Credit Risk Models, Capital Standards and Self-Regulation

Robert A. Jones*
Simon Fraser University

December 2000

I. Introduction

Credit risk models seek to characterize, quantify, forecast and evaluate the consequence of non-fulfillment of contractual obligations. In banking contexts, this typically refers to default on loans (including bonds) and payouts under credit guarantees. However financial innovations of the past two decades extends it to counterparty risk on derivatives, cash flows on stratified securitizations, and contingent payments on credit derivatives. Intermediaries that trade or make markets in such instruments are additionally exposed, not just to actual events of default, but also to fluctuation in the market’s perception of the risk-adjusted likelihood of such events.

Why should these models be of interest to regulators? First, regulators are concerned with the solvency of banks. Adverse credit outcomes by bank customers/counterparties can force leveraged banks to default on deposits. This costs taxpayers under government deposit insurance or bailout policies, and may also inflict negative real externalities through systemic disruption of the payment system. (Your insurance company and your neighbours both appreciate your having smoke detectors.) Any tools that help assess, aggregate and monitor credit exposure are potentially usable by regulators in their supervisory role and in the application of capital standards.

Why should such models be of interest to bankers? Taking their objective to be shareholder value maximization, the application of such models, by measuring and aggregating credit exposure, has a role to play in avoiding costs of financial distress through risk management and capital structure decisions, in performance measurement, and in pricing across related credit products. Credit exposure to a given customer can take a variety of forms, all considered

---

*Economics Department, Simon Fraser University, Canada V5A 1S6. tel: (604) 291-3367; email: rjones@sfu.ca

1 The notional amount of derivatives in US deposit insured commercial bank portfolios alone was 38.3 trillion as of third quarter 2000; notional amount of credit derivatives 379 billion (OCC Bank Derivatives Report, Comptroller of the Currency, 2000).
by him (or the ambitious relationship manager) simultaneously: fixed rate bullet (constant balance) loan; floating rate bullet loan; program guaranteeing the customer’s commercial paper; floating rate loan with spread reset according to credit quality; unsecured fixed/floating interest rate swap with the customer; third party credit default swap based on that customer; revolving credit line; standard credit line (maximum cumulative drawdowns independent of prepayments); option on any of these (loan commitment). Clearly an analytically consistent framework is needed to avoid internal arbitrage and to ensure the same compensation for bearing the same risk regardless of contractual form.

Why should such models be of interest to academics? First, better understanding the process of contract failure contributes to the positive economics objective of explaining endogenous institutional arrangements (ones that, like banks, would have no role in the hypothetical Arrow-Debreu world of no transaction, commitment, information or enforcement costs). Second, the actuality and perception of credit risk has efficiency and welfare implications via its impact on the allocation of real investment (through either relative pricing or credit rationing). Third, the extent of systematic factors in causing correlation in default, or comovements in market risk premia for bearing credit exposure, connects financial market phenomena with macroeconomic stability issues.

The increased use of senior/subordinated securitization and of credit derivatives will increasingly muddy the connection between an intermediary’s balance sheet and its actual credit exposure—much as already occurred with conventional derivatives and market (price) risk. These devices permit cost-effective transfer of credit exposure to entities more willing and able to bear it, a positive and desirable gain from trade. But, symmetrically, they permit large speculative credit exposures to be taken with little or no use of current financial capital. Or, in the case of securitization of portions of a bank’s loan portfolio, leave the illusion that most of the risk has been shed while in fact it is still there through retained junior/residual tranches or through contractual credit guarantees. Standard accounting procedures are ill-suited to pick up and aggregate such commitments.

Credit derivatives on the books of US deposit-insured commercial banks were of 379 billion notional amount as of third quarter of this year, growing at a 30%+ annual rate.\(^2\) In contrast with most conventional derivatives, credit derivatives typically have highly skewed payoffs: They entail small assured payments one way in return for a low probability large size payment the other (when default events occur). Thus the low notional amount compared to other derivatives likely understates their potential to cause mischief. Current ‘one size fits all’ BIS rules provide inadequate and frequently perverse measuring tools, with excessive opportunities for regulatory capital arbitrage. Well-thought-out credit risk models have the potential to do better here.

