partial \gamma_i(t,T)/\partial M(t) + \lambda_2(1-\delta_i)(T-t)/\sigma_m M(t)]v_l(t,T:i).$$

The variable ν is the value of risky zero coupon debt and gamma is the liquidity discount function representing the illiquidities often observed in the debt market. There are similar formulas in the Jarrow model for hedging coupon-bearing bonds, defaultable caps, floors, credit derivatives and so on.

Van Deventer and Imai [2003] show that the steps in hedging the macro factor risk for any portfolio are identical to the steps that a trader of options has been taking for 30 years (hedging his net position with a long or short position in the common stock underlying the options):

- Calculate the change in the value (including the impact of interest rates on default) of all retail credits with respect to interest rates
- Calculate the change in the value (including the impact of interest rates on default) of all small business credits with respect to interest rates
- Calculate the change in the value (including the impact of interest rates on default) of all major corporate credits with respect to interest rates
- Calculate the change in the value (including the impact of interest rates on default) of all bonds, derivatives, and other instruments
- Add these “delta” amounts together
- The result is the global portfolio “delta,” on a default adjusted basis, of interest rates for the entire portfolio
- Choose the position in interest rate derivatives with the opposite delta
- This eliminates interest rate risk from the portfolio on a default adjusted basis

We can replicate this process for any macro factor that impacts default, such as home prices, exchange rates, stock price indices, oil prices, the value of class A office buildings in the central business district of key cities, etc. We should how to do this using logistic regression to estimate default probabilities in a manner consistent with the Jarrow framework.

Most importantly

- We can measure the default adjusted transaction level and portfolio risk exposure with respect to each macro factor
- We can set exposure limits on the default adjusted transaction level and portfolio risk exposure with respect to each macro factor
- We know how much of a hedge would eliminate some or all of this risk

It can be shown that all other risk, other than that driven by macro factors, can be diversified away. This hedging and diversification capability is a powerful step forward from where most financial institutions found themselves at the close of the 20th century.

Fitting the Jarrow Model to Bond Prices, Credit Derivatives Prices, and Historical Default Data Bases

One of the many virtues of the reduced form modeling approach is its rich array of closed form solutions. These include closed form solutions for zero coupon bond prices, coupon bearing bond prices, credit default swap prices, first to default swap prices and many others. This means that there are many alternatives for fitting the Jarrow model parameters. We discuss a few of them in this section.

Fitting the Jarrow Model to Debt Prices

In Chapter 17, we take the reader through an example of how to fit the Jarrow model to observable prices of risky debt. Third party bond price vendors offer data bases with daily or weekly prices on more than 100,000 different risky debt issues. Prices can be obtained for almost every issuer with a rating from the legacy rating agencies.

Fitting to Current Price Data and Historical Price Data

If one is willing to accept the hypothesis that the impact of interest rates and macro-economic factors on the default probability of a particular company is fairly stable over time, one can improve the quality of parameter estimates by adding historical data to the estimation procedure. Let's assume that the analyst has five years of daily price data for, say, IBM bonds. The following procedure takes advantage of this data.

A. Assume a structure of the liquidity function in the Jarrow model. For example, a liquidity function that is effectively linear in years to maturity could be adopted. Say the assumed structure contains two parameters a and b.
B. Collect the N(i) observable bond prices on each observation date i, for the whole sample i=1,M observations.
C. Using a non-linear equation solver over the entire sample, find the combination of a, b, λ_0, λ_1, and λ_2 which minimize the sum of squared errors on the observable bonds. This parallels the interest rate parameter fitting we did in Chapter 14 for the risk free term structure.
D. λ_1 and λ_2 are the coefficients of interest rates and the macro factors on default. We hold these parameters constant over the whole sample and then go back over the sample and refit the parameters a, b, and λ_0 for each day in the sample, allowing them to change daily.
E. Eliminate any remaining bond pricing errors on a given day by making λ_0 a time (to maturity) dependent variable. The original value of λ_0 for that day, as above, is useful to obtain because it is the value which minimizes the amount of extension.

Fitting the Jarrow Model to Credit Derivatives Prices

The procedure for fitting the Jarrow model to credit derivatives prices is exactly the same as fitting the Jarrow model to bond prices, except of course that the closed form solution that is being used is the credit derivatives formula in the Jarrow model, not the bond price formula. Because most counterparties with traded credit derivatives will have only 1-3 maturities at which prices are visible, using a longer history of credit derivatives prices as described in the second method above will be even more helpful than it is in the bond market.

Fitting the Jarrow Model to a Historical Data Base of Defaults

In the first part of this chapter, the continuous time default "intensity" was the object of study, just like the continuous time interest rates we studied in Chapters 6 through 13. In both the interest rate arena and the credit arena, most applied risk analysis is done using discrete "rates," either interest rates or default probabilities. For the remainder of this chapter, the default probabilities we discuss all have a discrete time to maturity. In this section, we discuss how to fit a default probability function driven by macro-economic factors to a discrete historical data set of observations on public firms at various points in time, some of which constitute a "defaulting observation" and some of which constitute a "non-defaulting observation."

Fitting the reduced form model to historical bankruptcy data was explored extensively by Shumway [1998], Chava and Jarrow [2004], Campbell, Hilscher and Szilagyi [2008, 2011], Bharath and Shumway [2008]. Each of these authors used public firm data, but the identical approach is also applied to the construction of default models for all types of counterparties: retail borrowers, small business borrowers, large un-listed borrowers, municipal borrowers, and sovereign borrowers. This fitting process involves advanced hazard rate modeling, also called logistic regression, an extension of the credit scoring technology that has been used for retail and small business default probability estimation for many decades. The distinction between credit scoring and hazard rate modeling is very important. Typically, credit scoring analysis uses the most recent information available on a set of non-defaulting counterparties and the last set of information available on the defaulting counterparties. Hazard rate modeling makes use of all information about all counterparties at every available time period, in order to explain, for example, why Lehman Brothers defaulted in September 2008 instead of at some earlier time.

We illustrate the creation and testing of discrete default probability models using data from Kamakura Risk Information Services' KRIS public firm default model, first offered in 2002. The Kamakura Corporation's KRIS default probability data base includes 1.76 million monthly observations on over 8,000 companies for North America alone. In the North American sample, 2046 companies have failed since 1990. A company failure is denoted as a "1" in the bankruptcy variable, and the lack of failure is denoted as a "0". Since the entire history of a company is included, Lehman Brothers will have many observations that are zeros and only one observation which is a 1, the September 2008 observation (recorded as of the end of August 2008). Here is a sample of a simple logistic regression data base where the explanatory variables are market leverage, 3 month stock return volatility, the net income to total assets ratio, and the unemployment rate (percent). The variable that we are seeking to explain is the probability of default (the default flag). The default flag is set to zero if the firm does not default in the next month and it is set to one if the company does default over that time period. While the table below includes only entries for Lehman Brothers, all of the companies are modeled together and a set of explanatory variables are gathered which best predict default.

Figure 1

Logistic Regression Default Data Base
Sample Entries for Lehman Brothers Holdings

Ticker	Date	Company Name	Default Flag	Market Leverage	3 Month Stock Volatility	Net Income to Total Assets	Unemployment Rate (Percent)
lehmq	1/31/2008	Lehman Brothers Holdings	0	0.9495	0.6966	0.0013	5.4
lehmq	2/29/2008	Lehman Brothers Holdings	0	0.9610	0.6244	0.0013	5.2
lehmq	3/31/2008	Lehman Brothers Holdings	0	0.9699	1.5194	0.0006	5.2
lehmq	4/30/2008	Lehman Brothers Holdings	0	0.9689	1.5422	0.0006	4.8
lehmq	5/30/2008	Lehman Brothers Holdings	0	0.9740	1.5566	0.0006	5.2
lehmq	6/30/2008	Lehman Brothers Holdings	0	0.9858	1.0291	-0.0043	5.7
lehmq	7/31/2008	Lehman Brothers Holdings	0	0.9846	1.4784	-0.0043	6.0
lehmq	8/29/2008	Lehman Brothers Holdings	1	0.9857	1.6163	-0.0043	6.1

The result is a set of coefficients (the alpha and betas) of the logistic regression formula which predicts the unannualized probability of default at a given time t P[t] as a function of n explanatory variables:

$$P[t] = \frac{1}{1 + e^{-\alpha - \sum_{i=1}^{n} \beta_i X_i}}$$

In addition to interest rates and macro-factors that are part of the theory of the Jarrow model, other explanatory variables which are likely to be important are added as appropriate. These include financial ratios, other macro-economic variables, multiple inputs from the stock price and its history for each company, multiple inputs from the comprehensive stock market index for the relevant country, company size, industry variables, seasonality and so on. As we will see in the model testing section of this chapter, the result is a very powerful model highly capable of discriminating among companies by both long term and short term risk. Logistic regression is the tool that is used to estimate the parameters or coefficients multiplying each input variable. Appendix A and Chava and Jarrow [2004] discuss the use of default probabilities derived from a discrete data base in a continuous time model, just like interest rates smoothed from discrete observations are used in a continuous time term structure model. For a detailed review of logistic regression and its applications, see van Deventer and Imai (2003) and Hosmer and Lemeshow (2000).

Kamakura Risk Information Services produces more than a simple one month default probability. The KRIS service provides default probabilities for a full 10 year horizon, constructed from default probabilities projected for the next 120 months. We call the un-annualized monthly default probability for the first month P[1], for the second month (conditional on surviving the first month) P[2], and the nth month (conditional on surviving the first N-1 months P[N]. How are these default probabilities combined to form a term structure of default probabilities? Please see Appendix B for details on changing the periodicity of default probabilities for background.

