Loan Analytics in DROP

v4.36 30 September 2019

The Distribution of Loan Portfolio Value

Introduction and Overview

1. Portfolio Debt Securities Loss Capital: The amount of capital necessary to support debt securities depends on the probability distribution of the portfolio loss.
2. Portfolio Subject to Default Loss: Consider a portfolio of loans, each of which is subject to default resulting in a loss to the lender.
3. Financing Split of the Portfolio: Suppose that the portfolio is financed partly by equity capital and partly by borrowed funds.
4. Credit Quality of the Portfolio Notes: The credit quality of the lender's notes will depend on the probability that the loss on the portfolio exceeds the equity capital.
5. Maintenance of the Notes' Rating: To achieve a certain credit rating of its notes – say Aa on a rating agency scale – the lender needs to keep the probability of default on the notes at the level corresponding to that rating – about 0.001 for the Aa quality.
6. Estimating the Corresponding Equity Capital: This means that the equity capital allocated to the portfolio must be equal to the percentile of the distribution of the portfolio loss that corresponds to the desired probability.
7. Application of Portfolio Loss Distribution: In addition to determining the capital necessary to support a loan portfolio, the probability distribution of portfolio losses has a number of other applications. It can be used in regulatory reporting, measuring portfolio risk, calculating the VaR, portfolio optimization, and structuring and pricing of debt portfolio derivatives such as Collateralized Debt Obligations (CDO).
8. Primary Focus of this Chapter: Following Vasicek (2002), this chapter derives the portfolio loss distribution under certain assumptions.

9. Portfolio Size Loss Distribution Dependence: It is shown here that the distribution converges with increasing portfolio size to a limiting type, whose analytical form is give here.
10. Literature Related to Loss Distribution: The results of the first two sections of this chapter are contained in the technical notes provided by Vasicek (1987, 1991). For a review of literature on the subject, see, for instance, Pykhtin and Dev (2002).

The Limiting Distribution of Portfolio Losses

1. Loan Borrower Asset Value Model: Assume that a loan defaults if the value of the borrower's assets at the loan maturity T falls below the contractual value B of its obligations.
2. Loan Borrower Asset Value Dynamics: Let A_i be the value of the i^th borrower's assets, described by the process

$$\Delta A_i = \mu_i A_i \Delta t + \sigma_i A_i \Delta x_i$$

3. Evolution of the Asset Value: The asset value at T can be represented as

$$\log A_i(T) = \log A_i(0) + \mu_i T - \frac{1}{2}\sigma_i T^2 + \sigma_i \sqrt{T} X_i$$

   where X_i is a standard normal variable.
4. Default Probability for Loan i: The probability of default of the i^th loan is then

$$p_i = \mathbb{P}[A_i(T) < B_i] = \mathbb{P}[X_i < c_i] = \Phi(c_i)$$

where

$$c_i = \frac{\log B_i - \log A_i(0) - \mu_i T + \frac{1}{2}\sigma_i T^2}{\sigma_i \sqrt{T}}$$

and Φ is the cumulative normal distribution function.

5. Homogenous Portfolio Loan Component Specification: Consider a portfolio consisting of n loans in equal dollar amounts. Let the probability of default of any one loan be p, and assume that the asset values of the borrowing companies are correlated with a coefficient ρ for any two companies. It is further assumed that all loans have the same term T.
6. Single and Portfolio Loan Loss: Let L_i be the gross loss before recoveries on the i^{th} loan so that

$$L_i = 1$$

if the i^{th} borrower defaults and

$$L_i = 0$$

otherwise. Let L be the portfolio percentage gross loss

$$L = \frac{1}{n}\sum_{i=1}^{n} L_i$$

7. Independence of Component Loan Losses: If the events of the defaults of the loans in the portfolio were independent of each other, the portfolio loss distribution would converge, by the central limit theorem, to a normal distribution as the portfolio size increases.
8. Convergence of the Portfolio Loss: Because the defaults are not independent, however, the conditions of the central limit theorem are not satisfied, and L is not asymptotically normal. It turns out, however, that the distribution of the portfolio loss does converge to a limiting term, which shall now be derived.
9. Systemic and Idiosyncratic Factor Decomposition: The variables X_i in

$$\log A_i(T) = \log A_i(0) + \mu_i T - \frac{1}{2}\sigma_i T^2 + \sigma_i \sqrt{T} X_i$$

are jointly standard normal with equal pair-wise correlation ρ, and can therefore be represented as

$$X_i = Y\sqrt{\rho} + Z_i\sqrt{1-\rho}$$

where $Y, Z_1, \cdots, Z_n$ are mutually independent standard normal variables. This is not an assumption, but a property of the equi-correlated normal distribution.
10. Y as Systemic Common Factor: The variable Y can be interpreted as a portfolio common factor, such as an economic index, over the interval $(0, T)$.

11. <u>Z_i as Idiosyncratic Individual Factor</u>: The term $Y\sqrt{\rho}$ is the company's exposure to the common factor and the term $Z_i\sqrt{1-\rho}$ represents the company specific risk.
12. <u>Portfolio Loss Conditional on Y</u>: The probability of the portfolio loss as an expectation over the common factor Y will now be evaluated given the conditional probability Y.
13. <u>Interpretation of the Conditional Portfolio Loss</u>: This can be interpreted as assuming various scenarios for the economy, determining the probability of a given portfolio loss under each scenario, and weighting each scenario by its likelihood.
14. <u>Single Loan Conditional Default Probability</u>: When the common factor is fixed, the conditional probability of loss on any one loan is

$$p(Y) = \mathbb{P}[L_i = 1|Y] = \Phi\left(\frac{\Phi^{-1}(p) - Y\sqrt{\rho}}{\sqrt{1-\rho}}\right)$$

15. <u>Single Loan Unconditional Default Probability</u>: The quantity $p(Y)$ provides the loan default probability under the given scenario. The unconditional default probability p is just the average of the conditional probabilities over all the scenarios.
16. <u>Conditional on Y Distribution of L_i</u>: Conditional on the value of Y, the variables X_i are independent and equally distributed with a finite variance.
17. <u>Applying CLT to Large Portfolio</u>: The portfolio loss conditional on Y converges, by the law of large numbers, to its expectation $p(Y)$ as

$$n \to \infty$$

18. <u>Unconditional Large Portfolio Loss Distribution</u>: Then

$$\mathbb{P}[L \leq x] = \mathbb{P}[p(Y) \leq x] = \mathbb{P}[Y \geq p^{-1}(x)] = \Phi\left(-p^{-1}(x)\right)$$

and, on substitution, the cumulative distribution function of loan losses on a very large portfolio is in the limit

$$\mathbb{P}[L \leq x] = \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)$$

This result is given by Vasicek (1991).

19. Portfolio Weights with Unequal Components: The convergence of the portfolio loss distribution to the limiting form above actually holds even for portfolios with unequal weights.
20. Applying CLT to this Portfolio: Let the portfolio weights be $w_1, \cdots, w_n$ with $\sum_{i=1}^{n} w_i$ the portfolio loss

$$L = \sum_{i=1}^{n} w_i L_i$$

conditional on Y conditional on Y converges to its expectation $p(Y)$ whenever – and this is a necessary and sufficient condition –

$$\sum_{i=1}^{n} {w_i}^2 \to 0$$

21. Interpretation of the Weight Condition: In other words, if the portfolio contains a sufficiently large number of loans without it being dominated by a few loans much larger than the rest, the limiting distribution provides a good approximation for the portfolio loss.

