Portfolio Loss Analysis:

Extending the Large Pool Approximation

How to relax the Vasicek portfolio loss distribution assumptions of an infinite number of loans and of the homogeneity of loan characteristics.

By J. M. Pimbley

A ubiquitous challenge in the financial world is the analysis of the risk of a portfolio of financial instruments.[1] The basis of structured finance is the specification of an underlying portfolio of risky assets, such as residential and commercial real estate loans and consumer receivables (for credit card and auto loans). The future performance of the underlying portfolio directly determines the performance of the structured finance debt obligations of the transaction that hosts the portfolio.

Banks and bank regulators create mathematical models to project the likely behavior of such portfolios in order to assign capital to individual holdings or to entire bank portfolios.[2] To "assign capital," it is necessary to understand how much reserve to hold against potential loss. Estimation of this reserve requires specification of the cumulative distribution function (CDF) for portfolio loss. Debt ratings for structured finance transactions also require knowledge of this portfolio CDF.[3]

Vasicek Infinite Pool Size Result

Oldrich Vasicek provided the founding contribution to the determination of an approximate portfolio CDF for credit losses in a portfolio of debt instruments.[4] With the assumptions that each loan has identical default probability p, size, and loss-given-default (LGD), that each loan has a single asset correlation ρ $(0 < \rho < 1)$with every other loan and that the number of loans approaches infinity, Vasicek derived the CDF $F(x)$ to be:

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

The independent variable $x \in (0,1)$ in equation (1) is the portfolio loss expressed as a fraction of the total possible loss.[5] The CDF $F(x)$ is the probability that the actual (fractional) loss will be less than or equal to x. By this definition, then, $F(x)$ is a non-decreasing function of x, with and $F(0) = 0$ and $F(1) = 1$.[6] The symbol Φ ( ) denotes the standard normal CDF of the quantity within the parentheses while Φ^{-1}( ) represents the inverse of this standard normal CDF.

It is helpful to define and construct the probability density function (PDF) $f(x)$ for the loan loss distribution. Like the CDF, the PDF has an interpretation as a probability. The product $f(x)dx$ is the probability that the portfolio loss will ultimately lie in the interval $(x, x+dx)$ as dx approaches zero. The PDF is the derivative of the CDF with respect to the loss fraction x, as follows:

$$f(x) = F'(x) = \sqrt{\frac{1-\rho}{\rho}}\,\phi\left[\frac{K-\sqrt{1-\rho}\,\Phi^{-1}(x)}{\sqrt{\rho}}\right]\Big/\phi\left[\Phi^{-1}(x)\right] \qquad (2)$$

In equation (2), the symbol ϕ ( ) denotes the standard normal PDF of the quantity within the parentheses.

Vasicek's equation (1) result is directly useful since it provides the confidence interval for specific potential loan portfolio loss amounts. Thus, a bank may apply this $F(x)$ to determine economic capital (a key risk measure) of its holdings. A bank regulator may use $F(x)$ to specify the capital ("regulatory capital") it requires banks to hold against a lending portfolio.[7] A rating agency or investor in structured finance debt may use $F(x)$ to estimate the probability of loss of the debt.

Warning Regarding the Use of Models

Of course, whether a bank, bank regulator, rating agency, or investor performs this analysis, the result is an approximation, at best, given the restrictive assumptions. First, real loan portfolios do not consist of loans that are all equal in size, default probability and LGD. While it is reasonable to suspect that simplifying a real portfolio to its homogeneous counterpart (with mean size, mean default probability, and mean LGD) will lead to an adequate approximation for a sufficiently large number of loans, we have no guidance on how "large" this number of loans must be.[8]

Further, even if the loan portfolio already happens to be homogeneous, the Vasicek formulation does not indicate how many loans (100? 10,000?) the portfolio must have to be well represented by equation (1) — which assumes an infinite number of loans. Finally, the assumption of a single correlation ρ relating all loans is certainly an idealization, but we will retain it.

The contributions of this article are that we remove the constraints of the Vasicek model that the loan portfolio must be homogeneous and infinite. This generalization then allows us to explore the requisite (finite) portfolio size at which the Vasicek approximation becomes reasonably accurate. But the assumptions regarding correlation remain unchanged.