If one has a time horizon of N months, the probability that a company will NOT go bankrupt during those N months is

Q[N]=(1-P[1])(1-P[2])(1-P[3])…(1-P[N-1])(1-P[N])

The probability that the company WILL go bankrupt by the end of the Nth month is

P[N]=1-Q[N]

This is the cumulative probability of default for an N month time horizon. Since credit spreads and credit default swaps are all discussed or displayed on an annualized basis, we can annualize the cumulative probability of default for an N month horizon. The annualized default probability for an N month horizon is

$A[N]=1-(1-P[N])^{12/N}$

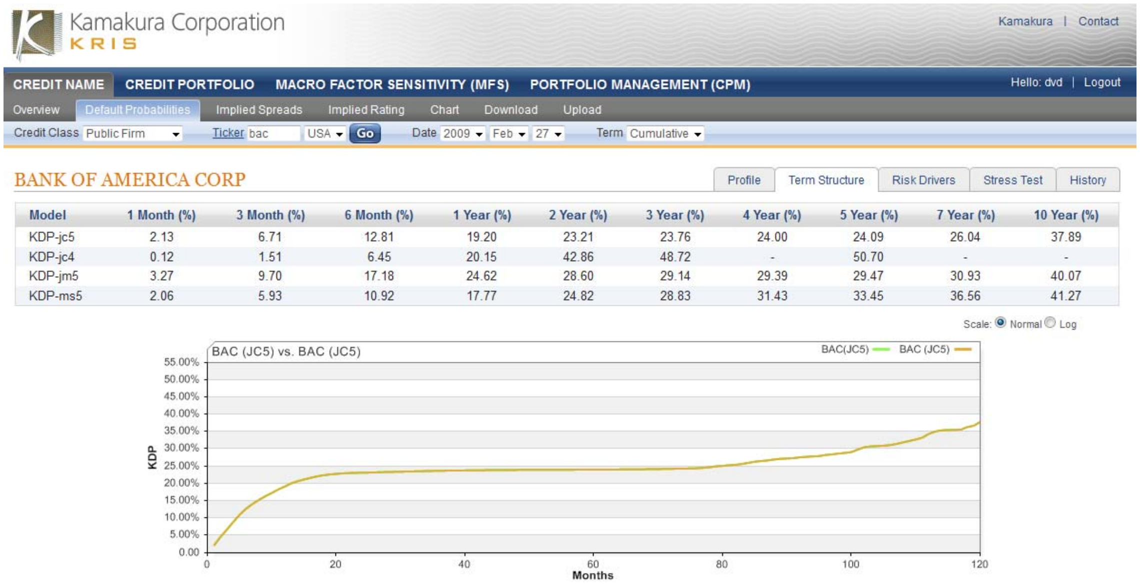
It is these annualized default probabilities that are displayed on the KRIS service. Figure 2 shows the annualized KRIS term structure of default for Bank of America on February 27, 2009, which was a dangerous period in the 2007 to 2010 credit crisis. The annualized 3 month default probability for Bank of America on this date was 24.27%:

Figure 2

Model	1 Month (%)	3 Month (%)	6 Month (%)	1 Year (%)	2 Year (%)	3 Year (%)	4 Year (%)	5 Year (%)	7 Year (%)	10 Year (%)
KDP-jc5	22.76	24.27	23.98	19.20	12.37	8.65	6.63	5.36	4.22	4.65
KDP-jc4	1.48	5.92	12.48	20.15	24.41	19.96	-	13.19	-	-
KDP-jm5	32.92	33.50	31.41	24.62	15.50	10.85	8.33	6.75	5.15	4.99
KDP-ms5	22.07	21.70	20.65	17.77	13.30	10.72	9.00	7.82	6.29	5.18

Figure 3 shows the same default risk for Bank of America, this time displayed on a cumulative basis. The graph shows that the 10 year cumulative failure risk of the bank was 37.89% over a ten year period, even after an injection of $45 billion into the organization under the U.S. government's Troubled Asset Relief Program.

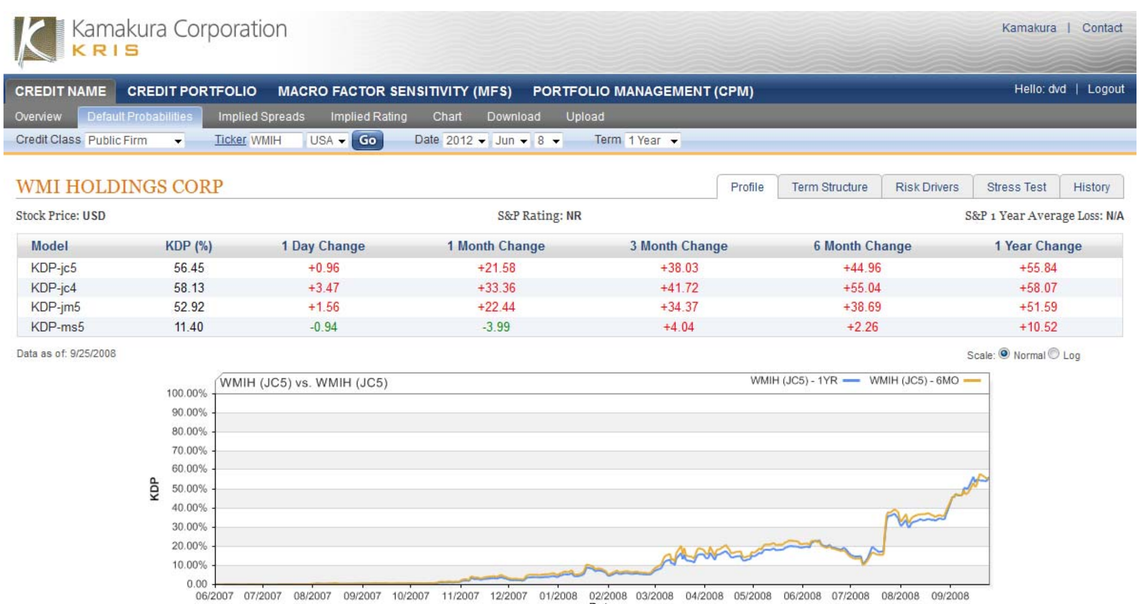
Figure 3



Model	1 Month (%)	3 Month (%)	6 Month (%)	1 Year (%)	2 Year (%)	3 Year (%)	4 Year (%)	5 Year (%)	7 Year (%)	10 Year (%)
KDP-jc5	2.13	6.71	12.81	19.20	23.21	23.76	24.00	24.09	26.04	37.89
KDP-jc4	0.12	1.51	6.45	20.15	42.86	48.72	-	50.70	-	-
KDP-jm5	3.27	9.70	17.18	24.62	28.60	29.14	29.39	29.47	30.93	40.07
KDP-ms5	2.06	5.93	10.92	17.77	24.82	28.83	31.43	33.45	36.56	41.27

Another firm that figured prominently in the 2007 to 2010 credit crisis was the aggressive mortgage lender Washington Mutual (labeled WMI Holdings in the graphs below). Washington Mutual's 1 year default probability hit 56.45% on the eve of September 25, 2008, the day that the Federal Deposit Insurance Corporation seized its main banking subsidiary and sold it in an assisted rescue to JPMorgan Chase.

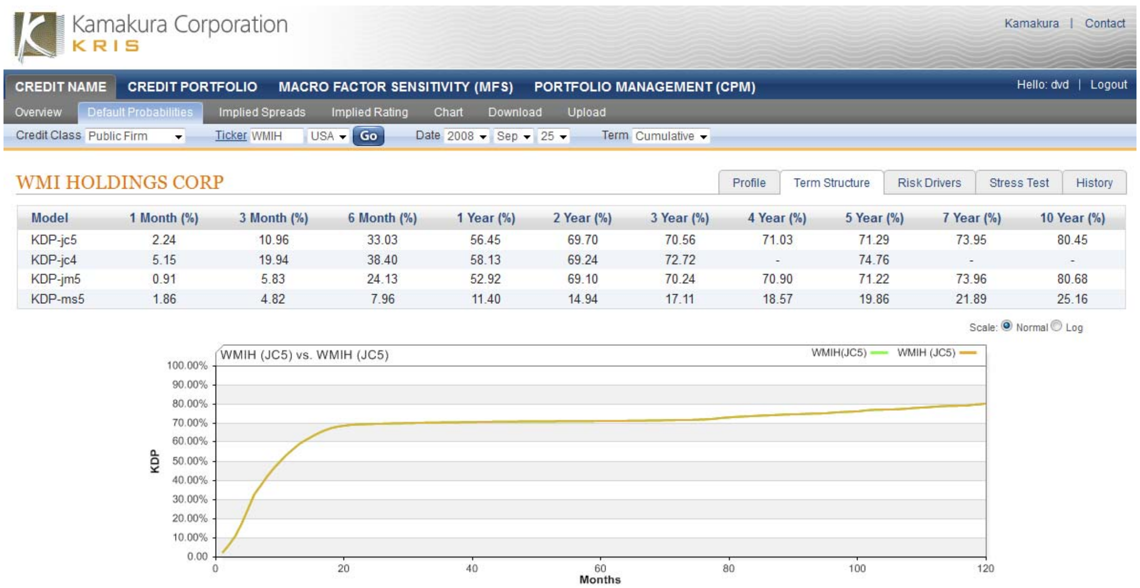
Figure 4



Model	KDP (%)	1 Day Change	1 Month Change	3 Month Change	6 Month Change	1 Year Change
KDP-jc5	56.45	+0.96	+21.58	+38.03	+44.96	+55.84
KDP-jc4	58.13	+3.47	+33.36	+41.72	+55.04	+58.07
KDP-jm5	52.92	+1.56	+22.44	+34.37	+38.69	+51.59
KDP-ms5	11.40	-0.94	-3.99	+4.04	+2.26	+10.52

On that day, the cumulative default risk of Washington Mutual was 80.45% for a time horizon of 10 years as shown in Figure 5:

Figure 5



Model	1 Month (%)	3 Month (%)	6 Month (%)	1 Year (%)	2 Year (%)	3 Year (%)	4 Year (%)	5 Year (%)	7 Year (%)	10 Year (%)
KDP-jc5	2.24	10.96	33.03	56.45	69.70	70.56	71.03	71.29	73.95	80.45
KDP-jc4	5.15	19.94	38.40	58.13	69.24	72.72	-	74.76	-	-
KDP-jm5	0.91	5.83	24.13	52.92	69.10	70.24	70.90	71.22	73.96	80.68
KDP-ms5	1.86	4.82	7.96	11.40	14.94	17.11	18.57	19.86	21.89	25.16

Fitting the Jarrow Model to Retail, Small Business, and Governmental Counterparties

While the Jarrow-Turnbull and Jarrow models have been discussed in the context of a corporate counterparty, there is no reason to restrict the models' application to corporations alone. They can be applied with equal effectiveness to retail client default modeling, small business default modeling, and to the modeling of defaults by sovereign entities. The only difference is the inputs to the logistic regression model like that which we discuss in the preceding paragraphs. Other than that, all of the analysis is identical.