Properties of the Loss Distribution

1. Support of the Portfolio Loss PDF: The portfolio loss distribution given by the cumulative distribution function

$$F(x; p, \rho) = \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)$$

is a continuous distribution concentrated on the interval

$$0 \le x \le 1$$

2. Two Parameter Family of Distribution: It forms a two-parameter family with the parameters

$$p > 0$$

and

$$\rho < 1$$

3. Limiting Behavior for ρ Values: When

$$\rho \to 0$$

it converges to a one-point distribution concentrated at

$$L = p$$

When

$$\rho \to 1$$

it converges to a zero-one distribution with probabilities p and $1 - p$, respectively.

4. Limiting Behavior for p Values: When

$$p \to 1$$

the distribution becomes concentrated at

$$L = 0$$

and

$$L = 1$$

respectively.

5. Symmetry Property of the Distribution: The distribution possesses a symmetry property

$$F(x; p, \rho) = 1 - F(1 - x; 1 - p, \rho)$$

6. Density of the Loss Distribution: The loss distribution has the density

$$f(x; p, \rho) = \sqrt{\frac{1 - \rho}{\rho}} e^{-\frac{1}{2\rho}\left\{\left[\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(p)\right]^2 - \frac{1}{2}\left[\Phi^{-1}(x)\right]^2\right\}}$$

which is unimodal at

$$L_{MODE} = \Phi\left(\frac{\sqrt{1-\rho}}{\sqrt{1-2\rho}}\Phi^{-1}(p)\right)$$

when

$$\rho < \frac{1}{2}$$

monotone when

$$\rho = \frac{1}{2}$$

and U-shaped when

$$\rho > \frac{1}{2}$$

7. Mean/Variance of the Loss Distribution: The mean of the distribution is

$$\mathbb{E}[L] = p$$

and the variance is

$$s^2 = \mathbb{V}[L] = \Phi_2(\Phi^{-1}(p), \Phi^{-1}(p), \rho)$$

where Φ_2 is the bivariate cumulative normal distribution function.

8. Usage in Cumulative p-Value Tables: The inverse of this distribution, that is, the -percentile value of L, is given by

$$L_\alpha = 1 - F(\alpha; 1 - p, \rho)$$

9. Skew and Kurtosis of the Distribution: The portfolio loss distribution is highly skewed and lepto-kurtic. The table below lists the α-percentile L_α expressed as a number of the standard deviations from the mean, for several values of the parameters. The α-percentiles of the standard normal distribution are shown for comparison.
10. Values of for the Portfolio Loss Distribution:

p	ρ	$\alpha = 0.9000$	$\alpha = 0.9900$	$\alpha = 0.9990$	$\alpha = 0.9999$
0.010	0.1	1.19	3.80	7.00	10.70
0.010	0.4	0.55	4.50	11.00	18.20
0.001	0.1	0.98	4.10	8.80	15.40
0.001	0.4	0.12	3.20	13.20	31.80
NORMAL		1.28	2.30	3.10	3.70

11. Non-normality of the Distribution: These values manifest the extreme non-normality of the loss distribution.
12. Sample Large Portfolio Loss Distribution: Suppose a lender holds a large portfolio of loans to firms whose pair-wise asset correlation is

$$\rho = 0.4$$

and whose probability of default is

$$p = 0.01$$

The expected portfolio loss is

$$\mathbb{E}[L] = 0.01$$

and the standard deviation is

$$s = 0.0277$$

If the lender wishes to hold the probability of default on his notes at

$$1 - \alpha = 0.001$$

he will need enough capital to cover 11.0 times the portfolio standard deviation. If the loss were normal, 3.1 times the standard deviation would suffice.

The Risk-Neutral Distribution

1. Actual World Loss Distribution Measure: The portfolio loss distribution given by

$$\mathbb{P}[L \leq x] = \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)$$

   is the actual probability distribution.
2. Probability of the Specified Loss Realization: This is the probability distribution from which to calculate the probability of loss of a certain magnitude for the purposes of determining the necessary capital or for calculating VaR.
3. Use in Structuring of CDO’s: This is the distribution to be used in structuring collateralized debt obligations; that is calculating the probabilities of loss and the expected loss for a given tranche.
4. Risk-Neutral World Asset Dynamics: For the purposes of pricing the traches, however, it is necessary to use the risk-neutral probability distribution. The risk-neutral distribution is calculated in the same way as above, except that the default probabilities are calculated under the risk-neutral measure

$$p^* = \mathbb{P}^*[A(T) < B] = \Phi\left(\frac{\log B - \log A - rT + \frac{1}{2}\sigma T^2}{\sigma\sqrt{T}}\right)$$

where r is the risk-free rate.

5. Risk-Neutral vs. Actual World: The risk-neutral probability is related to the actual probability of default by the equation

$$p^* = \Phi\left(\Phi^{-1}(p) + \lambda\rho_M\sqrt{T}\right)$$

where ρ_M is the correlation of the firm asset value with the market and

$$\lambda = \frac{\mu_M - r}{\sigma_M}$$

is the market price of risk.

6. Risk-Neutral Loss Distribution Measure: The risk-neutral portfolio loss distribution is then given by

$$\mathbb{P}^*[L \le x] = \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(x) - \Phi^{-1}(p^*)}{\sqrt{\rho}}\right)$$

7. Risk-Neutral Derivative Pricing: Thus, a derivative security – such as CDO tranche written against a portfolio – That pays at time T and amount $C(L)$ contingent on the portfolio loss is valued at

$$V = e^{-rT}\mathbb{E}^*[C(L)]$$

where the expectation is taken with respect to the distribution

$$\mathbb{P}^*[L \leq x] = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(p^*)}{\sqrt{\rho}}\right)$$

8. Tranche Pricing under this Measure: For instance, a default protection for losses in excess of L_0 is priced at

$$V = e^{-rT}\mathbb{E}^*[L - L_0] = e^{-rT}\left\{p^* - \Phi_2\left(\Phi^{-1}(p^*), \Phi^{-1}(L_0), \sqrt{1-\rho}\right)\right\}$$

The Portfolio Market Value

1. Forward Value for Expected Loss: So far, the treatment has been about the loss due to loan defaults. Now, suppose that the maturity date T_M is past the date T_H for which the portfolio is examined, i.e., the horizon date.

2. Impact of the Borrower Credit Quality: If the credit quality of the borrower deteriorates, the value of the loan will decline, resulting in a loss – this is often referred to as the loss due to *credit migration*.
3. MTM Change Impact on Portfolio Loss: This section investigates the distribution of the loss resulting from the changes in the MTM portfolio value.
4. Risk-Neutral Value of the Loan: The value of the loan at time 0 is the expected value of the loan payments under the risk-neutral measure

$$D = e^{-rT_M}(1 - Gp^*)$$

where G is the log given default, and p^* is the risk-neutral probability of default.

5. Value of the Loan at T_H: At time T_H the value of the loan is

$$D(T_H) = e^{-rT_H}\left\{1 - G\Phi\left(\frac{\log B - \log A(T_H) - r[T_M - T_H] + \frac{1}{2}\sigma[T_M - T_H]^2}{\sigma\sqrt{T_M - T_H}}\right)\right\}$$

6. Value of the Loss at T_H: Defining the loan loss L_i at time T_H as the difference between the risk-less value and the market value of the loan at one gets

$$L_i = e^{-r[T_M - T_H]} - D(T_H)$$

7. Alternative Definition for Loan Loss: This definition for loss is chosen purely for convenience. If the loss is defined in a different way – for instance, as a difference between

the accrued value and the market value – it will only result in a shift of portfolio loss distribution by a location parameter.