This treatment assumes pairwise correlations only. While the reliance on pairwise correlations is standard in the financial industry, it does ignore the possibility of more complex interactions for loan defaults. Further, even within the single-factor pairwise framework, it is well understood that a single correlation value is not appropriate for all points on the loss distribution. The Vasicek and extended Vasicek solutions we present here are useful tools, but the user should apply them in a manner that recognizes that one correlation value may not provide the best fit to all segments of the loss distribution.

A broader statement that applies to the results of this article and to all financial models generally is that one must exercise caution and discretion in the interpretation of the model output. Well-constructed models are best suited as guides to illustrate the dependence of the output (in this case, the portfolio capital requirement) on the input data and assumptions (such as default probabilities of loans and correlation among loans).[9] Uncertainty in the input data necessarily translates to uncertainty in the output results. Hence, decisions regarding capital requirements, credit ratings, investment suitability or other real-world uses should combine expert judgment with the model results.

Extended Vasicek Result for a Finite Homogeneous Pool

The purpose of this communication is to extend the Vasicek results for the CDF and PDF of a loan portfolio — which we also call the Large Pool Approximation (LPA) — to the case of a large (but finite), non-homogeneous portfolio. We consider first the case of a homogeneous portfolio with N loans with $N>>1$.

We impose correlation among the N loans in the same manner as Vasicek by positing an asset value V_n for each loan in terms of the correlation ρ, as follows:

$$V_n = \sqrt{\rho}\,Y + \sqrt{1-\rho}\,\varepsilon_n \qquad (3)$$

The random variates Y and ε_n are independent and normally distributed with a mean of zero and a variance of one. The V_n all share the common factor Y, which is the generator of correlation among the V_n. For example, direct calculation from equation (3) shows that $Corr(V_i, V_j) = \rho$ for $i \neq j$. One may interpret Y as "state of the economy," or "home price appreciation," or "unemployment," depending on the particular loan type one is attempting to approximate.

The asset values V_n are themselves normally distributed with a mean of zero and a variance of one. As stated above, all pairs of asset values have the same ("flat") "asset correlation" ρ. The creation of an "asset value" to represent each loan has no deep meaning.[10] It is merely a device to impose correlation.

Once constructed in this manner, we determine if a specific loan i has defaulted by comparing V_i to $\Phi^{-1}(p)$, where p is the common default probability of each loan. If $V_i < \Phi^{-1}(p)$, then we say the loan is in default.

Beyond providing the desired correlation, equation (3) is useful because it permits us to treat the N loans as independent *once we have fixed the value of* Y. Following the well trod path of Vasicek and others,[11] we analyze the portfolio loan loss distribution for a fixed Y and then integrate over all possible values of Y. The probability $\tilde{f}_n$ that exactly n of N loans will default, given a default probability of $\hat{p}$, is[12]

$$\tilde{f}_n = \binom{N}{n} \hat{p}^n (1-\hat{p})^{N-n} \qquad (4)$$

At this point the Vasicek method invokes the "law of large numbers" to stipulate that the number of defaults n will equal $\hat{p}N$ as $N \to \infty$ which is equivalent to the finding that the fraction $\hat{p}$ of all loans default.

This last reduction of equation (4) to the specification that $n = \hat{p}N$ is the step in the current LPA derivation that requires N to be infinite. We propose here an improvement that approximates equation (4) for large, but finite, N. By applying Stirling's approximation[13] ($\log N! \approx N \log N - N + \tfrac{1}{2}\log 2\pi N$, where "log" denotes the natural logarithm) to the factorial terms of (4), we derive the new approximation,[14] as follows:

$$\tilde{f}_n = \binom{N}{n} \hat{p}^n (1-\hat{p})^{N-n} \approx$$

$$\left[2\pi\hat{p}(1-\hat{p})N\right]^{-1/2} \exp\left[{-N(x-\hat{p})^2} \Big/ {2\hat{p}(1-\hat{p})} \right], \quad N >> 1 \qquad (5)$$

$$\text{with} \quad x = n/N$$

Since N is large, we can consider x to be a continuous variable, because the separation between permissible values of x is $1/N$. The conversion from the probability function $\tilde{f}_n$ to the PDF $\hat{f}(x)$ for the continuous variable x is $\hat{f}(x) = N\tilde{f}_n$. We use the "hat" notation in $\hat{f}(x)$ to indicate that this PDF still assumes (or "is conditioned on") a specific value of the common factor Y. From equation (5), then, we get

$$\hat{f}(x) = \left[\frac{N}{2\pi\hat{p}(1-\hat{p})}\right]^{1/2} \exp\left[{-N(x-\hat{p})^2} \Big/ {2\hat{p}(1-\hat{p})} \right] \qquad (6)$$