For example, consider sovereigns which issue debt but for whom there are no observable prices (say, Kazakhstan), historical estimation using logistic regression is the normal fall-back estimation procedure. Kamakura Risk Information Services offers a sovereign default service developed in exactly this way and launched in the fall of 2008.

In many countries there is a very large amount of bond issuance by cities, towns, municipal special purpose entities, and other government entities besides the national government itself. The prices of the bonds issued by these entities are sometimes heavily affected by special tax features, such as the exemption of municipal bonds from federal income tax in the United States. Jarrow default probabilities can be fitted to this group as well, both from bond prices (if they are high quality) and by logistic regression from a historical default data base.

Correlations in Default Probabilities

In the same way the Jarrow models fitted to historical default data bases illustrate the parallels between default probabilities and interest rates:

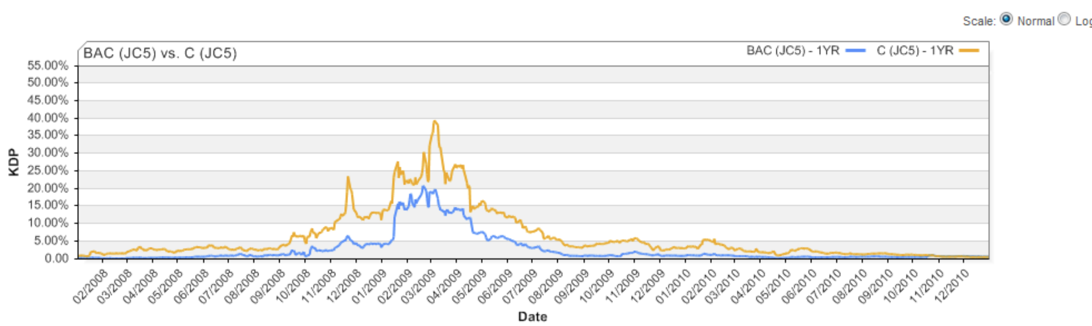
- Default probabilities have a term structure just like interest rates

- Changes in interest rates and macro-economic factors make the term structure of default probabilities rise and fall over time
- Like interest rates, short term default probabilities rise and fall by much more than long term default probabilities, which may be the effective average of default probabilities over many credit cycles
- The difference between a strong credit and a weak credit is not their short term default probability when times are good. In good times, both will have a low probability of default. The difference can be seen most clearly when times are bad.

Macro-factors cause correlated movements in the short term default probabilities for both strong and weak credits.

A few examples will illustrate this phenomenon. The first example, Figure 6, shows the one year default risk of Bank of America and Citigroup from 2008 to 2010:

Figure 6



The simple five year monthly correlation in their one year default probabilities for the period ending in June 2012 of their Jarrow default probabilities is 91.87%. That contrasts sharply with the 40.50% correlation in the 1 year default probabilities of Apple and Bank of America over the same five year period. Figure 7 shows the 1 and 5 year default probabilities for Apple from 2007 to 2012. While Apple's default risk was a small fraction of that of Bank of America, Apple also experienced a rise in 1 year default risk in late 2008 and early 2009 that was very similar to that of Bank of America. This correlated shift in risks, with a large difference in magnitude, is typical of the change in default risk for good credits versus bad credits in a credit portfolio:

Figure 7

The Jarrow and Jarrow-Turnbull Models: A Summary

In this section, we have introduced the Jarrow-Turnbull and Jarrow models of risky debt. They offer powerful closed form solutions for practical issues that are central to the integrated treatment of interest rate risk, market risk, liquidity risk and credit risk. We have shown how to fit the Jarrow model to historical default data bases. In Chapter 17, we show how to fit a credit model to observable spreads in the bond markets and derivatives markets. We have shown how the reduced form credit modeling approach, with different explanatory variables, can be applied to

- Retail counterparties
- Small business counterparties
- Municipalities
- Sovereigns

as well as to public firms.

Finally, we have shown through some very practical examples that the Jarrow reduced form models shed dramatic light on both the term structure of default probabilities for any type of counterparty and on the correlation of default probabilities for any two counterparties. This kind of correlated risk is what drives the "fat tails" in losses that endanger financial institutions. An understanding of this issue is essential to both the integrated measure of risk via the Jarrow-Merton put option and to the hedging of portfolio level credit risk with macro hedges using stock price index futures and derivatives on key macro-economic factors.
In the next section, we discuss in detail how to measure the effectiveness of default models, both in terms of absolute accuracy and relative accuracy compared to other models.

Tests of Credit Models Using Historical Data

This section is one of the key sections in this book. A debate about credit models without a factual basis for deciding which model is best is just as much of a beauty contest as the debate about smoothing methods that we discussed in Chapter 5. In this section, we replace the beauty contest with science. We define criteria for "best model," as we did in Chapter 5. We then measure candidate models by these criteria to reach our conclusions about which is best.

This approach is very consistent with the Basel II and Basel III capital regulations that have emerged in recent years from the Basel Committee on Banking Supervision. The Basel capital regulations place very heavy emphasis on the role of credit risk in measuring the safety and soundness of financial institutions. Solvency II for the insurance industry has a similar emphasis on model accuracy. For that reason, it is extremely critical from both a shareholder value point of view and a regulatory point of view that financial institutions have a clear understanding of the accuracy and performance of the credit models employed for that purpose.

Without knowledge of the accuracy of the models,

- How can regulators know whether the financial institution's capital is too high or too low?
- How can shareholders know whether the reserve for loan losses is adequate or not?
- How can the CEO certify the financial statements of the company as accurate?
- How can the institution have confidence that a hedge of its portfolio credit risk will work?
- How can the institution have confidence in its own survival?

These are critical questions, to say the least, and they are the major motivators of the emphasis on credit model testing in this section. The Basel Accords' contribution to credit risk modeling has been two-fold. First, the Accords stress the need to apply credit risk models to the entire balance sheet, not just to credit-related derivatives like credit default swaps. 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. See Allen and Saunders [2003] in working paper 126 of the Bank for International Settlements for more on the importance of cyclicality in credit risk modeling.

The result of these 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 December 10, 2003 announcement by the Federal Deposit Insurance Corporation that it has adopted a reduced form model-related Loss Distribution Model that we mentioned in Chapter 18.[8] The FDIC announcement is just another confirmation of a wholesale movement away from the older structural models of credit risk like the original Merton [1974] risky debt model and more recent extensions like Shimko, Tejima and van Deventer [1993], which we discuss in Chapter 18. We do something atypical in this book. Instead of explaining all of the candidate models first, like we did with yield curve smoothing, and then measuring "who's best," we do something different. As the reader will

[8] Please see Jarrow, Bennett, Fu, Muxoll, and Zhang, "A General Martingale Approach to Measuring and Valuing the Risk to the FDIC Insurance Funds," working paper, November 23, 2003 and available on www.kamakuraco.com or on the web site of the Federal Deposit Insurance Corporation.

see in the following pages, the credit model test results in both the academic and commercial public domain are so compelling that we reverse the order. After explaining the reduced form approach, but not legacy approaches, we then measure accuracy. Having seen "who's best" in this way justifies our relegation of older approaches to the "legacy" collection of credit models in Chapter 18, much like the legacy term structure models that we reviewed in Chapter 13.

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, fund manager, or insurance executive 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 section.

An Introduction to Credit Model Testing

One of the dilemmas facing the Basel Committee on Banking Supervision as it compiled the New Capital Basel Capital Accord (May 31, 2001) was the (then) paucity of data on credit model performance. In the following decade, however, and the years immediately prior to the publication of the New Capital Accord, an avalanche of studies on relative performance of credit models has reached the public domain. We refer the series reader to Shumway (1998), van Deventer and Imai (2003), Chava and Jarrow (2004), Bharath and Shumway (2008), and Campbell, Hilscher and Szilagyi (2008 and 2011) for some notable examples. Almost all of these papers were in wide circulation for 4 or 5 years prior to their formal publication.

Credit model testing is not a trivial exercise, for a number of reasons. The first is the considerable expense and expertise needed, both in terms of finance and computer science, to assemble the data and computer coding to provide a consistent methodology for testing credit models in a way that the original Basel II accord (and good corporate governance) requires. Van Deventer and Imai [2003] note that the Basel II requires that banks must prove to their regulatory supervisors that the credit models they use perform "consistently and meaningfully."[9] Typically, the only institutions who have the capability to assemble these kinds of data bases are extremely large financial institutions and commercial vendors of default probabilities. Prior to the commercialization of default probabilities by KMV (acquired by Moody's in 2002) in the early 1990s, studies of default were based on a very small number of defaulting observations. Falkenstein and Boral [2000] cite academic papers by Altman [1968] (33 defaults), Altman [1977] (53 defaults), and Blum [1974] (115 defaults) to illustrate the relatively small sample sizes used to draw inferences about bankruptcy probabilities prior to 1998. By way of contrast, Kamakura Corporation's commercial default data base for public firms includes more than 2,000 failed company observations, and its research data base, which spans a longer period, contains more than 2,400 failed companies for North America alone as of 2012.