8. Explicit Expression for Loan Loss: The loss on the ith loan can be written as

$$L_i = a\Phi\left(b\sqrt{\frac{T_M}{T_M - T_H}} - X_i\sqrt{\frac{T_M}{T_M - T_H}}\right)$$

where

$$a = Ge^{-r[T_M - T_H]}$$

$$b = \Phi^{-1}(p) + \lambda\rho_M\frac{T_M - T_H}{T_M}$$

and the standard normal variables X_i defined over the horizon T_H by

$$\log A_i(T_H) = \log A_i(0) + \mu_i T_H - \frac{1}{2}\sigma_i {T_H}^2 + \sigma_i\sqrt{T_H}X_i$$

are subject to

$$X_i = Y\sqrt{\rho} + Z_i\sqrt{1-\rho}$$

9. Explicit Expression for Condition Loss: The conditional mean of L_i given Y can be calculated as

$$\mu(Y) = \mathbb{E}[L_i|Y] = a\Phi\left(b\sqrt{\frac{T_M}{T_M - \rho T_H}} - Y\sqrt{\frac{\rho T_H}{T_M - \rho T_H}}\right)$$

10. Convergence of Portfolio Expected Loss: Let L be the market value loss at time T_H of a loan portfolio with weights w_i.The losses conditional on factor Y are independent, and therefore the portfolio loss L conditional on Y converges to its mean value

$$\mathbb{E}[L|Y] = \mu(Y)$$

as

$$\sum_{i=1}^{n} {w_i}^2 \to \infty$$

11. Limiting Distribution of Portfolio Loss: The limiting distribution of loss L is then

$$\mathbb{P}[L \leq x] = F\left(\frac{x}{a}; \Phi(b), \rho\frac{T_H}{T_M}\right)$$

12. Form of the Limiting Distribution: It can thus be seen that the limiting loss of the portfolio distribution is of the same type as

$$F(x; p, \rho) = \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)$$

whether the loss is defined as a decline in the market value or the realized loss at maturity. In fact, the results of the section on the distribution of loss due to default are just a special case of this section for

$$T_M = T_H$$

13. Risk-Neutral Measure for MTM: The risk-neutral distribution for the loss due to market value change is given by

$$\mathbb{P}^*[L \le x] = F\left(\frac{x}{a}; p^*, \rho\frac{T_H}{T_M}\right)$$

Adjustment for Granularity

1. <u>Zero Variance Criterion for CLT</u>: The limiting portfolio loss distribution of the loss L is then

$$\mathbb{P}[L \leq x] = \mathbb{P}[\mu(Y) \leq x] = F\left(\frac{x}{a}; \Phi(b), \rho\frac{T_H}{T_M}\right)$$

relies on the convergence of the portfolio loss L given Y to its mean value $\mu(Y)$, which means that the conditional variance

$$\mathbb{V}[L|Y] \rightarrow 0$$

2. <u>Adjustment for Non-zero Variance</u>: When the portfolio is not sufficiently large for the law of large numbers to take hold, on needs to take into account the non-zero value of $\mathbb{V}[L|Y]$.
3. <u>Sum of Quadratic Weights</u>: Consider a portfolio of uniform credits with weights $w_1, \cdots, w_n$ and set

$$\delta = \sum_{i=1}^{n} {w_i}^2$$

4. <u>Conditional Variance of the Portfolio Loss</u>: The conditional variance of the portfolio loss L given Y is

$$\mathbb{V}[L|Y] = \delta a^2 \left\{\Phi_2\left(\Phi^{-1}(U), \Phi^{-1}(U), \sqrt{\frac{(1-\rho)T_H}{T_M - \rho T_H}}\right) - \Phi^2(U)\right\}$$

where

$$U = b\sqrt{\frac{T_M}{T_M - \rho T_H}} - Y\sqrt{\frac{\rho T_H}{T_M - \rho T_H}}$$

5. Unconditional Portfolio Mean and Variance: The unconditional mean and variance of the portfolio loss are

$$\mathbb{E}[L] = a\Phi(b)$$

and

$$\begin{aligned}\mathbb{V}[L] &= \mathbb{V}\big[\mathbb{E}[L|Y]\big] + \mathbb{E}\big[\mathbb{V}[L|Y]\big] \\ &= \delta a^2 \Phi_2\left(b, b, \frac{T_H}{T_M}\right) + (1-\delta) a^2 \Phi_2\left(b, b, \rho\frac{T_H}{T_M}\right) - a^2 \Phi^2(b)\end{aligned}$$

6. Approximation of the Unconditional Variance: Taking the first two terms in the tetrachoric expansion of the bivariate normal distribution function

$$\Phi_2(x, x, \rho) = \Phi^2(x) + \rho\phi^2(x)$$

where ϕ is the normal density function, on has approximately

$$\mathbb{V}[L] = \delta a^2 \phi^2(b) \frac{T_H}{T_M} + a^2 \Phi_2\left(b, b, \{\rho + \delta[1-\rho]\} \frac{T_H}{T_M}\right) - a^2 \Phi^2(b)$$

7. Moment Matching Loss Distribution Fit: Approximating the loan loss distribution by

$$F(x; p, \rho) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)$$

with the same mean and variance, one gets

$$\mathbb{P}[L \leq x] = F\left(\frac{x}{a}; \Phi(b), \{\rho + \delta[1-\rho]\} \frac{T_H}{T_M}\right)$$

8. Asymptotic Behavior for n, δ: This expression is in fact exact for both extremes

$$n \to \infty$$

$$\delta = 0$$

and

$$n = 1$$

$$\delta = 1$$

9. Adjustment for the Granularity of Portfolio: The loss distribution

$$\mathbb{P}[L \leq x] = F\left(\frac{x}{a}; \Phi(b), \{\rho + \delta[1 - \rho]\}\frac{T_H}{T_M}\right)$$

provides an adjustment for the *granularity* of the portfolio. In particular, the finite portfolio adjustment to the distribution of the gross loss at the maturity date is obtained by setting

$$T_H = T_M$$

and

$$a = 1$$

to yield

$$\mathbb{P}[L \leq x] = F(x; \Phi(b), \rho + \delta[1 - \rho])$$

Summary

1. <u>Convergence of Portfolio Loss Distribution</u>: This chapter has shown that the distribution of loan portfolio loss converges, with increasing size, to the limiting type given by

$$F(x; p, \rho) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)$$

   This means that the distribution can be used to represent loan loss behavior of large portfolios.
2. <u>Spot/Forward Pricing of Loss</u>: The loan loss can be a realized loss on the loans maturing prior to the horizon date, or a market value deficiency on the loans whose term is longer than the horizon period.
3. <u>Derivation of the Limiting Loss Distribution</u>: The limiting probability distribution of the portfolio losses has been derived under the assumption that all loans in the portfolio have the same maturity, the same probability of default, and the same pair-wise correlation of borrower assets.
4. <u>Applicability to more General Portfolios</u>: Surprisingly, however, numerical simulations show that the family

$$F(x; p, \rho) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)$$

appears to provide a reasonably good fit to the tail of the loss distribution for more general portfolios.

5. Comparison with Monte Carlo Simulations: To illustrate this point, Vasicek (2002) provide the results of Monte-Carlo simulations on an actual bank portfolio.
6. Loan Portfolio Characteristics - Component Weight: The portfolio consisted of 479 loans in amounts ranging from 0.0002% to 8.7% with

$$\delta = 0.039$$

7. Loan Portfolio Characteristics - Maturity/PD: The maturities ranged from 6 months to 6 years and the default probabilities from 0.0002 to 0.064.
8. LGD + Asset Returns Common Factors: The loss give default averaged 0.54. The asset returns were generated with 14 common factors.
9. Simulated vs. Analytical CDF Comparison: Vasicek (2002) compares in detail the simulated cumulative distribution function of the loss in one year against the fitted limiting distribution function.