Equation (6) is convenient and intuitive in that it is a normal density function. Hence, it is entirely consistent with the Central Limit Theorem and it approaches a Dirac delta function as $N \to \infty$. The Vasicek LPA is this Dirac limit.[15] Hence, our extension of the Vasicek LPA is simply the substitution of equation (6) for the Dirac delta function.

With $\hat{f}(x)$ from equation (6), we've determined the PDF for a given value of Y. Recall that equation (3) gives the default behavior of each of the N loans. We now impose the asset correlation ρ to give us the correlated loan default PDF $f(x)$ by integrating $\hat{f}(x)$ over all possible values of the normally distributed Y, as follows:

$$f(x) = \int_{-\infty}^{+\infty} dy\, \phi(y)\, \hat{f}(x) \qquad (7)$$

As before, $\phi(\ )$ is the standard normal PDF. Equation (6) shows that $\hat{f}(x)$ depends on y through the default probability $\hat{p}$ which we clarify here to be

$$\hat{p} = \Phi\left(\frac{K - \sqrt{\rho}\, y}{\sqrt{1-\rho}}\right) \qquad (8)$$

with K already defined in equation (1).

Applying a change of variable to the integral in equation (7), we find

$$f(x) = \left(\frac{N}{2\pi}\right)^{1/2} \int_0^1 du\, \frac{\exp\left[-N(x-u)^2/2u(1-u)\right]}{\sqrt{u(1-u)}} f_\infty(u) \qquad (9)$$

The notation $f_\infty(\ )$ represents the ($N \to \infty$) Vasicek LPA expression that we stated in equation (2) and copy here:

$$f_\infty(x) = \sqrt{\frac{1-\rho}{\rho}}\, \phi\left[\frac{K - \sqrt{1-\rho}\, \Phi^{-1}(x)}{\sqrt{\rho}}\right] \Big/ \phi\left[\Phi^{-1}(x)\right] \qquad (10)$$

To get the CDF for the portfolio loss distribution, we go back to equation (7) and apply the property $F(x) = \int_0^x dt\, f(t)$ to find

$$F(x) = \int_{-\infty}^{+\infty} dy\, \phi(y)\, \Phi\left[\frac{x-\hat{p}}{\sqrt{\hat{p}(1-\hat{p})/N}}\right]$$

Applying the same change of variable that produced equation (9), we write

$$F(x) = \int_0^1 du\, \Phi\left[\frac{x-u}{\sqrt{u(1-u)/N}}\right] f_\infty(u) \qquad (11)$$

An unfortunate aspect of both equations (9) and (11) is that they require numerical integrations. These forms do not easily show the reader how different these extended LPA ("XLPA") results differ from the LPA ($N \to \infty$) counterparts. To reduce this difficulty, we recast equation (10) into the following less compact but more insightful form:

$$F(x) = F_\infty(x) + \int_0^1 du \left\{\Phi\left[\frac{x-u}{\sqrt{u(1-u)/N}}\right] - H(x-u)\right\} f_\infty(u) \qquad (12)$$

In equation (12), $H(\,)$ is the Heaviside function[16] and $F_\infty(x)$ is the LPA CDF, which we wrote in equation (1) and reproduce here, as follows:

$$F_\infty(x) = \Phi\left[\frac{-K + \sqrt{1-\rho}\,\Phi^{-1}(x)}{\sqrt{\rho}}\right] \qquad (13)$$

Granularity Adjustment

The integral term of equation (12) is the difference between the CDFs for the LPA and for the XLPA (which permits the number of loans N to be finite). Numerous studies over the past years have investigated this difference and given it the descriptive term "granularity adjustment" (GA).[17] These studies did not derive and evaluate the integral in equation (12); rather, they have developed asymptotic representations of the GA in powers of $1/N$.