For major financial institutions that have incurred the expense of a large default data base, the results of model testing are highly valuable and represent a significant competitive advantage over other financial institutions who do not have the results of credit model performance tests. For

[9] Section 302, p. 55, The New Basel Capital Accord, Basel Committee on Banking Supervision, May 31, 2001.

example, there is a large community of arbitrage investors actively trading against users of the Merton default probabilities when the arbitrage investors perceive the signals sent by the Merton model to be incorrect.

Among the vendor community, the majority of vendors offer a single default probability model. This presents a dilemma for potential consumers of commercial default probabilities. A vendor of a single type of credit model has two reasons not to publish quantitative tests of performance. The first reason is that the tests may prove that the model is inferior and ultimately may adversely affect the vendor's commercial prospects. Perhaps for this reason, most vendors require clients to sign license agreements that forbid the clients from publicizing any results of the vendor's model performance The second reason is more subtle. Even if quantitative performance tests are good, the fact that the vendor offers only one model means that the vendor's tests will be perceived by many as biased in favor of the model that the vendor offers.

Four former employees of Moody's Investors Service have set the standard for quantitative model test disclosure in a series of papers: Andrew Boral, Eric Falkenstein, Sean Keenan, and Jorge Sobehart. The authors respect the important contributions of Boral, Falkenstein, Keenan, and Sobehart to the integrity of the default probability generation and testing process. Much of the work in this section can be attributed to their influence and the influence of the authors mentioned above.

The need for such tests is reflected in the frequently heard comments of default probability users who display a naiveté with respect to credit models that triggered the massive losses from trading in collateralized debt obligations during the 2007-2010 credit crisis. We present some samples in the next section which illustrate the need for better understanding of credit model testing.

Misunderstandings About Credit Model Testing

A commonly heard comment on credit model performance goes like this:

"I like Model A because it showed a better early warning of the default of Companies X, Y and Z"

Many users of default probabilities make two critical mistakes in assessing default probability model performance. They choose a very small sample (in this case three companies) to assess model performance and use a naïve criteria for good performance. Assessing model performance on only three companies or 50 or even 100 in a universe of 8,000-10,000 in the total universe of US public firms needlessly exposes the potential user to: (a) an incorrect conclusion just because of the noise in the small sample; and (b) to the risk of data mining by the default probability vendor, who (like a magician doing card tricks) can steer the banker to the three or 50 or 100 examples which show the model in the best light. A test of the whole sample eliminates these risks. Analysts should demand this of both internal models and models purchased from third parties.

The second problem this banker's quote has is the performance criteria. The implications of his comment are two fold:

- I can ignore all false predictions of default and give them zero weight in my decision.
- If Model A has higher default probabilities than Model B on a troubled credit, then model A must be better than model B.

Both of these implications should be grounds for a failing grade by banking and insurance supervisors and the audit committee of the firm's board of directors. The first comment, ignoring all false positives, is sometimes justified by saying "I sold Company A's bonds when its default probabilities hit 20% and saved my bank from a loss of $1.7 million, and I don't care if other companies which don't default have 20% default probabilities because I would never buy a bond with a 20% default probability anyway." Why, then, did the bank have the bond of Company A in its portfolio? And what about the bonds that were sold when default probabilities rose, only to have the bank miss out on gains in the bond's price that occurred after the sale? Without knowledge of the gains avoided, as well as the losses avoided, the banker has shown a striking "selection bias" in favor of the model he is currently using. This selection bias will result in any model being judged good by a true believer. We give some examples below.

The second implication exposes the banker and the vendor to a temptation that can be detected by the tests we discuss below. The vendor can make any model show "better early warning" than any other model simply by raising the default probabilities. If the vendor of model B wants to win this banker's business, all he has to do is multiply all of his default probabilities by 6 or add an arbitrary scale factor to make his default probabilities higher than model A. The banker making this quote would not be able to detect this moral hazard because he does not use the testing regime mentioned below. The authors can't resist pointing out that using the summer temperature in Fahrenheit as a credit model will outpredict any credit model by this criterion if the company defaults in July or August in the Northern Hemisphere.

Eric Falkenstein and Andrew Boral[10] [2000] of Moody's Investors Service address the issue of model calibration directly:

"Some vendors have been known to generate very high default rates, and we would suggest the following test to assess those predictions. First, take a set of historical data and group it into 50 equally populated buckets (using percentile breakpoints of 2%, 4%, …100%). Then consider the mean default prediction on the x-axis with the actual, subsequent bad rate on the y-axis. More often than not, models will have a relation that is somewhat less than 45% (i.e., slope <1), especially at these very high risk groupings. This implies that the model purports more power than it actually has, and more importantly, it is mis-calibrated and should be adjusted."

We present the results of the Falkenstein and Boral test later in this section. We also present a second type of test to detect this kind of bias in credit modeling below. If a model has a bias to levels higher than actual default rates, it is inappropriate for Basel II, Basel III or Solvency II use because it will be inaccurate for pricing, hedging, valuation, and portfolio loss simulation.

Another quotation illustrates a similar point of view that is inconsistent with Basel II compliance in credit modeling:

"That credit model vendor is very popular because they have correctly predicted 10,000 of the last 10,500 small business defaults."

Again, this comment ignores false predictions of default and assigns zero costs to false predictions of default. If any banker truly had that orientation, the Basel II credit supervision process will root them out with a vengeance because the authors hereby propose a credit model at zero cost that outperforms the commercial model referred to above:

[10] Page 46.

100% Accurate Prediction of Small Business Defaults: The Default Probability for All Small Businesses is 100%

This naïve model correctly predicts 10,500 of the last 10,500 defaults. It is free in the sense that assigning a 100% default probability to everyone requires no expense or third party vendor since any one can say the default probability for everyone is 100%. And, like the banker quoted above, it is consistent with a zero weight on the prediction of false positives. When pressed, most financial institutions admit that false positives are important. As one major financial institution comments, one model "correctly predicted 1,000 of the last three defaults."
Once this is admitted, there is a reasonable basis for testing credit models.

The Two Components of Credit Model Performance

The performance of models designed to predict a yes or no - zero or one - event, like bankruptcy, has been the focus of mathematical statistics for more than fifty 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 modeling in recent years. Van Deventer and Imai [2003] provide an overview of these standard statistical tests for default probability modeling. Related papers are those cited at the beginning of this section: Shumway (1998), Chava and Jarrow (2004), Bharath and Shumway (2008), and Campbell, Szilagyi and Hilscher (2008, 2011). Jarrow, van Deventer and Wang [2003] 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.
Basel II and Basel III require that financial institutions have the capability to test credit model performance and internal ratings to ensure that they consistently and meaningfully measure credit risk. There are two principal measures of credit risk model performance. The first is a measure of the correctness of the ordinal ranking of the companies by riskiness. For this measure, we use the so-called receiver operating characteristics (ROC) Accuracy Ratio, whose calculation is reviewed briefly in the next section. The second is a measure of the consistency of the predicted default probability with the actual default probability, which Falkenstein and Boral [2000] call "calibration." This test is necessary to ensure the accuracy of the model for pricing, hedging, valuation and portfolio simulation. Just as important, it is necessary to detect a tendency for a model to bias default probabilities to the high side as Falkenstein and Boral note, which overstates the predictive power of a model by the naïve criteria of the first quote in the introduction.

Measuring Ordinal Ranking of Companies by Credit Risk

The standard statistic for measuring the ordinal ranking of companies by credit riskiness is the ROC accuracy ratio. The ROC accuracy ratio is closely related to, but different than, the cumulative accuracy profiles used by Jorge Sobehart and colleagues [2000], formerly at Moody's Investors Service and now at Citigroup, in numerous publications in recent years.

The receiver operating characteristics (ROC) curve was originally developed in order to measure the signal to noise ratio in radio receivers. The ROC curve has become increasingly popular as a measure of model performance in fields ranging from medicine to finance. It is typically used to measure the performance of a model that is used to predict which of two states will occur (sick or not sick, defaulted or not defaulted, etc.). van Deventer and Imai [2003] go into extensive detail on the meaning and derivation of the ROC accuracy ratio, which is a quantitative measure of model performance.

In short, the ROC accuracy ratio is derived in the following way:

- Calculate the theoretical default probability for the entire universe of companies in a historical data base that includes both defaulted and non-defaulted companies
- Form all possible pairs of companies such that the pair includes one defaulted "company" and one non-defaulted "company". To be very precise, one pair would be the December, 2001 defaulted observation for Enron and the October, 1987 observation for General Motors, which did not default in that month. Another pair would include defaulted Enron, December 2001, and non-defaulted Enron, November 2001, and so on.
- If the default probability technology correctly rates the defaulted company as more risky, we award one point to the pair.
- If the default probability technology results in a tie, we give half a point
- If the default probability technology is incorrect, we give zero points
- We then add up all the points for all of the pairs, and divide by the number of pairs[11]

The results are intuitive and extremely clear on model rankings:

- A perfect model scores 1.00 or 100% accuracy, ranking every single one of the defaulting companies as more risky than every non-defaulting company
- A worthless model scores 0.50 or 50%, because this is a score which could be achieved by flipping a coin
- A score in the 90% range is extremely good
- A score in the 80% range is very good, and so on

Van Deventer and Imai provide worked examples to illustrate the application of the ROC accuracy ratio technique. The ROC accuracy ratio can be equivalently summarized in one sentence:

It is the average percentile ranking of the defaulting companies in the universe of non-defaulting observations.

We turn now to the Kamakura Corporation database which we shall use to apply the ROC accuracy ratio technology. Results will be summarized for the North American data base, which includes all listed companies for which the explanatory variables were available, with no exceptions for any other reason.

The Predictive ROC Accuracy Ratio: Techniques and Results

The Kamakura Risk Information Services data base is a monthly data base. Other researchers have used annual data, including the work of Sobehart and colleagues noted above. Annual data has been used in the past because of the researchers' interest in long-term bankruptcy prediction. One of the purposes of this chapter is to show how monthly data can be used for exactly the same purpose, and to show how accuracy changes as the prediction period grows longer.