References

- Pykhtin, M. and A. Dev (2002): Credit Risk in Asset Securitizations: An Analytical Model *Risk* **15 (5)** 16-20
- Vasicek (1987): *Probability of Loss on Loan Portfolio* **KMV Corporation**
- Vasicek (1991): *Limiting Loan Loss Probability Distribution* **KMV Corporation**

- Vasicek (2002): The Distribution of Loan Portfolio Value

Vasicek Model Default Risk Simulation

Theoretical Background

1. Portfolio Loss Distribution Model Used: One-factor Gaussian Copula Credit Loss Model.
2. Incorporation of Asset Default Correlation: Default is driven by a common market factor Y - the systemic shock – and an idiosyncratic Gaussian Process Z_i:

$$X_i = Y\sqrt{\rho} + Z_i\sqrt{1-\rho}$$

3. Vasicek Expression for Loss Distribution: The famous Vasicek closed-form formula for calculating the loss dist4ribution under the Gaussian Copula framework:

$$VaR_\alpha = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)$$

4. Assumptions of the Vasicek Model:
   - All exposures are of the same amount – homogenous portfolio
   - All exposures have the same probability of default

- All exposures have a common set correlation

Model Implementation

1. Correlation to the Common Factor:

$$Factor_1 = \sqrt{\rho}$$

2. Correlation to Idiosyncratic Factor:

$$Factor_2 = \sqrt{1-\rho}$$

3. Default Threshold:

$$d_t = \Phi^{-1}(1-p)$$

4. Systemic Gaussian Process:

$$Y = \Phi^{-1}\big(rand(\cdot)\big)$$

5. Idiosyncratic Gaussian Process:

$$Z_i = \Phi^{-1}\big(rand(\cdot)\big)$$

6. Asset Value Default Incidence Simulation: If

$$Factor_1 \cdot Y + Factor_2 \cdot Z_i > d_t$$

   the exposure is treated as having defaulted.
7. Default Simulation across Counter-Parties: The systemic and the idiosyncratic variables for many paths are simulated across counter-parties, and for each, the total loss is summed in that exposure.
8. Risk Capital as Target Loss Percentile: The risk capital is the picked for the target percentile from these sums of total losses across all the simulation paths.

Sample Simulation Results

1. Uniform Exposure across all Ratings:

Uniform 1MM Exposure Size	Confidence	90%

Row Labels	No Hedge Cumulative Exposure	2Y PD	Correlation	Tail Loss
1	265,000,000	0.001050	0.238213433	661,866
2	264,000,000	0.004926	0.231836920	3,305,600
3	311,000,000	0.003169	0.231836920	2,477,183
4	276,000,000	0.012852	0.215292799	8,937681
5	311,000,000	0.065197	0.160048020	42,807,422
6	313,000,000	0.189831	0.121497866	100,943,573
7	260,000,000	0.481051	0.120000001	172,562,865
Grand Total	2,000,000,000			331,696,209

Average of Simulations	332,687,000
Simulation Difference Standard Deviation	6,637,385
Simulation Difference Standard Deviation %	2.00%
# > 0 Difference %	56
# < 0 Difference %	44

- Upper table is the closed for result, bottom table is the simulated result.
- As expected, the two results are close.

2. <u>Two Outsized Exposures in AA</u>:

Uniform 1MM Exposure Size			Confidence	90%
Row Labels	**No Hedge Cumulative Exposure**	**2Y PD**	**Correlation**	**Tail Loss**
1	265,000,000	0.001050	0.238213433	661,866
2	18,262,000,000	0.004926	0.231836920	228,662.386
3	311,000,000	0.003169	0.231836920	2,477,183
4	276,000,000	0.012852	0.215292799	8,937681
5	311,000,000	0.065197	0.160048020	42,807,422
6	313,000,000	0.189831	0.121497866	100,943,573
7	260,000,000	0.481051	0.120000001	172,562,865
Grand Total	19,998,000,000			557,052,995

Average of Simulations	334,552,000
Simulation Difference Standard Deviation	6,696,572
Simulation Difference Standard Deviation %	1.26%
# > 0 Difference %	0
# < 0 Difference %	100

- Upper table is the closed for result, bottom table is the simulated result.
- Closed form doesn’t work well in such a case.

3. Two Bigger Exposures in B:

Uniform 1MM Exposure Size			**Confidence**	**90%**
Row Labels	**No Hedge Cumulative Exposure**	**2Y PD**	**Correlation**	**Tail Loss**
1	265,000,000	0.001050	0.238213433	661,866
2	264,000,000	0.004926	0.231836920	3,305,600
3	311,000,000	0.003169	0.231836920	2,477,183
4	276,000,000	0.012852	0.215292799	8,937681
5	311,000,000	0.065197	0.160048020	42,807,422
6	701,000,000	0.189831	0.121497866	226,074,903
7	260,000,000	0.481051	0.120000001	172,562,865
Grand Total	2,000,000,000			456,827,539

Average of Simulations	514,818,000
Simulation Difference Standard Deviation	12,685,022
Simulation Difference Standard Deviation %	2.78%
# > 0 Difference %	100

# < 0 Difference %	0

- Upper table is the closed for result, bottom table is the simulated result.
- Simulated rating captures more idiosyncratic risk than the closed-form, indicating the importance of PD (rating) of the exposure.

Implication of the Simulation Tests

1. Addressing Vasicek Drawbacks using Simulation:
   - Vasicek expression cannot be used blindly to get the systemic default risk
   - Alternative ways of estimating systemic default risk can be done by breaking down portfolios into infinite number of smaller exposures and run large simulation
   - Drawback is that the implementation will be more complex and the run more time consuming
2. HHI as Portfolio Homogeneity Metric:
   - One indicator can help test the homogeneity of the portfolio
   - Introducing the Herfindahl-Hirschmann Index (HHI)

$$HHI = \sum_{i=1}^{n} \left(\frac{EAD_i}{\sum_{j=1}^{n} EAD_j} \right)^2 \rightarrow 0$$

as

$$n \to \infty$$

- For an infinitely large homogenous portfolio, the HHI should approach 0
- Our exposure without hedge HHI is 0.9% with hedge is 0.4%

Limitations of the Model

1. Model Parameter Calibration Limitation Description: in regards to calibration of model parameters, the probabilities of default (PD's) and the default correlations are calibrated using the Vasicek (2002) Credit Model whereas the Loss Given Default (LGD's) are calibrated using historical observations. The calibration of LGD's and PD's is based on limited historical data and the assumption of the homogenous base portfolio. The calibrated model inputs are difficult to observe and may change with time. The model parameters may also differ between alternative energy sources or sampling methodologies.
2. Model Parameter Calibration Severity Assessment: The impact of this limitation on the model parameter is limited. The model will be replaced in the near future by a six-factor model similar to the one commonly in use in CCAR IDL.
3. Model Parameter Calibration Remediation Action: The calibration of the model inputs, i.e., PD's and LGD's, should be benchmarked with the historical data available from third party vendor's such as S&P's and Moody's on an ongoing annual basis as part of Ongoing Performance Analysis (OPA) and the calibration process for the model parameters should be reviewed.