In this talk, I first review recent academic approaches to modelling credit risk, then try

to link them with the regulatory issue of capital standards. A central question is the extent to which the Modigliani-Miller capital structure irrelevance propositions apply to banks. We will then consider what might be plausible research priorities in this still developing area.

II. Credit Risk Models

1. Academic developments

There is a considerable and rapidly growing literature in finance applying modern contingent claims analysis (‘option pricing’) to modelling default and credit risk. A portion is listed in the attached bibliography. The majority of it is theoretical, struggling for empirically relevant, yet computationally tractable, ways of looking at the problem. The approach is inherently dynamic. It thus has the potential to characterize not just the likelihood of default at future times from the perspective of today, but also how those likelihoods might fluctuate over time. The latter is necessary for measuring risk in a mark-to-market or mark-to-model sense, in the absence of default, and also necessary for devising hedging strategies.

Data is a problem. As reported in the BIS 1999 study, *Credit Risk Modelling: Current Practices and Applications*, banks and researchers alike find data limitations to be a key impediment to the design and implementation of credit risk models. Credit instruments are not generally marked to market, and long-term historical data within banks is typically in the form of aggregate chargeoff rates within various business lines. Publicly available data on default-risky securities consists mainly of price/default histories of traded corporate bonds, and rating transition/loss experience provided by major credit rating agencies. As a result, most published empirical work is based on such data and theoretical work is oriented towards explaining it.

Traded corporate bond price data display several stylized characteristics that useful credit risk models must accommodate. First, default-free interest rates fluctuate over time. This indicates that multi-factor risk models, with their attendant analytical and empirical difficulties, are likely required. Second, ‘spreads’ between contractual yields to maturity on default-free and risky bonds are positive even at short maturities. Third, these spreads vary with maturity, but not always monotonically. Fourth, these spreads fluctuate over time. Fifth, bond prices appear to ‘jump’ at times even without default (e.g., summer ’98 junk bond market meltdown). Pure diffusion models of credit quality evolution may have difficulty capturing this. Sixth, there is noticeable positive correlation in changes in spreads across issuers. Whether this reflects correlated changes in the objective probability of default due to common macroeconomic factors, or whether it reflects changes in a market required risk premium (and hence risk-neutral probability of default) is unclear. And finally, actual losses due to default appear to be ‘clumped’ in moderately short and infrequent historical windows (e.g., ’91-92 recession). As we shall see, the theoretical valuation literature has developed in ways that attempt to
accomodate these facts. However empirical work linking these models to data is in relative infancy.

The principle that, in equilibrium, no riskless arbitrage opportunities exist—either static or dynamic—imposes a useful discipline on modelling security values: It relates the infinite number of potential securities whose risk derives from the same underlying random factors. The arbitrage-free value may be equally thought of as the price that would prevail in financial market equilibrium, as the forecast cost of replicating or hedging a security through trading in others driven by the same underlying factors, or the price which if paid would provide the same premium for bearing equivalent risk as that offered by alternative securities.

The analytical implication of this principle is that fair market values should equal the expected discounted prospective cash flows of a security—with two provisos. First, the discounting must be done using the path followed by short-term default-free interest rates (whose future course is ex ante uncertain). Second, the probabilities used in computing the expected values must generally differ from the objective odds of the payoff-relevant events occurring. It is this difference between the objective or empirical odds that would be revealed by historical statistical analysis, and the risk-adjusted or risk-neutral odds implied by asset prices, that embodies the reality that investors are averse to non-diversifiable risk. Recognizing this distinction is important not just for valuing derivatives contracts, but also for appraising lending/investment opportunities, for determining effect of capital structure decisions on share value, and for extracting ‘market expectations’ from observed market prices.

Arbitrage-free credit risk models are classified as either structural or reduced form according to whether the underlying state variable determining default is the value of the obligor’s assets (value of the firm) or whether it is something else more specifically indicating the likelihood of default (credit quality).

**Structural approach**

The initial application of arbitrage-free valuation principles to valuing corporate debt was by Black and Scholes (1973) and Merton (1974). They took the value of the firm’s assets, following a lognormal diffusion, to be the exogenous state variable, considered bonds that could only default at maturity, and assumed strict priority rules prevailed in bankruptcy. The value of the defaultable bond was the value of an identical default-free bond minus a call option on the firm’s assets, expiring at bond maturity. Geske and Johnson (1974, 1984) extend this to bonds that could default on any coupon payment date, characterizing them as compound options.