The basic Jarrow-Chava [2004] model uses logistic regression technology to combine equity market and accounting data for default probability prediction consistent with the reduced form modeling approach. Jarrow and Chava provide a framework that shows these default probability estimates

[11] This calculation can involve a very large number of pairs. The current commercial database at Kamakura Corporation involves the comparison of 1.55 billion pairs of observations, but on a modern personal computer as of this writing processing time for the exact calculation is very quick.

are consistent with the best practice reduced form credit models of Jarrow [1999-2001] and Duffie and Singleton [1999]. We turn now to the ROC accuracy ratios for the Jarrow-Chava reduced form modeling approach.

The Predictive Capability of the Jarrow-Chava Reduced Form Model Default Probabilities

This section shows that the technology for measuring credit model predictive capability is transparent and straightforward. We show that the reduced form modelling approach has a high degree of accuracy even when predicting from a distant horizon.

The standard calculation for the ROC accuracy ratio is always based on the default measure in the period of default (which could be a month if using monthly data, or a year if using yearly data). In order to address Basel and financial institutions management questions about the "early warning" capability of default probability models, it is useful to know the accuracy of credit models at various time horizons, not just in the period of default.

We know that, in their distant past, companies which defaulted were at some point the same as other companies and their default probabilities were probably no different from those of otherwise similar companies. In order words, if we took their default probability in that benign era and used that to calculate the ROC accuracy ratio, we would expect to see an accuracy ratio of 0.50 or 50%, because at that point in their history the companies are just "average" and so are their default probabilities.

Therefore, our objective is to see how far in advance of default the ROC accuracy ratio for the defaulting companies rises from the 50% level on its way to achieving the standard ROC accuracy ratio in the month of default.

Measuring the Predictive ROC Accuracy Ratio

We calculate the time series of predictive ROC accuracy ratios in the following way:

- We again form all possible pairs of one defaulted company and one non-defaulted company. Instead of defining the "default date" and associated default probability for Enron as December, 2001, however, we step back one month prior to default and use the November time period for Enron as the default date and the November 2001 default probability as the "one-month predictive" default probability.
- We calculate the ROC accuracy ratio on this basis.
- We lengthen the prediction period by one more month again.
- We recalculate which default probability we will use in the ROC comparison for each of the defaulted companies (in the second iteration, we will use the October 2001 default probability for Enron, not November).
- We repeat the process until we have studied the predictive period of interest and have calculated the predictive ROC accuracy ratio for each date.

Reduced Form Model Versus Merton Model Performance

We now supplement the results if Bharath, Campbell, Chava, Hilscher, Jarrow, Shumway, and Szilagyi with results from the Kamakura Risk Information Services testing results for version 5.0, which uses monthly data spanning the period from 1990 to 2008. We compare the accuracy of the ordinal ranking of credit risk for three credit models distributed by the Kamakura Corporation:

Reduced form credit model

- **KDP-jc**, Kamakura Default Probabilities from an advanced form of the Jarrow-Chava [2004] approach. Version 5 of the model is denoted KDP-jc5.

Structural credit models

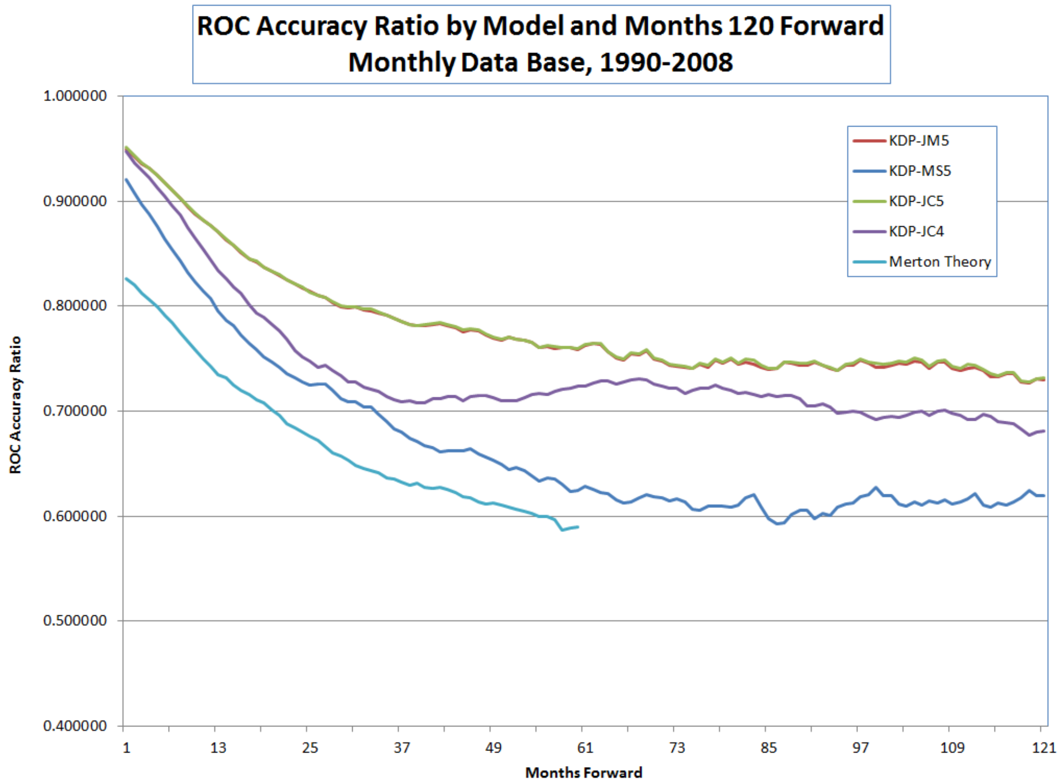
- **KDP-ms**, Kamakura Default Probabilities using the "best" Merton econometric approach. This model is labeled KDP-ms5 in version 5.

Hybrid credit models

- **KDP-jm**, Kamakura Default Probabilities combining the Jarrow and Merton approaches in a hybrid model within the logistic regression framework. A transformation of the KDP-ms Merton default probability is added as an additional explanatory variable to the Jarrow-Chava variables in KDP-jc to form KDP-jm, denoted KDP-jm5 for version 5.

The chart below shows the ROC accuracy ratios for version 5 of all three of the models based on monthly data from 1990 to 2008, which includes the most dramatic failures in the 2007 to 2010 credit crisis. For comparison, we also plot the accuracy of an older reduced form model, version 4, labeled KDP-JC4, and a "pure theory" implementation of the Merton model. The chart shows clearly that the accuracy of the credit risk ranking of both the KDP-jc5 Jarrow-Chava reduced form model and the hybrid KDP-jm5 model are consistently superior to the KDP-ms Merton model for all forecasting horizons for 120 months prior to default. In the final month before default ("month 1" to model credit modelers), the ROC accuracy ratios are as follows: 95.10 for KDP-JC5 (Jarrow-Chava reduced form model version 5), 94.93 for KDP-JM5 (Jarrow-Merton hybrid model version 5), 94.69 for KDP-JC4 (Jarrow-Chava reduced form model version 4), 92.09 for KDP-MS5 (Merton structural model, econometric version 5), and 82.59 for the Merton theoretical model. For a detailed explanation of the explanatory variables used in the KRIS models, please contact info@kamakuraco.com. The Merton model implementation strategies are explained in Chapter 18. The differential of more than 12 percentage points versus the Merton theoretical model comes as no surprise as it is consistent with the early work of Jorge Sobehart and colleagues and Bohn et al [2005], as well as the many papers by Chava and Jarrow, Bharath and Shumway, and Campbell et al. Please note that the ROC accuracy ratio for month N is the accuracy of predicting default in month N, conditional on the firm not defaulting in months one through N-1. This is a much more difficult prediction that predicting if the firm defaulted in months 1 through N. Accuracy ratios for that cumulative default prediction would be much higher for all of the models.

Figure 8



With this kind of precision in ranking of models by accuracy for any user-defined forecasting horizon, there should be no question about the superiority of Model A vs. Model B when it comes to the Basel II and III requirements for credit model testing.

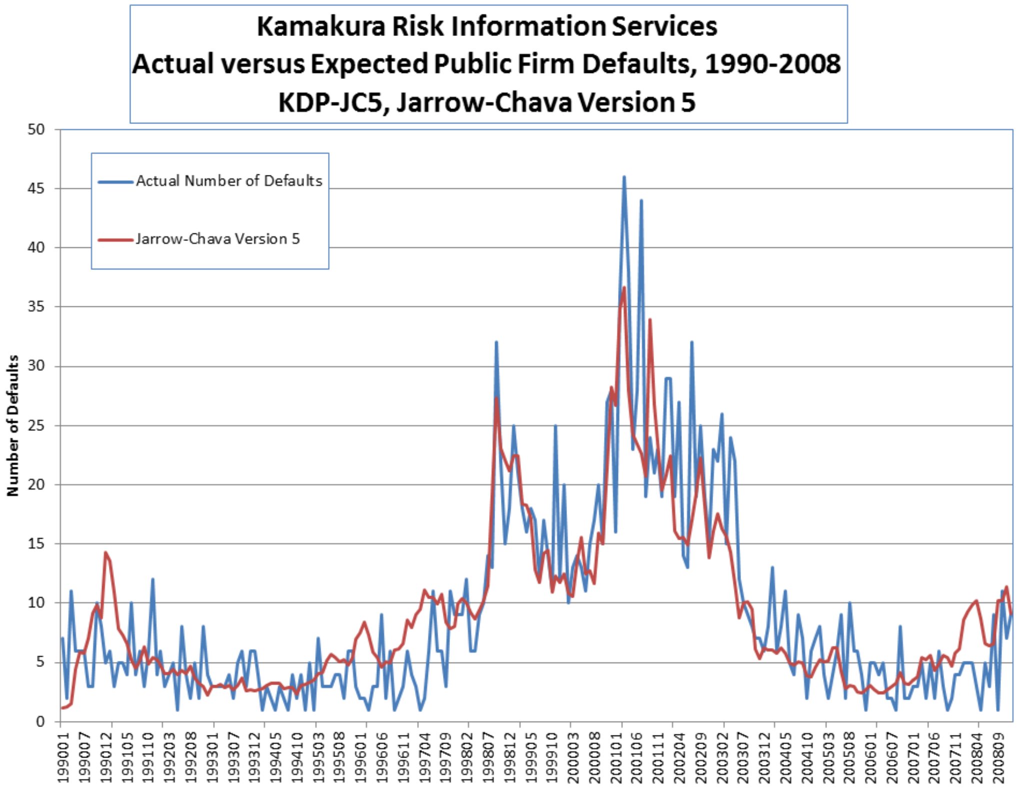
Consistency of Estimated and Actual Defaults

Falkenstein and Boral [2000] correctly emphasize the need to do more than measure the correctness of the ordinal ranking of companies by riskiness. One needs to determine whether a model is correctly "calibrated" (in the words of Falkenstein and Boral), that is, whether the model has default probabilities that are biased high or low. As noted in the example in the introduction, a naïve user of credit models can be convinced a model has superior performance just because it gives higher default probabilities for some subset of a sample. A test of consistency between actual and expected defaults is needed to see whether this difference in default probability levels is consistent with actual default experience or just an *ad hoc* adjustment or noise. Anecdotal evidence suggests that default models for public firms in common use are biased high by a factor of 3 or 4 times actual default levels.

Recent Results from North America

In this and subsequent sections, we report test results for the KRIS version 5.0 default probabilities on a sample of 1.76 million observations from North America from 1990 to 2008. We plot actual defaults, month by month, over this period versus the expected default produced by the model. Figure 9 shows the results for the Jarrow-Chava KDP-JC5 version 5 model:

Figure 9



If we run the regression

Actual defaults = A + B (Expected Defaults)

as suggested by van Deventer and Wang [2003], the adjusted R-squared of the regression is a good quantitative measure of the consistency between actual and expected defaults. In Figure 10, we see that the figure for the KDP-JC5 Jarrow Chava model is 74.30%, well above the Merton model, which explains less than 43% of the month to month variation in actual defaults.

Figure 10

Actual Defaults Measured as a Function of Predicted Defaults by theModel

Actual Defaults = A + B * Expected Defaults

Version	**Model**	**Adjusted R^2**	**Data Set**
Version 5.0	**KDP-jc5 Jarrow-Chava Model**	**74.30%**	**Version 5.0, January 1990-December 2008, Monthly**
Version 5.0	**KDP-jm5 Jarrow-Merton Hybrid**	**73.88%**	**Version 5.0, January 1990-December 2008, Monthly**
Version 5.0	**KDP-ms5 Merton Logistic Model**	**42.38%**	**Version 5.0, January 1990-December 2008, Monthly**

The Falkenstein and Boral Test

Eric Falkenstein and Andrew Boral [2000], formerly of Moody's Investors Service, 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 default probability level
- Create N "buckets" of these observations with an equal number of observations in each bucket
- Measure the default probability boundaries that define the low end and high end of each

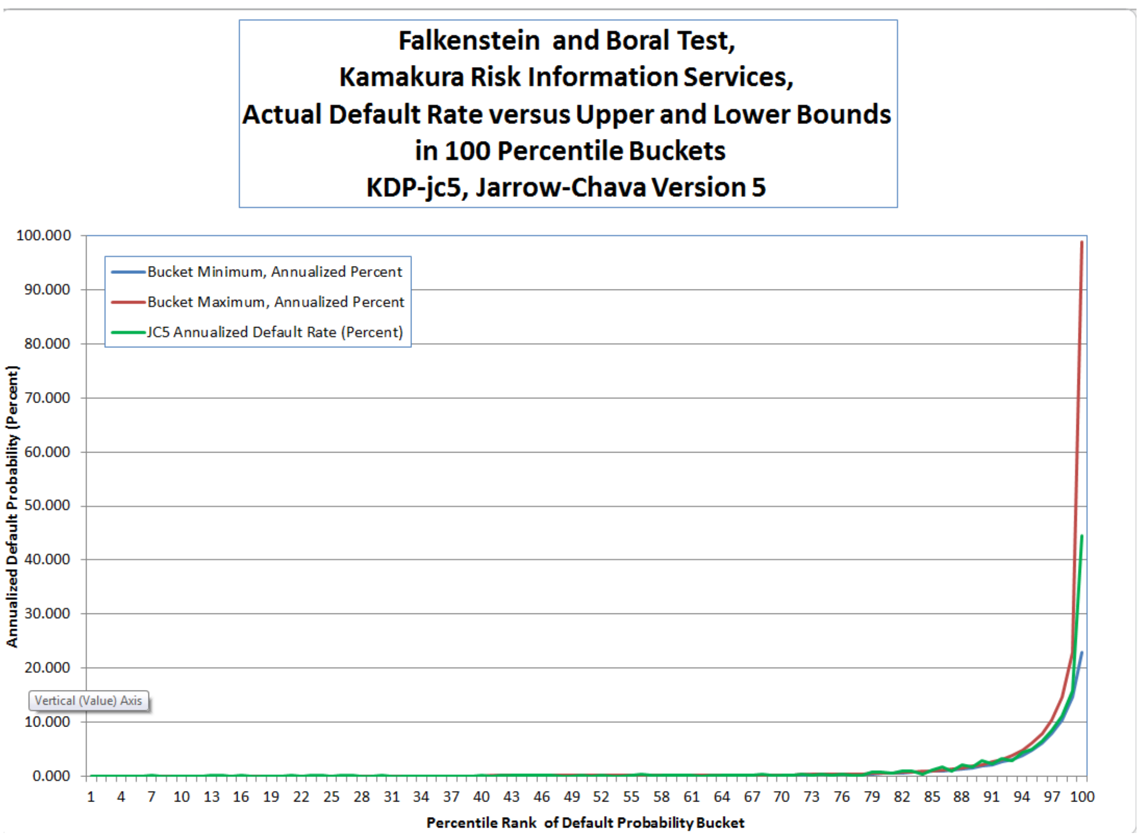
bucket

- Measure the actual rate of default in each bucket
- 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 like 5 for the reasons discussed above. 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 and final version of Basel II, Basel III, and Solvency II.

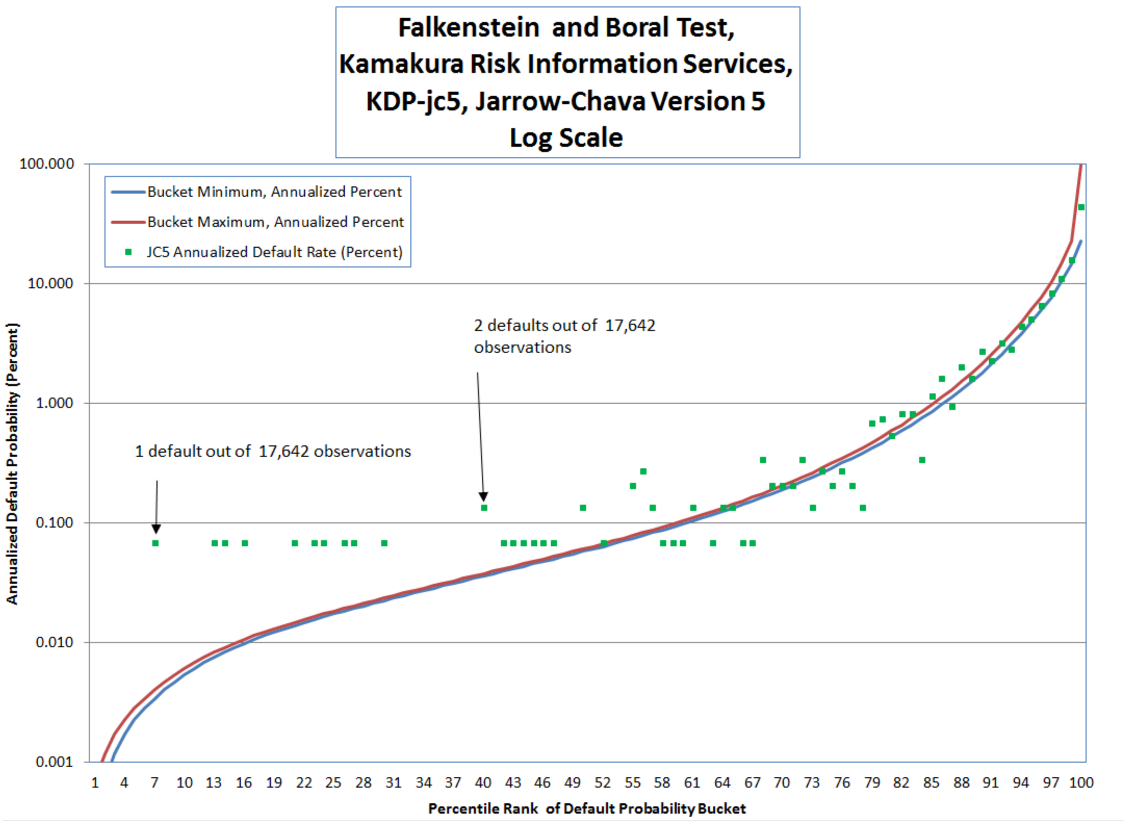
If there is an accidental or intentional bias in default probabilities, the adjusted R^2 test in the prior section will detect date-related or credit-cycle related biases. If the bias is related to the absolute level of the of the default probabilities, the Falkenstein and Boral test applies. The graph below applies the Falkenstein and Boral test using 100 time buckets for version 5.0 of Kamakura Corporation's KDP-JC5 Jarrow Chava reduced form model:

Figure 11



The top line defines the upper default probability in each of the 100 buckets. Each bucket has 17,642 observations. The lower line defines the lower default probability bound of each bucket. The kinked middle line is the actual frequency of defaults. The graph shows a very high degree of consistency in the KDP-jc modeling of actual versus expected defaults. The following graph displays the same results on a logarithmic scale.

Figure 12

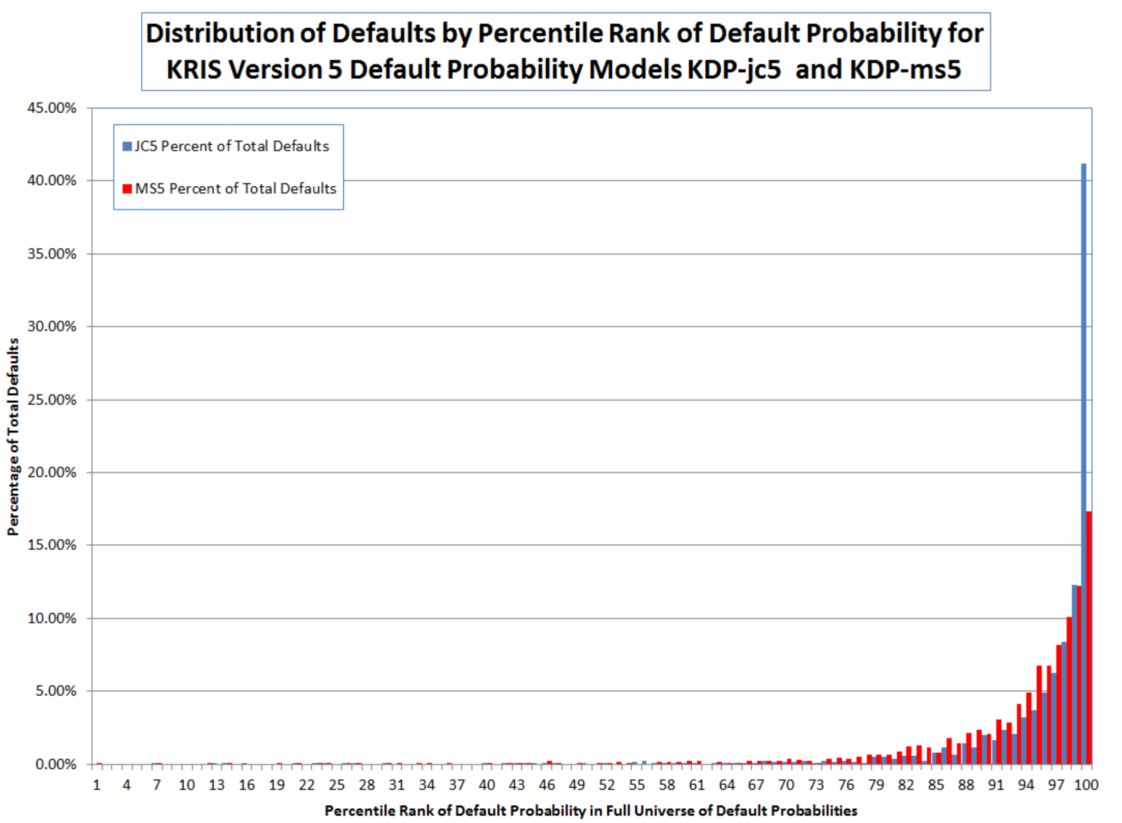


The Falkenstein and Boral test does not work well for the lower default probability buckets due to small sample size issues having to do with the discreteness of a default event. As indicated in this graph, it is impossible for the actual default frequency to fall between the upper and lower bounds of the bucket because 0 defaults out of 17,642 is below the lower bound and 1 default out of 17,642 is above the upper bound. Consequently, this test is most useful for the intermediate and higher buckets. The Jarrow-Chava version 5 model shows good consistency, when measured by default probability level, between actual and expected defaults.

Another test suggested by Eric Falkenstein[12] is to plot the distribution of actual defaults by the percentile of the default probabilities. The 1st percentile (the lowest 17,642 of the 1.76 million default probabilities in the sample) should have the lowest number of defaults and the 100th percentile should have the highest.

[12] Private conversation with one of the authors, fall 2003.

Figure 13



The above graph shows these test results for version 5 of the Merton model and Jarrow-Chava model. The Jarrow-Chava model shows a better performance because the actual default experience is more heavily concentrated in the highest default probability percentile buckets. In contrast, percentage of defaults captured by the Merton model in the 99th percentile is nearly 20 percentage points lower than the 42% captured by the Jarrow-Chava model. The difference in discrimination between the two models can be measured quantitatively using a Chi-squared test as suggested in van Deventer and Imai [2003].

Performance of Credit Models vs. Naïve Models of Risk

Knowing the absolute level of the accuracy of a credit model 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, comes as a surprise to many academics and financial services regulators so we present representative results here[13].

ROC Accuracy Ratios for Merton Model Theoretical Version versus Selected Naïve Models

As mentioned above, the 'pure theory' implementation of the Merton model has an ROC accuracy ratio over a one month horizon of 0.8259. Among the explanatory variables used in the Jarrow-Chava version 5 model, seven variables have a higher stand-alone accuracy ratio than the Merton model. Three of those ratios are summarized here:

0.8803 3 month standard deviation of stock price returns

[13] Based on 1.76 million monthly observations consisting of all listed companies in North America for which data was available from 1990-2008, Kamakura Risk Information Services data base, version 5.0.

0.8687	Excess return over S&P 500 equity index for the last year
0.8522	Percentile rank of firm's stock price on given date

This observation that the Merton model under-performs 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 modeling approach on a stand-alone basis?***

Of course, the answer should be no. This finding is another factor supporting the use of reduced form models which by construction outperform any naïve model depending on a single financial ratio.[14]

Tests of Credit Models Using Market Data

The first part of this chapter introduced reduced form credit models and a testing regime to show how well such models perform relative to the Merton model and other legacy credit models that we review in more detail in Chapter 18. Our interest in the models and how well they work is more than academic. We are looking for a comprehensive framework for enterprise risk management that creates true risk-adjusted shareholder value. We want to be able to accurately value the Jarrow-Merton put option on the value of the firm's assets that is the best comprehensive measure of integrated credit risk, market risk, liquidity risk and interest rate risk. If risk is outside of the "safety zone" discussed in earlier chapters, we want to know the answer to a critical question: what is the hedge? If we cannot answer this question, our credit risk modeling efforts are just an amusement with no practical use. In this section, we trace recent developments in using market data to test credit models and outline directions for future research.

Testing Credit Models: The Analogy with Interest Rates

In Chapters 6 through 13, we examined competing theories of movements in the risk-free term structure of interest rates. In Chapter 14, we reviewed various methods of estimating the parameters of those models. In this section, our task is similar except that the term structure models we are testing is a term structure with credit risk. We devote all of Chapter 17 to that task, but we introduce the topic here.

Market Data Test 1:

Accuracy in Fitting Observable Yield Curves and Credit Spreads

Continuing the term structure model analogy, one of the easiest and most effective market data tests of a credit model is to see how well the model fits observable yield curve data. All credit models take the existing risk free yield curve as given, so we can presume we have a smooth and perfectly fitting continuous risk free yield curve using one of the techniques of Chapter 5. Given this risk free curve, there is an obvious test of competing credit models: which model best fits observable credit spreads?

As we discuss in Chapter 17, the market convention for credit spreads is a crude approximation to the exact day count precision in credit spread estimation that we describe in that chapter. It would

[14] 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 favor of the multi-variable models presented.

be better if we could measure model performance on a set of data that represents market quotations on the basis of credit spreads that we know are "clean." There are many sources of this kind of data, ranging from interest rate swap quotations (where the credit issues are complex) to credit default swaps, a data source of increasing quality concerns because of low trading volume and risks of manipulation. Van Deventer and Imai [2003][15] present another source, the credit spreads quoted by investment bankers weekly to First Interstate Bancorp, which at the time was the seventh largest bank holding company in the United States. This data series is given in its entirety in an appendix by van Deventer and Imai.

They then propose a level playing field for testing the ability of the reduced form and Merton models to fit these credit spreads by restricting both models to their two parameter versions, with these two parameters re-estimated for each data point. The parameters were estimated in a "true to the model" fashion. For the Merton model, for instance, it is assumed that the equity of the firm is an option on the assets of the company as we discuss in Chapter 18. Van Deventer and Imai assume this assertion is true, and solve for the implied value of company assets and their volatility such that the sum of squared error in pricing the value of company equity and the two year credit spread was minimized. The two year credit spread was chosen because it was the shortest maturity of the observable credit spreads and therefore closest to the average maturity of the assets held by a bank holding company like First Interstate.

For the Jarrow model, van Deventer and Imai select the simplest version of the model in which the liquidity function discussed above is a linear function of years to maturity and in which the default intensity $\lambda(t)$ is assumed constant. These assumptions imply (see Jarrow [1999]) that the zero coupon credit spread is a linear function of years to maturity.

Van Deventer and Imai show that the Jarrow model substantially outperforms the Merton model in its ability to model credit spreads for First Interstate. The Merton model tended to substantially under price credit spreads for longer maturities, a fact noted by many market participants. The Merton model also implies very short term credit spreads will be zero for a company that is not extremely close to bankruptcy, while the Jarrow liquidity parameter means the model has enough degrees of freedom to avoid this implication. These problems are discussed extensively in David Lando's [2004] excellent book on credit modeling

In the tests we have described so far, we are essential testing the credit model equivalent of the Vasicek term structure model, i.e. we are using the pure theory and not "extending" the model to perfectly fit observable data like we do using the extended Vasicek/Hull and White term structure model. We now turn to a test that allows us to examine model performance even for models that fit the observable term structure of credit spreads perfectly.

Market Data Test 2: Tests of Hedging Performance

From a trader's perspective, a model that can't fit observable data well is a cause for caution but not a fatal flaw. The well-known "volatility smile" in the Black-Scholes options model doesn't prevent its use (even though it should), but it affects the way the parameters are estimated. After this tweaking

[15] See Chapter 7.

of the model, it provides useful insights on hedging of positions in options. How do the Jarrow and Merton models compare in this regard?

Jarrow and van Deventer [1998] present such a test, again using data from First Interstate Bancorp. They implement basic versions of the Jarrow and Merton models and assume that they are literally true. They then calculate how to hedge a one week position in First Interstate two year bonds, staying true to the theory of each model. In the Merton model, this leads to a hedge using U.S. Treasuries to hedge the interest rate risk and a short position in First Interstate common stock to hedge the credit risk. In the Jarrow model, it was assumed that the macro factors driving the default intensities $\lambda(t)$ were interest rates and the S&P 500 index, so U.S. Treasuries and the S&P 500 futures contracts were used to hedge.

Jarrow and van Deventer report that the hedging error using the Jarrow model was on average only 50% as large as that of using the Merton model over the entire First Interstate sample.

More importantly, they report the surprising conclusion that the Merton model fails a naïve model test like those we examined in Chapter 19: a simple duration model produced better hedging results than the Merton model. This surprised Jarrow and van Deventer and many other observers, so Jarrow and van Deventer turned to the next test to understand the reasons for this poor performance.

Market Data Test 3: Consistency of Model Implications with Model Performance

Jarrow and van Deventer investigated the poor hedging performance of the Merton model in detail to understand why the hedging performance of the model was so poor. They ran a linear regression between hedging error and the amount of Merton hedges (in Treasuries and common stock of First Interstate). The results of this standard diagnostic test were again a shock—the results showed that instead of selling First Interstate common stock short to hedge, performance would have been better if instead one had taken a long position in common stock. While the authors don't recommend this as a hedging strategy to anyone, it illustrates what Jarrow and van Deventer found—the movements of stock prices and credit spreads are much more complex than the Merton model postulates. As one can confirm from the first derivatives of the value of risky debt given in Chapter 18, the Merton model says that when the value of company assets rises, stock prices should rise and credit spreads should fall. The opposite should happen when the value of company assets falls, and neither should change if the value of company assets is unchanged.

If the Merton model is literally true, this relationship between stock prices and credit spreads should hold 100% of the time. Van Deventer and Imai report the results of Jarrow and van Deventer's examination of the First Interstate credit spread and stock price data—only 40-43% of the movements in credit spreads (at 2 years, 3, years, 5 years, 7 years and 10 years) were consistent with the Merton model. This is less than one could achieve with a credit model that consists of using a coin flip to predict directions in credit spreads. This result explains why going short the common stock of First Interstate produced such poor results. Van Deventer and Imai report the results of similar tests involving more than 20,000 observations on credit spreads of bonds issued by companies like Enron, Bank One, Exxon, Lucent Technologies, Merrill Lynch and others. The results consistently show that only about 50% of movements in stock prices and credit spreads are consistent with the Merton model of risky debt. These results have been confirmed by many others and are presented more formally in a recent paper by Jarrow, van Deventer and Wang [2003].[16]

Why does this surprising result come about? And to which sectors of the credit spectrum do the findings apply? First Interstate had credit ratings which ranged from AA to BBB over the period studied, so it was firmly in the investment grade range of credits. Many have speculated that the Merton model would show a higher degree of consistency on lower quality credits. This research is being done intensively at this writing and we look forward to the concrete results of that research.

For the time being, we speculate that the reasons for this phenomenon are as follows:

- Many other factors than the potential loss from default affect credit spreads, as the academic studies outlined in Chapter 17 have found. This is consistent with the Jarrow specification of a general "liquidity" function that can itself be a random function of macro factors
- The Merton assumption of a constant, unchanging amount of debt outstanding, regardless of the value of company assets, may be too strong. Debt covenants and management's desire for self-preservation may lead a company to overcompensate, liquidating assets to pay down debt when times are bad.
- The Merton model of company assets trading in frictionless efficient markets may miss the changing liquidity of company assets, and this liquidity of company assets may well be affected by macro-factors.

What are the implications of the First Interstate findings for valuation of the Jarrow-Merton put option as the best indicator of company risk? In spite of the strong conceptual links with the Merton framework, the Merton model itself is clearly too simple to value the Jarrow-Merton put option with enough accuracy to be useful. We explore these problems in greater detail in our credit model attic, Chapter 18.

What are the implications of the First Interstate findings for bond traders? They imply that even a trader with perfect foresight of changes in stock price one week ahead would have lost money more than 50% of the time if they used the Merton model to decide whether to go long or short the bonds of the same company. Note that this conclusion applies to the valuation of risky debt using the Merton model—it would undoubtedly be even stronger if one were using Merton default probabilities because they involve much more uncertainty in their estimation, as we shall see in Chapter 18.

What are the implications of the First Interstate findings for the Jarrow model? There are no such implications, because the Jarrow model per se does not specify the directional link between company stock prices and company debt prices. This specification is left to the analyst who uses the Jarrow model and derives the best fitting relationship from the data.

[16] The treasury department of a top ten U.S. bank holding company reported, much to its surprise, that their own "new issue" credit spreads showed the same low level of consistency with the Merton model as that reported for First Interstate.

Market Data Test 4: Comparing Performance with Credit Spreads and Credit Default Swap Prices

There is a fourth set of market data tests that one can perform which is the subject of a very important segment of current research: comparing credit model performance with credit spreads and credit default swap quotations as predictors of default. We address these issues in Chapter 17 in light of recent concerns about manipulation of the Libor market and similar concerns about insider trading, market manipulation and low levels of trading activity in the credit default swap market.

Appendix A: Converting Default Intensities to Discrete Default Probabilities

The reduced form models are attractive because they allow for default or bankruptcy to occur at any instant over a long time frame. Reduced form default probabilities can be converted

- From continuous to monthly, quarterly or annual
- From monthly to continuous or annual
- From annual to monthly or continuous

This section shows how to make these changes in periodicity on the simplifying assumption that the default probability is constant over time. When the default probability is changing over time, and even when it is changing randomly over time, these time conversions can still be made although the formulas are slightly more complicated.

Converting Monthly Default Probabilities to Annual Default Probabilities

In estimating reduced form and hybrid credit models from a historical data base of defaults, the analyst fits a logistic regression to a historical default data base that tracks bankruptcies and the explanatory variables on a monthly, quarterly or annual basis. The default probabilities obtained will have the periodicity of the data (i.e. monthly, quarterly or annual). We illustrate conversion to a different maturity by assuming the estimation is done on monthly data. The probability that is produced by the logistic regression, P, is the probability that a particular company will go bankrupt in that month. By definition, P is a monthly probability of default. To convert P to an annual basis, assuming P is constant, we know that

$$\text{Probability of no bankruptcy in a year} = (1-P)^{12}$$

Therefore the probability of going bankrupt in the next year is

$$\text{Annual probability of bankruptcy} = 1 - (1-P)^{12}$$

The latter equation is used to convert monthly logistic regression monthly probabilities to annual default probabilities.

Converting Annual Default Probabilities to Monthly Default Probabilities

In a similar way, we can convert the annual probability of default, say A, to a monthly default probability by solving the equation

$$A=1-(1-P)^{12}$$

for the monthly default probability given that we know the value of A.

$$\text{Monthly probability of default } P = 1-(1-A)^{1/12}$$

Converting Continuous Instantaneous Probabilities of Default to an Annual Default Probability or Monthly Default Probability

In the *Kamakura Technical Guide* for reduced form credit models by Robert Jarrow, Professor Jarrow writes briefly about the formula for converting the instantaneous default intensity $\lambda(t)$ (the probability of default during the instant time t) to a default probability that could be monthly, quarterly or annual:

"For subsequent usage the term structure of yearly default probabilities (under the risk neutral measure) can be computed via

$$Q_t(\tau \le T) = 1 - Q_t(\tau > T) \quad \text{where} \quad Q_t(\tau > T) = E_t\left(e^{-\int_t^T \lambda(u)du}\right)$$

and $Q_t(\cdot)$ is the time t conditional probability."

In this section we interpret Professor Jarrow's formula. $Q_t(\tau \le T)$ is the probability at time t that bankruptcy (which occurs at time τ) happens before time T . Likewise, $Q_t(\tau > T)$ is the probability that bankruptcy occurs after time T. Professor Jarrow's formula simply says that the probability that bankruptcy occurs between now (time t) and time T is one minus the probability that bankruptcy occurs after time T. He then gives a formula for the probability that bankruptcy occurs after time T:

$$Q_t(\tau > T) = E_t\left(e^{-\int_t^T \lambda(u)du}\right)$$

E_t is the expected value of the quantity in parentheses as of time t. Because we are using the simplest version of the Jarrow model in this appendix, $\lambda(t)$ is constant so we can simplify the expression for $Q_t(\tau>T)$:

$$Q_t(\tau > T) = e^{-\lambda(T-t)}$$

Converting Continuous Default Probability to an Annual Default Probability

The annual default probability A when λ is constant and T-t=1 (one year) is

$$A = 1 - Q_t(\tau > T) = 1 - e^{-\lambda}$$

Converting Continuous Default Probability to a Monthly Default Probability

The monthly default probability M when λ is constant and T-t=1/12 (one twelfth of a year) is

$$M = 1 - Q_t(\tau > T) = 1 - e^{-\lambda(1/12)}$$

Converting an Annual KDP to a Continuous Default Intensity

When the annual default probability A (the KDP is known) and we want to solve for λ , we solve the equation above for λ as a function of A:

$$\lambda = -\ln(1 - A)$$

Converting a Monthly Default Probability to a Continuous Default Intensity

In a similar way, we can convert a monthly default probability to a continuous default intensity by reversing the formula for M. This is a calculation we would do if we had a monthly default probability from a logistic regression and we wanted to calculate λ from the monthly default probability:

$$\lambda = -12\ln(1 - M)$$