We believe results of this existing analysis are consistent with this new XLPA result. For example, one can show by direct analysis that the difference between $f(x)$ and $f_\infty(x)$ in equation (9) is $O(1/N)$. Since $F(x)=\int_0^x dt\, f(t)$, the difference between $F(x)$ and $F_\infty(x)$ and would also behave as $O(1/N)$ for large N. Numerical evaluations of equation (12) confirm this behavior.

Comparison of the LPA and XLPA

Figures 1a and 1b compare the PDF for the LPA (equation (10)) with the PDF for various values of the number of loans N for the XLPA (equation (9)), with a loan default probability of 10% and an (asset) correlation of 5%.[18]

Figure 1a: PDF for 10% PD and 5% Correlation

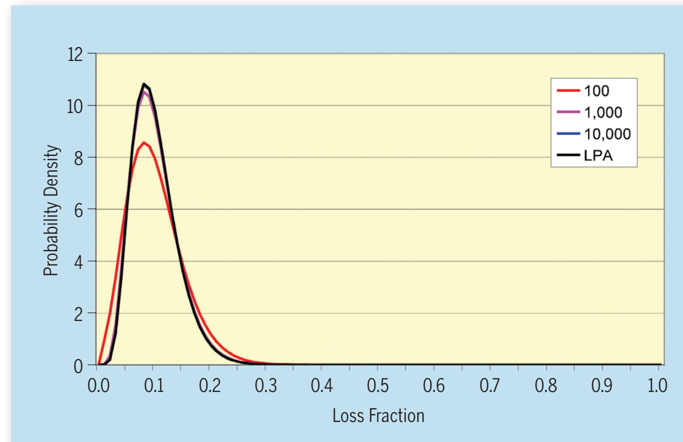

Figure 1b: PDF for 10% PD and 5% Correlation

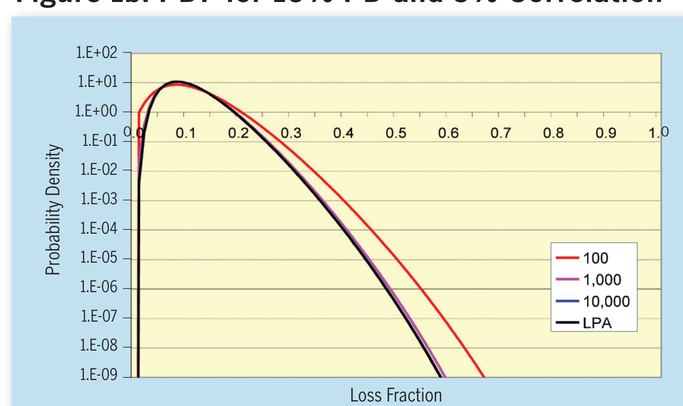

Figure 1a specifies a linear y-axis while figure 1b is logarithmic in order to see the extended tail as the loss fraction x increases. The XLPA PDF with 10,000 loans is visually indistinguishable from the LPA PDF. With 1,000 loans, the difference between the XLPA and LPA is visible; while for 100 loans, the difference is significant. Roughly speaking, then, the LPA is a good approximation to a finite loan pool of 1,000 loans or greater with default probability of 10% and correlation of 5%. The LPA is not a good approximation when there are only 100 loans.

Figures 2a and 2b (see pg. 18) change the correlation from 5% to 20%. The shapes of the curves change considerably, with the peak of the PDF moving to lower values of loss fraction and the high-loss-fraction tail increasing. Yet the LPA appears to become a better approximation with higher correlation (all else being equal).

Figure 2a: PDF for 10% PD and 20% Correlation

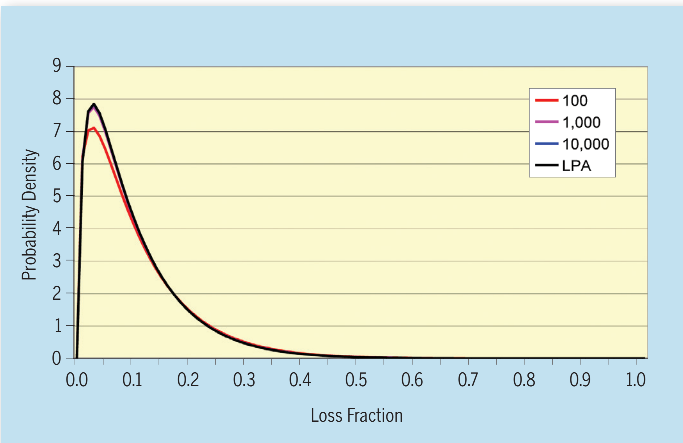

Figure 2b: PDF for 10% PD and 20% Correlation

Extended Vasicek Result for a Finite, Non-homogeneous Pool

To this point, the XLPA has simply extended the LPA result to the case of a finite number of loans, while still requiring the homogeneity of uniform loan size, default probability, and LGD. Relaxing this last requirement for these portfolio parameters is not as challenging as it may seem, as long as the pool remains reasonably diversified.