Recognizing that other obligations of a firm and/or debt covenants might induce bankruptcy at times unconnected with the contract being valued, other authors characterize the arrival of default as the first passage time of firm value through an exogenously specified barrier. Included here is the work of Black and Cox (1976), Brennan and Schwartz (1980) and Longstaff
and Schwartz (1995). All these models give rise to a term structure of credit spreads, linked to the current value of a firm relative to its default barrier. The more recent recognize that recoveries in bankruptcy do not necessarily follow strict contractual priority rules.

But what determines where the default barrier lies? Constant barriers are more analytically tractable but difficult to justify. Work by Fisher, Heinkel and Zechner (1986), Jones (1995) and Leland and Toft (1996) portray default as rational exercise of an American option to cease putting additional cash into a limited liability enterprise. Again, the firm’s assets are what shareholders forfeit in default, and is the central state variable. These models endogenize bankruptcy, with interesting consequences for credit rationing and loan contract design. Rational default boundaries are revealed to be non-constant in the presence of finite-maturity debt; numerical methods must be used for solution.

The above treatments presume that anything less than full contractual payment forces bankruptcy or liquidation of collateral. But in the presence of liquidation costs, the commitment by a lender to do this is not necessarily credible nor desirable. Works by Anderson and Sundaresan (1995) and by Theunissen (1998) add a game-theoretic component to the problem, permitting borrowers to offer partial payments that just suffice to make the lender postpone foreclosure. Theunissen shows that this possibility mitigates ex ante credit rationing, and can lead to the unexpected phenomenon of partial default when collateral value gets too high (as well as too low)—a sort of ‘unilateral renegotiation’ downward of loan rates inappropriate for the now-safe borrower.

All of these frameworks generate a term structure of contractual credit spreads facing a particular borrower. All assume that the random process followed by borrower assets is a diffusion: value over time moves continuously. As a result, the theoretical credit spread for zero maturity bonds must be zero, since the odds of diffusing a discrete distance in short time is negligible.

Yet we observe positive credit spreads for low-grade borrowers at even short maturities. To reconcile this with the theory, there must be the possibility of significant surprises, news that causes sudden revision in the perceived value of borrower assets. Building on work of Merton (1976), Zhou (1997) accommodates this aspect of the data through adding Poisson arriving jumps to the asset value diffusion. These must be log-normally distributed if analytical tractability is to be maintained.

The use of firm asset value as the central state variable is appealing for several reasons. First, it should approximately equal the total market value of a firm’s liabilities, including equity, and is thus potentially observable. Second, the theoretical link between it and default is reasonably transparent and intuitive. Third, its level is readily connected with creditor recoveries in bankruptcy. And finally, one can show that the risk-adjustment of probabilities needed to obtain fair market values are those that make the firm’s assets just yield the risk-free interest rate. For many pricing and hedging applications, this lets one avoid specifying...
(risking mis-specifying) and estimating many details of the return process.

However it must be remembered that the probabilities of default so-obtained will *not* be the objective probabilities of concern to regulators, or give accurate Value-at-Risk to managers. For that, one requires a statistically correct picture of the *objective* process followed by debtor assets, to be combined with the *risk-neutral* process that forms the basis for rational default.

Further, virtually all structural approaches require knowledge of the full range of the firm’s obligations, and that these are either fixed or changing in a perfectly foreseen way over time, to identify where the default boundary plausibly lies. This substantial information requirement and restriction suggests this approach might best apply only for very simple corporate structures (e.g., real estate development and holding companies).

**Reduced form approach**

In contrast to the structural approach, the so-called reduced form approach takes ‘credit quality’, somehow construed, to be the underlying state variable. Default is always a surprise and can occur at any time. It is analytically portrayed as a Poisson process: the probability of occurrence over a short interval of length $dt$ is $\lambda(t)\ dt$, where $\lambda(t)$ is the arrival intensity at time $t$. $\lambda$ may be constant, follow a deterministic time path, or be randomly varying.

Again, considerations of market equilibrium and lack of arbitrage imply that fair security values equal expected discounted cash flows up to default. Objective and risk-neutral default intensities need not coincide if this ‘jump risk’ is not completely diversifiable (the only restriction being that the probabilities agree on what events are ‘impossible’). Explicit assumption must specify what will be recovered when default occurs.