4. Ratings Based Calibration Limitation Description: The PD's constitute a key model input, provided by the BHC credit model. Default probabilities are often calculated by risk rating so that all names in the corresponding ratings category take the expected default probability. As a result, the default probabilities may not be distinguished between industries and sectors. In addition, the calculation of the default risk for the CVA is based on a single factor Gaussian copula for the estimation of the correlation among the counter-parties. While this may be sufficient for the Risk Capital (RC) estimation as an industry standard, it is a simplistic approach. As a result the model may not capture the diversity of different markets and industry sectors. Also, it may not reflect the accurate correlation between the counter-parties.
5. Ratings Based Calibration Severity Assessment: This limitation is related to model framework and assessment. The impact on the model output is potentially moderate. A multi-factor model, similar to the one often used in CCAR IDL, is a candidate for replacing this.
6. Ratings Based Assessment Remediation Action: The multi-factor model, which is capable of diversifying the correlation between different industries, sectors, and regions, is a viable candidate for replacement. More granular and industry representative default probabilities may be considered in the model development.
7. Input Sensitivity Tests Limitation Description: To further assess the efficacy and the impact of the model, the following sensitivity analysis must be performed, preferably on a recently available data:
    - Evaluate the changes in the risk capital with respect to the PD's.
    - Evaluate the changes in the risk capital with respect to the changes in the default correlation input parameters.
8. Input Sensitivity Tests Severity Assessment: This limitation is to recognize more tests, and does not indicate model weakness. A multi-factor mode, similar to the one used in CCAR IDL, is a candidate for replacing this. Additional tests will need to be performed for this new model.
9. Input Sensitivity Tests Remediation Action: The sensitivity tests should be performed against the mode input parameters – e.g., PD's and different correlations – on a recently available data, and the performance should be assessed.

References

- Vasicek (2002): The Distribution of Loan Portfolio Value

Market Place Lending Credit Model Methodology

Overview of Credit Model Methodology

1. Risk Estimation in MPL Platforms: DROP (2015) has introduced version 2.0 of the Credit Model for estimating prepayment and default risk on a pool of US unsecured consumer loans issued by a select set of marketplace lending ("MPL") platforms.
2. Valuation, Prepay, and Default Drivers: This extends the version 1.0 Credit Model published previously in August 2015. In both cases the goal is to enable the market participants understand the drivers behind valuation, prepayment, and default risk across their holdings in a transparent and robust manner.
3. Usage of Markov Logit Models: A main objective is to apply the rigor and forward looking tools of mortgage analytics, namely the Multinomial Logit and the Markov Chain approach to modeling the MPL collateral.
4. Simulation of Prepay and Default: The second objective is to have the ability apply simulations to prepayment and default rates, and ultimately the cash flows.
5. Usage of Statistical Learning Techniques: The final objective is to use machine learning techniques such as regularization and cross-validation to improve the robustness of the modeling.

Credit Methodology – Purpose and Introduction

1. MPL Growth/Institutional Investors Participation: The dramatic growth in Marketplace Lending (MPL) has been paired with the transition of the consumer and the SME credit to the

capital markets. Institutional investors are increasingly funding credit risk in whole loan, structured credit, or warehouse formats.

2. MPL Risk and Analytics Tools: A new set of analytics and risk pricing tools are necessary to enable institutional capital efficiently access, price, and exchange risk. A core component of risk pricing is independent 3rd party credit models that forecast cash flows on pools of loans.
3. Improvement in Liquidity and Transparency: By releasing a new credit model, the aim is to improve transparency and standards in MPL by providing risk management tools promoting independent pricing and whole loan and ABS market liquidity.
4. Valuation and Risk MPL Common Language: Independent credit models increase activity by allowing market participants to speak a common language in reference to valuation and risk.
5. Independent 3rd Party Credit Models: Several sources have described on various occasions the robust scaling and growth in the US and global marketplace lending sector. Large institutional investors require a 3rd party credit model to understand their credit risk exposures, and to participate in size.
6. Short Duration, High Yield Risk: Further, credit spreads have generally tightened since the 2008 credit crisis in a backdrop of quantitative easing, healthy global economy, and stringent regulation. As a result, investors see MPL offering an attractive short-duration, high-yield credit risk compared to the alternatives.
7. Engagement in the MPL Space: Investors have been engaged in the market through a variety of activities including, but not limited to:
    a. Investing in marketplace lending platform equity
    b. Facilitating funding for the platforms via a provision of credit facilities
    c. Investing in MPL securitizations
    d. Directly lending to borrowers.
8. Proliferation among MPL Asset Classes: In addition, there has been a proliferation of marketplace lenders across a multitude of asset classes including consumer, purchase finance, education finance, real estate, merchant cash advance, and small businesses.
9. Asset Pricing and Regulation Obligation: As a consequence, institutional investors, diversified financial services firms, and funding providers require an independent 3rd party to help them price their holdings or satisfy other regulatory obligations.

10. MPL Valuation Standards Methodology Enhancement: Finally it is imperative that 3rd parties commit to improving and enhancing their standards and methodologies to serve what is a rapidly growing and evolving market. The 2.0 Credit Model is an attempt at enhancements over several key areas foe projections of cash flows on historical data above 1.0.
11. Additional Loan Specific Risk Factors: 2.0 addresses the need to incorporate additional loan specific measures of risk. Version 1.0 segments the loan by 6 factors:
    a. The originator
    b. Loan credit quality grade as provided by the originator
    c. Origination vintage
    d. Loan term
    e. Loan status
    f. Loan age
12. Incorporating Macro-economic Driving Factors: 2.0 provides a structure for incorporating macro-economic factors that drive the estimates of default and prepayment.
13. Reducing Dependence on Recent Issuance: Reducing the reliance on the most recently issued set of loans to drive expectation of prepay and default is another objective (1.0 applied the default and prepay experience of the most recently issued cohort corresponding to the risk factors above).
14. Usage of Statistical Learning Techniques: Version 2.0 applies advanced statistical and machine learning techniques to develop a predictive model for prepay and default. Thus Credit Model Version 2.0 aims to address and overcome the shortcomings from the Model 1.0 above, and bring rigorous techniques to an expanding asset class with a growing investor base.

Scope of the 2.0 Model

1. Loans Originated by Lending Club: For the purposes of demonstration, PeerIQ (2015) illustrate the construction and performance of Model Version 2.0 on public data from loans originated by Lending Club ("LC") with reporting months from 1 January 2010 to 1 July 2015.
2. The Data Model: PeerIQ's data model is proprietary and unified in its methodology for cleaning, enriching, and housing data across all MPL originators, and gets expanded as additional asset classes and originators are on-boarded.
3. Similarity with Lending Club Model: Although DROP does tailor the Model to specific data classes and originators, the methodology and the model structure is substantially similar to the Lending Club model.

Data Model Construction Rules

1. Amendments to Originator Generated Payments: DROP has made specific amendments (or transformations) on the raw originator-generated payments and balances for loans.
2. Consistency and Accuracy across Cohorts: These changes have been made in a rule-based fashion based on conversations with marketplace lenders to ensure the calculation of the cohort payments and balances in a consistent and accurate manner.
3. Reconciliation between Borrowers and Originators: Many of the above rules help reconcile between a borrower snapshot file (the borrower file) and a cumulative payments file (the payments file) published by the originators.

Loan Data Quality Rules

1. Inaccurate Originator Loan Level Record: Some originators publish inaccurate records for loans that have previously charged or paid off. These records are excluded from the cleaned data set and calculations.
2. Identification and Removal of Duplicates: If the loan has more than one record with the same originator loan ID, loan month, month on book, and outstanding principal BOP balance, it is assumed that the subsequent records are duplicates and that they must be removed.
3. Combine Payments for a Given Month: If the loan as multiple payments for a given month, then these must be combined to form a single payment. The formula applied is: combine all rows where count of loan month, originator loan ID > 1.
4. Entry for Maximum Loan Month: Each loan that has a non-zero EOP balance for a month prior to the maximum loan month should have a record for the maximum loan month. For example, if the maximum loan month on file is February 2015, and the loan is current in January 2015 but does not have a February 2015 record, a record will need to be created.
5. Loan Month Issue Date Consistency: Each loan ID should have a record where the loan month equals the issue date. Further, a record must exist for each loan between the issue date and the current file date, charge off date, or fully paid date, whichever is earlier.
6. Loan Age/Days Past Due: Days Past Due value should never be negative. The expected loan age is calculated as the loan month minus the issue date.
7. Charge Off/Fully Paid Fields:
   a. All instances where the

$$CO\ Amount\ != 0$$

      should have the charge off flag set to 1.
   b. All loans that have a status of fully paid should have an end-of-period balance > 0.
   c. BOP principal minus principal received and charge-off amount should equal to 0.
8. Principal, Interest, and Fee Payments:
   a. Interest paid for the given loan on a given month should equal between the borrower and the originator

   b. Amount paid should equal the sum of principal, interest, and fees paid
   c. All principal payments, interest payments, and fee payments should be positive.
9. Field Unchanged through the Loan Life: The following fields should remain unchanged and populated through the loan life: loan purpose, loan interest rate, loan grade, loan term, loan state, original principal, issuance date.
10. Consistency of the Recovery Fields: All recoveries should be positive, and should be recorded at the month the loan charges off.