Consider the result in equation (6) for the homogeneous pool loss PDF $\hat{f}(x)$ given a fixed value of the common factor Y. This PDF is a normal density function with mean $\hat{p}$ and standard deviation $\sqrt{\hat{p}(1-\hat{p})/N}$, which we will define as σ. While we cannot directly derive the PDF for the non-homogeneous pool, we can derive the mean and standard deviation of this unknown PDF and find:[19]

$$\text{Mean} = \sum_i L_i \hat{p}_i \lambda_i \Big/ \sum_i L_i \lambda_i \tag{14a}$$

$$\text{Standard Deviation} = \left[\sum_i L_i^2 \lambda_i^2 \hat{p}_i (1-\hat{p}_i)\right]^{1/2} \Big/ \sum_i L_i \lambda_i \tag{14b}$$

In equations (14a) and (14b), L_i, λ_i, $\hat{p}_i$ and are the known par amount, assumed LGD and specified default probability, respectively, for loan i, with i ranging from 1 to N. As Appendix B shows, equation (14b) becomes the homogeneous pool value of $\sqrt{\hat{p}(1-\hat{p})/N}$ in the special case that all the L_i, λ_i and $\hat{p}_i$ equal the common values of L, λ and $\hat{p}$ respectively, as one would expect.

Let us define the non-homogeneous pool PDF to be $\hat{f}_{nh}(x)$. The mean of $\hat{f}_{nh}(x)$ is $\hat{p}$ — the same as the mean of the homogeneous pool PDF $\hat{f}(x)$ — since we define $\hat{p}$ for the non-homogeneous portfolio to be the weighted average in equation (14a). The standard deviation of $\hat{f}_{nh}(x)$ will, in general, be larger than that of $\hat{f}(x)$ for the homogeneous pool of the same size N and mean values of L, λ, and $\hat{p}$. For convenience, then, we write the standard deviation of $\hat{f}_{nh}(x)$ as $\gamma\sqrt{\hat{p}(1-\hat{p})/N}$ where, from equation (14b), we find

$$\gamma = \left[\frac{N}{\hat{p}(1-\hat{p})}\sum_i L_i^2 \lambda_i^2 \hat{p}_i (1-\hat{p}_i)\right]^{1/2} \Big/ \sum_i L_i \lambda_i \tag{15}$$

Given an actual loan portfolio, it is straightforward to compute the numerical value of γ from equation (15).

The probability density function resulting from the addition of a large number of independent contributions approaches a normal density function.[20] Just as we required the number of loans N to be large so that equation (6) provided a form of the homogeneous pool PDF $\hat{f}(x)$ more useful than that of equation (4), the large N requirement permits us to specify the non-homogeneous pool PDF $\hat{f}_{nh}(x)$ as the normal density function with desired mean and standard deviation, as follows:

$$\hat{f}_{nh}(x) = \frac{1}{\gamma}\left[\frac{N}{2\pi\hat{p}(1-\hat{p})}\right]^{1/2} \exp\left[-N(x-\hat{p})^2 \Big/ 2\gamma^2\hat{p}(1-\hat{p})\right] \tag{16}$$

In other words, solving the non-homogeneous pool problem is not difficult. One need only compute the γ value from the actual portfolio parameters and modify the XLPA expressions we derived previously. The only qualification is that, just like the original Vasicek LPA, there exists no clear guideline on how large N must be for the vague constraint "large N" to be satisfied.

Solving the non-homogeneous pool problem is not difficult. One need only compute the γ value from the actual portfolio parameters and modify the XLPA expressions we derived previously.

As a trivial example, we found earlier that 10,000 loans constitute a sufficiently large pool for the Vasicek LPA to be valid for the homogeneous pool with the parameters we specified. (In fact, the single greatest attribute of the XLPA is simply that it provides a check on the LPA.) Suppose we now add just one additional loan with par amount equal to that of the entire pre-existing portfolio.

Even if the LGD and default probability of this new loan equal those of the other loans, this is a huge non-homogeneity! In this extreme case, the PDF of the loss for this 10,001-loan portfolio would not be well approximated by a normal density function. Hence, equation (16) would not be appropriate. Rather, equation (16) would only become appropriate for a much larger number of loans if this type of extreme par amount idiosyncrasy is present.

Just as equation (16) extends the equation (6) homogeneous PDF to the case of a more realistic non-homogeneous portfolio, we give the extensions to equations (9), (11), and (12), for the final PDF and CDF results, as

$$f_{nh}(x) = \left(\frac{N}{2\pi}\right)^{1/2} \int_0^1 du \, \frac{\exp\left[-N(x-u)^2/2\gamma^2 u(1-u)\right]}{\gamma\sqrt{u(1-u)}} f_\infty(u) \quad (17)$$

$$F_{nh}(x) = \int_0^1 du \, \Phi\left[\frac{x-u}{\gamma\sqrt{u(1-u)/N}}\right] f_\infty(u) \quad (18)$$

$$F_{nh}(x) = F_\infty(x) + \int_0^1 du \left\{\Phi\left[\frac{x-u}{\gamma\sqrt{u(1-u)/N}}\right] - H(x-u)\right\} f_\infty(u) \quad (19)$$

The functions $f_\infty(\,)$ and $F_\infty(\,)$ retain their earlier meanings as the PDF and CDF, respectively, of the Vasicek LPA.

FOOTNOTES

1. What we now call "modern portfolio theory" dates back to the seminal work of Harry Markowitz. See H. M. Markowitz, "Portfolio Selection," *The Journal of Finance* 7(1), 77-91, March 1952. See also H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, 1959.

2. See, for example, M. B. Gordy, "A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules," *J. Financial Intermediation* 12, 199-232, 2003.

3. See, for example, DBRS, Inc., "Master European Granular Corporate Securitisations (SME CLOs)," June 2011, at http://dbrs.com/research/240236/master-european-granular-corporate-securitisations-sme-clos.pdf .

4. O. Vasicek. "Probability of Loss on Loan Portfolio," KMV Working Paper, 1987. See also the derivation in P. J. Schönbucher, *Credit Derivatives Pricing Models*, John Wiley & Sons Ltd., 2003.

5. The total possible loss is the sum over all portfolio loans of the product of loan size and loan LGD.

6. As a qualification to this statement, it is possible for $F(0)$ to be greater than zero if we permit discontinuous density functions. The interpretation of such a positive value is simply that there is a non-zero probability that the portfolio loss is precisely zero.

7. See, for example, P. S. Calem and J. R. Follain, "The Asset-Correlation Parameter in Basel II for Mortgages on Single-Family Residences," *Background for Public Comment on the Advanced Notice of Proposed Rulemaking on the Proposed New Basel Capital Accord,* Board of Governors of the Federal Reserve System, November 2003.

8. Non-homogeneous loan portfolios have arbitrary distributions of loan size, default probability and LGD. The nature of these distributions will impact whether the homogeneous counterpart provides a reasonable approximation for a specific, large number of loans. The substitution of "mean" values to create the homogeneous portfolio will need to employ suitably weighted means.

9. In this role, models serve as excellent tests of data quality since erroneous input data often present vividly in the model output.

10. It's likely that the Merton model for investigating the default probability of a corporate entity's debt motivated Vasicek's treatment. In the Merton approach, the "asset value" really does represent something close to the value of the assets of the firm. But this background thought process is not necessary.

11. Up to this point, we are recounting the discussion the reader will find in O. Vasicek, "Probability of Loss on Loan Portfolio,"

KMV Working Paper, 1987, or in P. J. Schönbucher, *Credit Derivatives Pricing Models*, John Wiley & Sons Ltd., 2003.

12. We use the symbol $\hat{p}$ here, rather than p, to indicate that this is the default probability contingent on a specific value of Y rather than the final loan default probability. The probability function $\tilde{f}_n$ is the discrete variable analog to the probability density function $f(x)$ for the continuous variable x.

13. See, for example, the web reference http://planetmath.org/encyclopedia/StirlingsApproximation.html for this approximation of $\log N!$ that improves as N increases. Even with N as small as 5, the accuracy is better than 0.5%. The accuracy surpasses 0.1% at N equal to 10 and continues to improve thereafter.

14. Appendix A provides more detail on the derivation of equation (5) from equation (4).

15. In O. Vasicek, "Probability of Loss on Loan Portfolio," KMV Working Paper, 1987, or in P. J. Schönbucher, *Credit Derivatives Pricing Models*, John Wiley & Sons Ltd., 2003, these authors invoke the "law of large numbers" to say that a precise fraction of loan defaults will occur as $N\to\infty$. The mathematical translation of this statement is that the PDF for the fraction x is a Dirac delta function.

16. A more descriptive name for the discontinuous Heaviside function is the "step function." $H(x)$ is 1 for $x>0$, 0 for $x<0$ and one-half for $x=0$.

17. An excellent and recent discussion is M. B. Gordy and J. Marrone, "Granularity Adjustment for Mark-to-Market Credit Risk Models," Finance and Economic Discussion Series 2010-37, Federal Reserve Board, June 2010

18. We evaluated equation (9) numerically by a sequential Simpson's Rule technique.

19. See Appendix B for a derivation of these results.

20. This is the Central Limit Theorem. See, for example, the web reference http://en.wikipedia.org/wiki/Central_limit_theorem.

Dr. J.M. ("Joe") Pimbley is Principal of Maxwell Consulting, a consulting firm he founded in 2010, and a member of Risk Professional's editorial board. His recent and current engagements include financial risk management advisory, underwriting for structured and other financial instruments, and litigation testimony and consultation. In a prominent engagement from 2009 to 2010, Joe served as a lead investigator for the Examiner appointed by the Lehman bankruptcy court to resolve numerous issues pertaining to history's largest bankruptcy.

He acknowledges many helpful conversations with G. Leppert of DBRS Inc., which contributed to this article.

Appendix A

Beginning with equation (4), we seek to use Stirling's approximation to derive equation (5). Taking the natural logarithm of both sides of (4), we find

$$\log \tilde{f}_n = \log\left[\binom{N}{n}\hat{p}^n(1-\hat{p})^{N-n}\right]$$

$$\log N! - \log n! - \log(N-n)! + n\log\hat{p} + (N-n)\log(1-\hat{p}) \quad . \quad \text{(A1)}$$

Applying

$$\log N! \approx N\log N - N + \frac{1}{2}\log 2\pi N$$

to the first term of (A1) and similar approximations to the other log- factorial terms, we find

$$\frac{1}{N}\log\tilde{f}_n \approx (1-x)\log\frac{1-\hat{p}}{1-x} + x\log\frac{\hat{p}}{x} - \frac{1}{2N}\log[2\pi x(1-x)N] \quad . \quad \text{(A2)}$$

In (A2), we eliminated n by replacing it with Nx, so that x is the loss fraction. Both $\hat{p}$ and x have values in the range (0,1). Inspection of the first two terms on the right-hand side of (A2) shows a maximum value of zero at $x = \hat{p}$ with local quadratic behavior, as follows:

$$(1-x)\log\frac{1-\hat{p}}{1-x} + x\log\frac{\hat{p}}{x} \approx \frac{-(x-\hat{p})^2}{2\hat{p}(1-\hat{p})} \text{ for } x \text{ near } \hat{p}\,. \quad \text{(A3)}$$

We neglected the $\log x(1-x)$ term in (A2) for purposes of approximating the x-dependence since this term is divided by N, which makes it small relative to the first two terms. The result in (A3) is useful since it shows that, from (A2), $\tilde{f}_n$ will include, as follows, the exponential of a large (since multiplied by N) and negative quadratic function:

$$\tilde{f}_n \approx [2\pi\hat{p}(1-\hat{p})N]^{-1/2}\exp\left[-N(x-\hat{p})^2 \big/ 2\hat{p}(1-\hat{p})\right] \quad . \quad \text{(A4)}$$

Equation (A4) is identical to equation (5) which is what we sought to prove. To get (A4), we made one other approximation. We replaced $x(1-x)$ by $\hat{p}(1-\hat{p})$ in the term preceding the exponential. The justification for this replacement is that the exponential term in (A4) will make $\tilde{f}_n$ very small for any value of x not sufficiently close to $\hat{p}$, given that these approximations are reasonable only for $N>>1$. The replacement is extremely helpful because it produces both the familiar Gaussian functional form for $\tilde{f}_n$ and the correct normalization for $\tilde{f}(x)$ in equation (6) as a PDF.

Let's consider how the standard deviation expression for a non-homogeneous loan portfolio (equation (B5)) changes when we make all loan par amounts equal ($L_i \rightarrow L$), all LGD values equal ($\lambda_i \rightarrow \lambda$), and all default probabilities equal ($p_i \rightarrow p$). The denominator of (B5) becomes $N L\lambda$ while the summation in the numerator becomes $NL^2\lambda^2 p(1-p)$.

Appendix B

Our goal is to derive the mean and standard deviation of the fractional loss of a non-homogeneous portfolio, as shown in equations (14a) and (14b), respectively. With N loans of par amount L_i, loss-given-default (LGD) λ_i and default probability p_i, the portfolio loss is

$$Loss = \sum_i L_i \lambda_i \varepsilon_i \quad . \quad \text{(B1)}$$

In (B1), the ε_i term equals 1 (probability p_i) when loan i is in default and zero (probability 1- p_i) otherwise. All summations range from 1 to N. The expected loss is

$$E\{Loss\} = E\left\{\sum_i L_i \lambda_i \varepsilon_i\right\} = \sum_i L_i \lambda_i E\{\varepsilon_i\} = \sum_i L_i \lambda_i \hat{p}_i \quad . \quad \text{(B2)}$$

Dividing this expected loss by the total possible loss of $\sum_i L_i \lambda_i$, we find the expected loss fraction is precisely the expression in (14a).

To get the standard deviation of the loss fraction (14b), we must first find the variance of the portfolio loss, which leads us to determine the expected square of the loss, as follows:

$$E\{Loss^2\} = E\left\{\sum_{i,j} L_i L_j \lambda_i \lambda_j \varepsilon_i \varepsilon_j\right\} = \sum_{i,j} L_i L_j \lambda_i \lambda_j E\{\varepsilon_i \varepsilon_j\}$$

$$= \sum_i L_i^2 \lambda_i^2 \hat{p}_i + \sum_{i \neq j} L_i L_j \lambda_i \lambda_j \hat{p}_i \hat{p}_j \quad . \quad \text{(B3)}$$

Since the variance of the loss is the expectation of the squared loss minus the square of the expected loss, we use (B2) and (B3) to write

$$Var\{Loss\} = E\{Loss^2\} - [E\{Loss\}]^2$$

$$= \sum_i L_i^2 \lambda_i^2 \hat{p}_i + \sum_{i \neq j} L_i L_j \lambda_i \lambda_j \hat{p}_i \hat{p}_j - \sum_{i,j} L_i L_j \lambda_i \lambda_j \hat{p}_i \hat{p}_j$$

$$= \sum_i L_i^2 \lambda_i^2 \hat{p}_i (1 - \hat{p}_i) \quad . \quad \text{(B4)}$$

Taking the square root of this variance to get the standard deviation of the portfolio loss and then dividing by the total possible loss $\sum_i L_i \lambda_i$, we find the standard deviation of loss fraction to be

$$= \left[\sum_i L_i^2 \lambda_i^2 \hat{p}_i (1 - \hat{p}_i)\right]^{1/2} \Big/ \sum_i L_i \lambda_i \qquad \text{(B5)}$$

which is identical to (14b).

As a final exercise, let's consider how the standard deviation expression for a non-homogeneous loan portfolio (equation (B5)) changes when we make all loan par amounts equal ($L_i \rightarrow L$), all LGD values equal ($\lambda_i \rightarrow \lambda$), and all default probabilities equal ($p_i \rightarrow p$). The denominator of (B5) becomes $NL\lambda$ while the summation in the numerator becomes $NL^2\lambda^2 p(1-p)$. Consequently, taking the square root of this term and dividing by the denominator, we find the standard deviation of loss fraction for the "homogeneous portfolio limit" to be $\sqrt{\hat{p}(1-\hat{p})/N}$, which corresponds to our earlier results for the homogeneous portfolio.