The approach has appealing features. First, since default can occur arbitrarily soon, positive observed credit spreads at all points on the yield curve can be accommodated. Second, it can be analytically and computationally very tractable: When recoveries are either deterministic (e.g., fixed fraction of par) or a fixed fraction of the contract value immediately before default, the effect on value is the same as if the riskless discount rate had been raised. Third, it lets one subsume in the specification of $\lambda$ all the complexities and non-observability of the actual financial structure of firms. Duffie and Lando (1999) give a formal portrayal of the reduced form approach as a structural approach with incomplete accounting information.

The literature in this area is rather more recent. Hull (1989, 1995) assumes a constant default intensity, with credit state thus a binary variable (default or non-default) to examine counterparty risk in the derivatives market. For bond yields, this implies a credit spread that is constant over time and across maturities for a given issuer. Lando (1994), Jarrow and Turnbull (1995), Duffie and Huang (1995), and Duffie and Singleton (1995, 1997) use a deterministically time-varying intensity to accommodate maturity-dependent credit spreads when looking at term structure, counterparty risk, and options on defaultable securities. This permitted accurate ‘fit’ to a cross-section of market prices at a single point in time.
But credit spreads evolve in a way that is not perfectly foreseen. One way to handle this is to assume a finite set of credit states with Markov switching between them. This avenue is developed in the credit ratings-based model of Jarrow, Lando and Turnbull (1997). Switches between ratings (including default, which is absorbing) arrive as Poisson processes with different constant intensities depending on the state currently occupied. The model could then be calibrated to historical default and ratings-transition data from Moody’s. Recovery rates were specified as averages recorded by the rating agency. A one-parameter ‘adjustment’ distinguishes between risk-neutral and objective probabilities.

The approach is promising in that it permits use of actual default experience in determining default probabilities over various time horizons, and conditions prices on an observable state variable (agency credit rating). However it implies identical term structure of credit spreads for firms with the same rating, and near perfect comovement of ratings and credit spreads. Bond yield data accord with these implications at best very crudely.

The relevant empirical question here is whether agency or other credit ratings confer any useful information about the likelihood of default beyond what is in market prices. In situations where no market price data is available, it is of course better than nothing. But it would be worrisome if substantial credit decisions and mark-to-model valuations of many institutions were driven largely by the simplistic and, of necessity, standardized computations a few (junior?) analysts without financial stake in the outcome.

To accommodate less ‘notchy’ fluctuations in market credit spreads, Duffie (1998) and Duffie and Lando (1999) allow each firm’s instantaneous intensity of default \( \lambda \) to follow a mean-reverting diffusion—a ‘Cox process’. Lando (1998) nicely augments a ratings-transition model with one having a continuum of credit qualities within each category, each following a diffusion. This approach permits more accurate representation of actual market prices and fluctuations at the expense of making credit quality a partially latent or unobserved variable.

Two recent papers suggest frameworks that capture both the normal continuous fluctuation of credit spreads and their relatively infrequent sudden changes in the absence default. Duffie and Garleanu (1999) take the instantaneous Poisson default intensity as state variable, and assume it follows its own mixed jump-diffusion process. When the diffusion is affine, as in mean-reverting square-root diffusion, and the jump sizes are exponentially distributed, a quite tractable model emerges with closed form expressions for discount bond values.

My own work (Jones, 2000) assumes the existence of an unobserved scalar credit quality variable which follows a mixed jump-diffusion. Default is associated with it crossing an exogenously given barrier, zero, either by jump or through diffusion. Solution for security and contract values is entirely by numerical methods, so the specification of the diffusion, the jump intensity and the jump size distribution as a function of credit quality can be whatever you choose. Credit quality, given a parametrization of these components, is inferred from whatever market price (e.g., fair terms on a bullet loan) can be observed or conjectured. With
Peter Chau at Wells Fargo Bank, we have fit the model to a sample of 41,000 monthly bond price observations on 737 US firms over the period 1993-97. Agency credit ratings ranged from B to AAA; maturities from a few months to 27 years. Results are encouraging. With the four estimated model parameters, assumed constant across firms and time, and credit quality varying across firms and time, bond yield prediction residuals were less than 5 basis points for 53% of the observations and less than 25 basis points 95% of the time. Focussed as it was on market pricing, however, it must be emphasized that the parameter estimates were exclusively of the risk-neutral process. Joint estimation both risk-neutral and objective parameters requires a longer data set spanning periods in which defaults actually occurred.

Reduced form approaches are basically ways of characterizing the probability distribution of time to insolvency, conditional on current credit state, and how that distribution might evolve over time. Assuming that default occurs simultaneously on all obligations, they can be readily applied to problems of counterparty default risk (with additional factors capturing movement in the default-free value of the derivative), VaR of credit guarantees, cost of loan commitments, credit derivatives on a single security, and so on. Further, recognizing that drawdowns on a credit line rationally fluctuates with the borrower’s credit state (which influences his cost of borrowing elsewhere), they can be used to value and quantify credit exposure when this optionality is present. Jones (2000), for example, shows that the profitability of a line can fall as contractual lending spreads rise when this is taken into account.

**Default correlation**

Much of the impetus for recent work comes from the increased securitization of bank loan portfolios. To accommodate diverse risk-attitudes and expectations of potential buyers, minimize temptation to dump just the ‘lemons’ into the pool, and maintain bank monitoring incentives, claims on the pool are typically stratified into senior, junior and residual tranches; the bank retains those portions taking the first default hits. The problem then is what are theoretical fair values for the different tranches, what is the distribution of credit loss associated with each, and how do these vary with the terms of the securitization? The answers, together with the projected cash from the sale, and regulatory capital relief if that is a consideration, determine whether the bank goes ahead.

This is a tough problem. The answers should be of interest to regulators (to determine what, if any, capital relief should be granted), to rating agencies (who may be commissioned to rate the tranches), to investors (who must decide whether it is a better deal than straight bonds), and of course to the bank. We note here that the problem is similar to that of determining the value and loss distribution of a bank itself: A bank is, in part, one big securitization.

The answers hinge critically on the correlation of default across pool assets. On the one hand, if defaults are independent and the pool sufficiently diversified, total default losses are
known with virtual certainty: senior tranches can be treated as default-free bonds and the most junior as likely worthless. On the other hand, if defaults are perfectly correlated (all or none default) and recoveries are zero, then all tranches bear identical default risk, and should be valued the same. Reality lies somewhere between.

The problem has been approached in a variety of ways. In the context of structural models, one can assume correlation between the diffusions governing firm values. This was used by Wells Fargo for real estate loan securitizations, and formed the basis for KMV’s loan valuation software. In the context of reduced form models, Duffie and Singleton (1998) develop a notion of correlated jumps (default) and provide computationally efficient methods for simulating them. Duffie and Garleanu (1999) postulate correlation in the diffusion of default jump intensities. Given the complex payout structure of securitizations and the diversity of assets within the pool, Monte Carlo simulation is typically needed to estimate value of individual tranches.

2. Further remarks

- probability of bank failure depends on dynamic asset management policy followed; portfolios are not static
- mixing and confusing risk-neutral vs objective probabilities can give misleading price, (long-term) VaR projections, and assessments of counterparty risk
- optionality in credit products such as loan commitments and credit lines is important
- perils of extrapolating from limited-experience recent data to likelihood of meltdown scenarios: case of a portfolio credit default swap

III. Capital Standards and Self-Regulation

The intent of this section is to raise issues about the applicability of capital structure irrelevance to banks. The hypothesis advanced is that higher capital requirements might plausibly impose negligible costs on shareholders, and thus be the attractive policy for bank regulation.

1. Confusions about capital

The term capital is used to refer to a variety of very different things in banking, with great opportunities for confusion as a result: physical capital, financial capital, equity capital, regulatory capital, ‘economic’ capital. Physical capital are real goods that takes real resources to produce. Financial capital is the ‘pot of money’ or other assets, or command over purchasing power, that one has available to transfer to others for purchase of goods, services, securities. Equity capital is a type of liability, a residual contingent claim, reflecting the contractual terms under which investors advanced funds to a firm. Regulatory capital is a government creation,
measuring the extent of bank liabilities with contractual terms both subordinate to selected ‘senior’ claimants (e.g., depositors, deposit insurer, employees, government) and sufficiently flexible to preclude being a trigger for insolvency. ‘Economic’ capital is a concept closely related to Value at Risk, intended to aid investment and allocation decisions within a firm, or serve as a basis for performance measurement.\(^3\) It is a quantile of the return distribution on a risky position or activity—the financial cushion that would cover losses with specified probability.