Lending Club Loan Level Data

1. Lending Club Loan Types Considered: As a starting point for the demonstration of the modeling approach, DROP uses the loan level public data from Lending Club. As such, the loan products considered are fixed rate, fixed term, fully amortizing 36 month and 60 month loans issued by Lending Club.
2. Number and Size of Loans: In all, PeerIQ (2015) uses over 9 million loan months of Lending Club data in constructing the model. The table below contains high level descriptive statistics for select items from the dataset.
3. Descriptive Statistics for LC Data: Source: PeerIQ Research

Field	Mean	Standard Deviation	Minimum	Maximum	First Quartile	Median	Third Quartile
Age (Months on Balance)	10.1	8.9	0.0	60.0	3.0	8.0	15.0
Vintage	February 2013	NA	February 2007	February 2015	April 2012	March 2013	February 2014

Original Principal	$14,254	$8,254	$500	$35,000	$8,000	$12,000	$20,000
Monthly Gross Income	$6,066	$4,599	$250	$725,549	$3,750	$5,167	$7,333
Term (Months)	42.7	10.8	36.0	60.0	36.0	36.0	60.0
Coupon	13.7%	4.3%	5.3%	29.0%	10.6%	13.5%	16.3%
FICO Origination	699	31	612	847	677	692	717
DTI (ex-mortgage)	16.6%	7.8%	0.0%	39.0%	11.0%	16.0%	22.0%
Total Borrower Accounts	25	11	1	162	16	23	31
Revolving Utilization Rate	57%	24%	0%	892%	40%	58%	75%
Inquiries in Last 6 Months	0.9	1.2	0.0	33.0	0.0	0.0	1.0
DQ Accounts in Last 2 Years	0.3	0.8	0.0	39.0	0.0	0.0	0.0

Months since Last DQ	34	22	0	188	16	31	50
Months since Last Public Record	76	29	0	129	55	79	102
Total Open Credit Lines	11	5	0	90	8	10	14

4. Period of LC Loan Origination: Overall, the sample used contains 245,243 distinct loans originated between February 2007 and February 2015. Average loan size is a little over $14,000, varying between $500 and $35,000 for Lending Club.
5. Variation among the Underwriting Parameters: There is also considerable variation among other under-writing information, including DTI and revolving debt utilization rates, for example.
6. Defaults/Prepayments by Origination Year: Ultimately the goal is derive insight into termination events (defaults and prepays) from loan pools, and the table below lists some simple summaries of defaults and prepayments by origination year.
7. LC Prepay and Default Exits: Source: PeerIQ Research.

Origination Year	**Origination Volume ($mm)**	**Prepays ($mm)**	**Cumulative Defaults ($mm)**	**Cumulative Defaults to Date (%)**
2010	132	26	12	8.82%
2011	262	62	27	10.49%
2012	718	164	75	10.39%

2013	1,982	405	135	6.83%
2014	3,504	387	88	2.51%

8. Peaking of Defaults by Vintage: The table above shows that the defaults in the current cycle have peaked for the 2011 and the 2012 vintages.

Loan Credit Model Implementation

1. Discrete Loan Level Status Codes: The approach to modeling starts with the observation that at a given point in time, a loan has several status codes, or 'state's: current, delinquent, fully prepaid, or charged off.
2. Explicit Transition between Loan States: The key conceptual idea is to model transitions between the various states explicitly. For example, a loan may move from a state of 'current' to '30-day delinquent', or move further down the delinquency queue (e.g., move from 30 day to 60 day delinquent). Alternatively, a loan may 'cure' and move from a status of current to delinquent.
3. Exit from the Transition Graph: The only exits for a loan from this network are a transition to prepay or default, which are of course the end statuses we are most interested in. The graph table below illustrates the transition network.
4. Directed State Transition Network Graph: Source: PeerIQ Research.

State	**C**	**P**	**D3**	**D6**	**D6+**	**D**
C	Y	Y	Y	N	N	Y
P	N	Y	N	N	N	N

D3	Y	Y	Y	Y	N	N
D6	N	N	Y	Y	Y	Y
D6+	N	N	N	Y	Y	Y
D	N	N	N	N	N	Y

5. State Transition Network Graph Annotation: Y's indicate directionality of the possible transitions. Note that prepay and default here are "absorbing" states, i.e., states from which exits are not possible. For modeling purposes, and due to data considerations, the 90-180 day delinquent states have been combined into the D6+ category.
    a. C => Current
    b. P => Prepay
    c. D3 => 30 day Delinquent
    d. D6 => 60 day Delinquent
    e. D6+ => 90 and more days Delinquent
    f. D => Default
6. Cross-State Transition Probabilities: Mathematically the probabilities of these moves can be represented in a transition matrix thereby describing the propensity of the borrower to move from one state to another.
7. State -> State Transition Setup:

$$P_{i,j}(t) = \begin{bmatrix} P_{C,C}(t) & \cdots & P_{C,D}(t) \\ \vdots & \ddots & \vdots \\ P_{6+,C}(t) & \cdots & P_{6+,D}(t) \end{bmatrix}$$

Each entry in the matrix represents the probability of the borrower moving from the row state to the column state in a particular month. For example, $P_{C,C}(t)$ represents the probability of the borrower moving from current to current (that is, staying current) over month t, and so on.

8. Transition Probabilities from MNL Frameworks: The next step is to decide on the parameterization of the entries in the transition matrix, i.e., how does one model the transition probabilities $P_{i,j}(t)$ (commonly referred to as the 'roll rate's?).
9. Transition Matrix in Markov Process: The MNL framework has features prominently in the mortgage modeling literature in modeling prepay and default. Such a modeling setup is known as the *Markov* process, which is practice means that each monthly observation of a loan's transitions is independent of any prior observations.
10. MNL Formulation in Logit Framework: Under the logit assumption inherent in MNL, the transition probabilities assume the following form:

$$P_{i,j}(t) = \frac{e^{\overrightarrow{\beta_0} + \overrightarrow{\beta_1} \cdot \overrightarrow{X_{\iota,J}}(t)}}{1 + e^{\overrightarrow{\beta_0} + \overrightarrow{\beta_1} \cdot \overrightarrow{X_{\iota,J}}(t)}}$$

11. Loan Variables as Regressor Factors: Here $\overrightarrow{X_{\iota,J}}$ represents the matrix of regressors (or predictors) observed at a particular time (such as consumer credit variables, loan age, cohort, for example, and macro-economic variables such as inherent rates). Note that we are not restricted to using the same set of predictors for every transition.
12. Normalization under the MNL Framework: The parametrization above implies that the probabilities $P_{i,j}(t)$ will lie between zero and one, as they should. Further, since the probabilities across a given row should sum up to one, there will always be a status s for which the probabilities are determined as

$$P_{i,s}(t) = 1 - \sum_j P_{i,j}(t) = \frac{1}{1 + \sum_k e^{\overrightarrow{\beta_0} + \overrightarrow{\beta_1} \cdot \overrightarrow{X_{\iota,k}}(t)}}$$

13. Reduction of the Transition Probabilities: In estimating the model, one does not generally estimate all the probabilities from one state to another, as this would make the models unnecessarily complex, especially if the transitions are rare.
14. Current to Charge-Off Transition: While there are various reasons a loan can theoretically transition from current to charge-off (skipping intermediate statuses such as delinquency), such as due to the death of the borrower, these tend to be rare empirically.
15. Current-to-Default Transition Likelihood: This, it is quite intuitive that the probability of going from a state of current to a state of charge-off in a month should be quiet low. As shown in PeerIQ (2015) this monthly transition rate can be seen to be at most 0.03% for 60 month loans. Therefore this probability is not estimated.
16. Estimation of the Sparse Matrix: Thus, a sparse matrix, which is a subset of all the possible transitions – is estimated. The grid below gives the transitions that are estimated in the model, and those that are not.
17. Sparse Transition Matrix and Determinants: Source: PeerIQ Research.

States	**C**	**D3**	**D6**	**D6+**	**D**	**P**
C	Y	Y	N	N	N	Y
D3	Y	Y	Y	N	Y	Y
D6	N	Y	Y	Y	N	N
D6+	N	N	Y	Y	Y	N

Clearly the number of transitions to be estimated will vary depending on the originator and the observed frequency.

18. Transition Probabilities Using Separate Models: Since there is no restriction to using the same set of regressors for each probability, it is instructive to think of each of the entries estimated in the sparse matrix as separate models.

DROP Credit Model Selection Methodology

1. Cross Validation Based Model Selection: DROP applies a rigorous out-of-sample based testing procedure to estimate each of the models discussed above. The procedure for the model selection is described as follows.
2. Candidate Variables for Model Selection: DROP has identified a set of candidate variables for selection. Those variables and their data types are listed in the table below.
3. Candidate Variables and their Types: Source: PeerIQ Research.

Variable	Implementation Type
Age (Months Remaining)	Evenly Spaced Linear Splines
Term (36 or 60 months)	Categorical Variable
Interaction Variable (Age Spline * Term)	Linear Spline * Categorical Variable
DTI	Continuous Variable
FICO at Origination	Categorical Variable
Vintage (Origination Year)	Categorical Variable
Seasonality (Month of Year)	Categorical Variable
Original Loan Size Bucket	Categorical Variable
Coupon Stack Bucket	Categorical Variable
Loan Purpose	Categorical Variable
Employment Length	Categorical Variable
Inquiries in the Past 6 Months	Untransformed

Monthly Gross Income	Untransformed
Total Outstanding Accounts	Untransformed
Revolving Credit Utilization	Untransformed
Delinquent Accounts in last 2 Years	Untransformed
Total Open Credit Lines	Untransformed

4. Model Variables from Candidate Regressors: Altogether the combinations of the categorical and the numerical variables generate 86 candidate regressors from which the best choice for each transition probability is optimized. The details of the procedure for model variable selection follow.
5. Likelihood Estimation for Traditional MNL: Variable selection was achieved via a regularized logistic regression procedure. In traditional logistic regression one typically optimizes for the values of $\overrightarrow{\beta_0}$ and $\overrightarrow{\beta_1}$ in the equation above. Applying maximum likelihood across borrowers, the formulation for the transition probability between states (for e.g., borrowers moving from current to prepay) follows, as shown below.
6. Example Current -> Prepay MLE Setup:

$$Max\left[L\left(\overrightarrow{\beta_0},\overrightarrow{\beta_1}\right)\right] = \sum_{m=1}^{N} \log P_{c,p,m}$$

where $P_{c,p,m}$ represents the probability of the m^{th} borrower moving from current to prepay (dropping time subscripts for convenience) with the probabilities given from

$$P_{i,j}(t) = \frac{e^{\overrightarrow{\beta_0}+\overrightarrow{\beta_1}\cdot\overrightarrow{X_{i,j}}(t)}}{1 + e^{\overrightarrow{\beta_0}+\overrightarrow{\beta_1}\cdot\overrightarrow{X_{i,j}}(t)}}$$

7. Penalty Based Regularized Logistic Regression: Regularized logistic regression is similar to the equation above, but it imposes a penalty on $\overrightarrow{\beta_0}$ and $\overrightarrow{\beta_1}$ to ensure that the coefficients that are not predictive of the transition probability are not unnecessarily added (and thus eliminated from the regression).
8. Current Prepay Penalized MLE Setup:

$$Max\left[L\left(\overrightarrow{\beta_0},\overrightarrow{\beta_1}\right)\right] = \sum_{m=1}^{N} \log P_{c,p,m} + \alpha \sum_{j=1}^{p} \left|\beta_j\right|$$

   where β_j represents the set of predictors in $\overrightarrow{\beta_1}$, α represents the parameter that determines the strength of the penalty function.

9. Optimal Penalty Loading Selection Algorithm: The penalty loading parameter α is selected using the following step:
   a. Select a possible range of values for α
   b. For each value of α in the range above, train the logistic regression model on a training dataset, and test the predictive power of the model on a separate test dataset
   c. The test data is defined by sampling every third observation to minimize the risk of over-fitting the data
   d. Choose the value of α which results in the 'best' performance on the testing dataset
10. Classifier Performance Using Sample AUC: The AUC is a well-known measure of predicting accuracy of a classification technique (Receiver Operating Characteristic (Wiki)). Thus, the 'best' in the algorithm above is decided by looking at that value of α that gives the best out-of-sample AUC, e.g., on classifying borrowers who move from current to prepay in a given month.

Empirical Results – Regressor Contribution Weights

1. C -> P Predictor Loading Strength Order:

Strong Positive	Strong Inverse	Weak/Neutral
a. Age_k0.0	a. Revolving_utilization	a. Coupon_20+
b. Total_accounts	b. Total_open_credit_lines	b. Cohort_2014
c. Inq_6m	c. Age_k0.0_term_60	c. Cohort_2013
	d. Age_k18.0	d. Fico_750_800
	e. Dti_ex_mortgage	e. Lp_debt_consolidation
		f. Coupon_15-20
		g. Cohort_2015
		h. Mo_3
		i. Mo_5
		j. Mo_2
		k. Mo_4
		l. Size_15_20
		m. Mo_10
		n. El_10+years
		o. El_<1year
		p. Mo_8
		q. Mo_9
		r. Mo_11
		s. Size_gte20
		t. Fico_650_700
		u. Lp_small_business
		v. Coupon_<10
		w. Cohort_2011

		x. Cohort_2009 y. Cohort_2008 z. Cohort_2010 aa. El_n/a bb. Mo_12 cc. Term_60 dd. Age_k12.0

2. C -> D3 Predictor Loading Strength Order:

Strong Positive	**Strong Inverse**	**Weak/Neutral**
a. Age_k0.0 b. Inq_6m	a. Age_k6.0 b. Monthly_gross_income c. Age_k12.0	a. Coupon_20+ b. Coupon_15-20 c. Dti_ex_mortgage d. Dq_accounts_past_2_years e. Lp_small_business f. Lp_educational g. Cohort_2008 h. Lp_moving i. El_n/a j. Mo_10 k. Lp_other l. Fico_650_700 m. Mo_7 n. Lp_medical o. El_<1year p. Mo_6 q. Mo_11 r. Size_5_10

		s. Mo_9 t. El_10+years u. Lp_major_purchase v. Lp_debt_consolidation w. Mo_3 x. Size_lte5 y. Cohort_2015 z. Mo_2 aa. Mo_5 bb. Fico_750_800 cc. Lp_wedding dd. Cohort_2014 ee. Total_accounts ff. Fico_800_850 gg. Term_60 hh. Lp_home_improvement ii. Age_k24.0 jj. Cohort_2012 kk. Cohort_2013 ll. Age_k24.0term60 mm. Coupon_<10 nn. Total_open_credit_lines oo. Revolving_utilization

3. D3 -> D3 Predictor Loading Strength Order:

Strong Positive	**Strong Inverse**	**Weak/Neutral**
a. Age_k0.0 b. Dq_accounts_past_2_years	a. Cohort_2015 b. Cohort_2014	a. Coupon_15-20 b. Size_15_20

c. Cohort_2009	c. Age_k24.0term_60	c. Fico_650_700
d. Cohort_2008	d. Cohort_2013	d. El_4years
e. Cohort_2010	e. Size_lte5	e. Mo_7
f. Revolving_utilization	f. Total_open_credit_lines	f. Mo_10
g. Size_gte20	g. Age_k18.0term_60	g. Mo_8
h. Coupon_20+	h. Cohort_2012	h. Fico_750_800
	i. Term_60	i. Mo_12
	j. Size_5_10	j. El_3years
	k. El_7years	k. Mo_9
		l. Mo_5
		m. Coupon_<10
		n. Mo_11
		o. Lp_credit_card
		p. El_5years
		q. El_6years
		r. El_8years

4. <u>D3 -> C Predictor Loading Strength Order</u>:

Strong Positive	**Strong Inverse**	**Weak/Neutral**
a. Monthly_gross_income	a. Total_open_credit_lines	a. Revolving_utilization
b. Dq_accounts_past_2_years	b. Cohort_2014	b. Age_k12.0term_60
	c. Cohort_2015	c. Mo_3
	d. Dti_ex_mortgage	d. Mo_4
	e. Cohort_2013	e. Lp_moving
		f. Fice_650_700
		g. Size_lte5
		h. Lp_car
		i. Cohort_2008

		j. Lp_medical k. Cohort_2010 l. Size_5_10 m. Coupon_<10 n. Size_gte20 o. Cohort_2009 p. Lp_other q. Age_k12.0 r. Lp_debt_consolidation s. El_7years t. Mo_5 u. El_5years v. El_8years w. Size_15_20 x. El_9years y. Mo_6 z. Inq_6m aa. El_10+years bb. Mo_12 cc. Mo_7 dd. Term_60 ee. Cohort_2012 ff. Age_k30.0 gg. Mo_11 hh. El_n/a

5. <u>D3 -> P Predictor Loading Strength Order</u>:

Strong Positive	**Strong Inverse**	**Weak/Neutral**

a. Total_accounts b. Age_k6.0 c. Monthly_gross_income d. Fico_800_850 e. Cohort_2009 f. El_10+years g. El_8years h. El_5years i. Cohort_2015 j. El_6years	a. Age_k30.0 b. Age_k0.0term_60 c. Dti_ex_mortgage d. Lp_car e. Mo_8 f. Lp_small_business g. Fico_700_750 h. Size_lte5	a. Size_gte20 b. Mo_10 c. Lp_credit_card d. Mo_6 e. Mo_7 f. Cohort_2014 g. Cohort_2013 h. Size_15_20 i. Cohort_2010 j. Term_60

6. <u>D3 -> D Predictor Loading Strength Order</u>:

Strong Positive	**Strong Inverse**	**Weak/Neutral**
a. Age_k0.0 b. Total_accounts c. Age_k6.0 d. Dti_ex_mortgage e. Size_lte5 f. Cohort_2013 g. Cohort_2014 h. El_10+years	a. Coupon_15_20 b. El_3years c. Coupon_20+	a. Size_5_10 b. Cohort_2012 c. Mo_5 d. Mo_6 e. El_7years f. Fico_700_750 g. Term_60 h. El_<1year i. Mo_12 j. Mo_2 k. Size_15_20 l. Size_gte20 m. El_4years

7. <u>D6 -> D6 Predictor Loading Strength Order</u>:

Strong Positive	Strong Inverse	Weak/Neutral
a. Age_k0.0	a. Cohort_2014	a. Cohort_2009
b. Age_k6.0	b. Cohort_2015	b. Coupon_15-20
c. Revolving_utilization	c. Cohort_2013	c. El_<1year
d. Coupon_20+	d. Cohort_2012	d. Fico_750_800
e. Total_accounts	e. Cohort_2011	e. Mo_5
f. El_4years	f. Age_k24.0term_60	f. Size_gte20
	g. Age_k30.0term_60	g. Lp_home_improvement
	h. Size_lte5	h. Mo_2
	i. Lp_major_purchase	i. Lp_other
	j. Term_60	j. Lp_debt_consolidation
		k. Mo_3
		l. Mo_4
		m. El_10+years
		n. Age_k18.0term_60
		o. Fico_700_750
		p. Lp_credit_card
		q. Mo_10
		r. Mo_12
		s. Mo_11
		t. Size_15_20
		u. El_7years
		v. El_9years
		w. Mo_7
		x. El_6years
		y. Size_5_10
		z. El_8years
		aa. Dti_ex_mortgage

8. <u>D6 -> D6+ Predictor Loading Strength Order</u>:

Strong Positive	**Strong Inverse**	**Weak/Neutral**
a. Age_k30.0term_60	a. Age_k30.0	a. Cohort_2012
b. Age_k24.0term_60	b. Total_accounts	b. Coupon_20+
c. Age_k18.0term_60	c. Age_k0.0	c. Mo_11
d. Total_open_credit_lines	d. Cohort_2008	d. El_9years
	e. Age_k6.0	e. Cohort_2015
		f. Size_gte20
		g. Mo_7
		h. Fico_650_700
		i. El_4years
		j. Cohort_2013
		k. Mo_8
		l. Lp_credit_card
		m. Age_k36.0term_60
		n. Mo_12
		o. Mo_2
		p. Coupon_<10
		q. Monthly_gross_income
		r. El_<1year
		s. Mo_6
		t. Fico_750_800
		u. Mo_9
		v. Term_60
		w. El_10+years
		x. Cohort_2009
		y. El_n/a
		z. Cohort_2010
		aa. Mo_4

		bb. Age_k24.0 cc. Size_5_10 dd. Revolving_utilization ee. Size_lte5 ff. Mo_3

9. D6+ -> D6+ Predictor Loading Strength Order:

Strong Positive	**Strong Inverse**	**Weak/Neutral**
a. Age_k12.0 b. Cohort_2008 c. Cohort_2009	a. Age_k0.0 b. Cohort_2013 c. Cohort_2014 d. Cohort_2012 e. Cohort_2011	a. Mo_11 b. Mo_9 c. Revolving_utilization d. Coupon_20+ e. Fico_650_700 f. Size_gte20 g. Mo_7 h. Coupon_15-20 i. Age_k18.0 j. Mo_12 k. Lp_debt_consolidation l. El_3years m. El_6years n. Mo_6 o. Size_5_10 p. Lp_other q. Mo_4 r. Mo_10 s. Mo_3 t. Size_lte5

		u. Age_k0.0term_60

10. D6+ -> D Predictor Loading Strength Order:

Strong Positive	Strong Inverse	Weak/Neutral
a. Cohort_2012	a. Age_k6.0	a. Lp_car
b. Cohort_2013	b. Age_k12.0	b. Size_15_20
c. Cohort_2014	c. Dq_accounts_past_2_years	c. Dti_ex_mortgage
d. Mo_7	d. Cohort_2008	d. Mo_10
e. Term_60	e. Cohort_2009	e. El_6years
f. Age_k24.0term_60	f. Total_accounts	f. El_9years
g. Mo_12	g. Mo_2	g. El_n/a
h. Mo_6		h. Mo_5
i. Cohort_2011		i. El_5years
		j. El_7years
		k. El_3years
		l. