Modelling default distributions

Joe Pimbley looks at the use of models by Gauss and Vasicek in the analysis of distributions of default, and at how a combination of the two can add further detail to the picture they give

Pricing and risk analysis for debt securities backed by portfolios of assets such as receivables, bonds and loans require the assumption of a default distribution for future defaults by the underlying obligors. It is challenging enough to estimate the expected (i.e., mean) number of defaults over time using historical performance of similar assets or plausible scoring models. But today's structured finance analyst has less guidance – and correspondingly less confidence – in projecting a default distribution.

Definition of the default distribution

The default distribution is simply the tabulation of all possible default outcomes and their respective probabilities. For example, if a portfolio has N obligors, then it is possible that none of these obligors will default during the life of the transaction. It's also possible that every obligor will default or that any number n with $0 < n < N$ will default. For the two endpoints (0 and N) and all values of n in between, specifying the probability function $P(n)$ defines the default distribution.

There are surprisingly few good methods for assuming (or guessing) an analytical default distribution. The most basic beginning point requires all obligors to be independent of each other and have a common default probability p. With these restrictions, the probability function becomes the binomial distribution:

$$P(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \qquad (1)$$

In addition to the independence and common default probability restrictions, equation (1) is unwieldy for large values of n and N. Even with N as low as 20, $N!$ is greater than 2×10^{18}. For portfolios where N is large (greater than 40 or so), an improvement to equation (1) is the corresponding Gaussian distribution:

$$f(x) = \left[\frac{N}{2\pi p(1-p)}\right]^{1/2} \exp\left[-N(x-p)^2 \Big/ 2p(1-p)\right] \qquad (2)$$

Here we use x as the fraction n/N of defaulting obligors and treat this $x(0 \le x \le 1)$ as a continuous variable, which is both convenient and plausible for large N. Instead of being a simple probability, the function $f(x)$ is a probability density function (PDF).

An additional, extremely useful mathematical representation of a default distribution for the continuous default variable x is the cumulative distribution function (CDF), which we write as $F(x)$. $F(x)$ is the probability that the actual default fraction is less than x. Hence, $F(0)$ is zero, since the fraction of the portfolio in default must be greater than or equal to zero. Also, $F(1)$ is one, since the default fraction will certainly be less than or equal to 100%. The PDF $f(x)$ is the (calculus) derivative of the CDF $F(x)$: $f(x) = F'(x)$. In deriving or analysing a default distribution, it is often more convenient to work with the PDF (or CDF) and then derive the CDF (or PDF) with the relationship $f(x) = F'(x)$. The CDF for the PDF of equation (2) is, in terms of the standard normal distribution $\Phi(\ )$,

$$F(x) = \Phi\left[\frac{x-p}{\sqrt{p(1-p)/N}}\right] \qquad (3)$$

Vasicek allows for correlation

Of the several shortcomings of equations (1) to (3), the most severe is the assumption that the large number N of obligors default independently of one another. This is the zero correlation assumption, and it is unwarranted. In the financial world there is not much that we know about correlation – but we know that it's not zero!

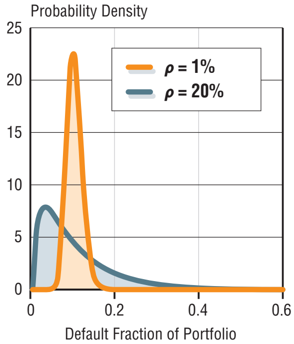
Oldrich Vasicek devised a default distribution with the assumption that all obligors have mutual asset correlation ρ with one another. The CDF for the Vasicek distribution is, in terms of the standard normal distribution $\Phi(\ )$ and its inverse $\Phi^{-1}(\ )$,

$$F(x) = \Phi\left[\frac{-\Phi^{-1}(p) + \sqrt{1-\rho}\,\Phi^{-1}(x)}{\sqrt{\rho}}\right] \qquad (4)$$

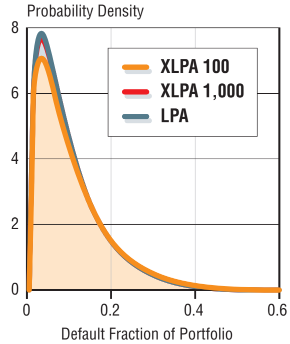
This result maintains the restrictive assumption that all obligors have the same default probability p and adds the new – and also restrictive – assumption that the number of obligors N becomes infinite. Hence, a common name for this Vasicek distribution is the large pool approximation (LPA). The PDF of the Vasicek LPA is, in terms of the standard normal density $\phi(\ )$ and standard normal distribution inverse $\Phi^{-1}(\ )$,

$$f(x) = \sqrt{\frac{1-\rho}{\rho}}\,\phi\left[\frac{\Phi^{-1}(p) - \sqrt{1-\rho}\,\Phi^{-1}(x)}{\sqrt{\rho}}\right] \Big/ \phi\left[\Phi^{-1}(x)\right] \qquad (5)$$

1: LPA PDF for 10% Default Probability



2: PDF with 10% Default Probability and 20% Correlation



Despite its limitations, the LPA is a vast improvement over binomial or Gaussian representations. One can see directly the influence of correlation ρ in equation (5). The density function is fat-tailed as is evident in Figure 1, which compares a typical asset correlation of 20% with a near-zero value of 1% (with obligor default probability p of 10%).

Extension to the Vasicek large pool approximation

Equations (4) and (5) show the impact on the default distribution of the imposition of a uniform asset correlation ρ on the obligor portfolio. It is useful to have this Vasicek closed-form analytical solution even if the assumptions do not fit real-world situations. In recent work, we have created an extended LPA (XLPA) that relaxes the assumptions of infinite pool size and common default probability. The Gaussian function of equation (2) embeds the finite pool size N. In essence, we combine the Gaussian with the LPA to get the XLPA CDF:

$$F(x) = F_\infty(x) + \int_0^1 du \left\{ \Phi\left[\frac{x-u}{\gamma\sqrt{u(1-u)/N}} \right] - H(x-u) \right\} f_\infty(u) \qquad (6)$$

In equation (6), $F_\infty(\ )$ and $f_\infty(\ )$ are the Vasicek CDF and PDF of equations (4) and (5), respectively. $H(\ )$ is the Heaviside function. One computes the parameter γ in equation (6) from the specific default probabilities of the subject portfolio. When one creates a *loss* distribution rather than a *default* distribution in this framework, the value of γ also includes variation in the obligor concentrations and loss severities.

The XLPA CDF of equation (6) is more difficult to read than the Vasicek LPA CDF of equation (4). The purpose of the XLPA is not so much to replace the LPA as it is to provide a demonstration of where the LPA remains accurate. As an example, industry practitioners apply the LPA to real portfolios that, of course, have finite pool size N and nonuniform obligor default probabilities. Figure 2 compares the Vasicek LPA PDF with that of the XLPA for pool sizes of 100 and 1,000 obligors. With mean default probability p of 10% and asset correlation of 20%, we see that a pool of 1,000 obligors is sufficiently large such that the LPA provides a good approximation. When the pool size falls to 100, though, it is better to have the XLPA correction.

Analytical solutions and simple correlation prescriptions

Expert practitioners must apply caution and judgment in the analysis of structured transactions whether the goal be pricing, risk assessment, or credit rating determination. Even if one considers the numerous assumptions for the LPA or XLPA of equations (4), (5), and (6) to be valid, the input variables of mean default probability p and asset correlation ρ are merely estimates. Hence, the resulting default distribution is, at best, a plausible estimate that will be least reliable precisely where it is most needed – in one of the tails of the distribution.

Industry best practice does not rely solely on analytical solutions. Some alternative approaches combine numerical solution techniques, such as Monte Carlo simulation, with greater refinement in correlation specification or with default clustering that is not expressible in terms of pair-wise correlation. Other approaches are of the stress test variety, which subject each portfolio asset to specific stress scenarios and/or apply a global default stress to all assets.

A full analysis should combine both the refined model solution and the stress tests, since neither provides a complete answer. The refined model default distribution will still be unreliable in the tails due to uncertainty in the inputs. The stress tests provide no guidance on the probabilistic default distribution since, by their nature, the likelihood of each scenario is unknown.

The absence of a fully satisfactory solution to the problem of structured finance analysis has two immediate ramifications. First, the application of analytical solutions such as the LPA and XLPA to simplified problems remain important because they help analysts assess the results of more complicated models and procedures. Second, pricing and risk assessment in structured finance will never be fully model driven. Out-of-model considerations – such as illiquidity and supply/demand, input uncertainty and potential downside events (involving bank counterparties, the sovereign or the currency) – all require the addition of expert judgment.

Notes for this piece may be found online.

Joe Pimbley is the principal of Maxwell Consulting (maxwell-consulting.com). His expertise includes enterprise risk management, structured products, derivatives and quantitative modelling.