The point here is that only physical capital carries a social opportunity cost through foregone consumption or diversion from alternative productive use. Regulatory and ‘economic’ capital are artificial constructs not directly relevant for welfare analysis. The amount of equity capital, on the other hand, reflects the division of a bank’s liabilities between contracts of debt form (deposits, CP, bonds) and contingent form (preferred shares, common stock, deferable maturity bonds). In the aggregate, varying the ratio of one contractual form to the other does not alter total savings available for consumption loans or investment. It is thus not obvious that there is an efficiency reason to wish banks to be as highly leveraged as they are.

2. Do Modigliani-Miller propositions apply to banks?

In a 1995 symposium on the role of capital in banking, Merton Miller gave a paper entitled “Do the M&M propositions apply to banks?” The abstract reads “Yes and no”. My reading is that he really wanted to say Yes, but, from experience at an earlier conference with bankers, thought it more gracious to leave some uncertainty. The proposition to which he refers is that, in an idealized world, the value of a firm (bank), is insensitive to its degree of leverage.\(^4\)

If true, this implies that higher capital requirements should be a matter of indifference to bank shareholders. Yes, the average cost of equity capital (expressed as market required expected rate of return) would be higher than the average cost of debt or deposit finance, if the bank’s assets display positive systematic risk. But the marginal cost of additional equity, when account is taken of both the reduction in leverage on equity and lower default risk on debt, will reflect the use to which the additional funds will be put. With no leakages, the total value of the bank’s liabilities must equal the total value of its assets. If the only social concern is with negative externalities inflicted by bank failures, higher capital requirement seems the natural policy solution.

There is a large academic literature in finance identifying deviations from the idealized world which cause the proposition to fail: differential tax treatment of interest and dividends; bankruptcy costs; high transaction costs of issuing equity relative to debt; information asymmetries that raise adverse selection, moral hazard and signalling possibilities. Additionally for banks, the cost of deposits may be implicitly subsidized by deposit insurance, access to

\(^3\)See, for example, Stoughton and Zechner (1999) for discussion and further references.

\(^4\)Miller and Modigliani (1958). See also Stiglitz (1969, 1974).
central bank lending, or effectively reduced by the liquidity services that are a joint product of bank deposits.

Do these considerations suggest higher capital requirements would place an undue burden on banks? Less leverage implies lower expected financial distress costs, a benefit to shareholders. It is difficult to see how retaining more earnings incurs high transaction costs. If capital ratios are raised to meet higher legal requirements, any normally adverse signal implication of lower leverage or of temporarily reduced dividends should not apply. Any moral hazard premium on remaining debt financing should shrink.

What remains are corporate income tax incentives for leverage and reduced claim on possible deposit insurance subsidy. In the Canadian case, there appears little evidence that deposit insurance is much of a subsidy. Rather, its beneficial role, if any, is likely more in maintaining the favourable no-banking-panic equilibrium in the sense of Diamond and Dybvig. And if there were in fact a subsidy, it is difficult to see why policymakers should be concerned with maintaining it. Regarding tax incentives, these are again less strong in the Canadian case by virtue of the tax-free flow of dividends between Canadian corporations and dividend tax credit at the personal level. The ‘capital charges’ of BIS standards do not appear to be charges at all, but rather a misinterpretation by banks and regulators alike of principles of corporate finance.

It is interesting to note that capital ratios have historically been much higher than they are now. In the US, the capital/asset ratio of banks averaged more than 50% in 1840, trending downward to about 20% at the turn of the century, and the 5-10% range of 1980 to the present.

It may be useful to explore what policies would encourage the functioning of the M&M propositions, thereby promoting self-regulation of the banking sector through the discipline of market forces.

IV. Priorities for Research

- empirical work that recovers both objective and risk-neutral distributions
- valuation model and empirical work addressing the problem of correlation of default within portfolios
- measuring the actual marginal cost of equity capital to banks, and determining the extent to which M&M propositions apply

\footnote{If the marginal investor is either tax-exempt or foreign, of course, the usual argument applies full strength.}

\footnote{Berger, Herring and Szego (1995).}
References and Related Literature


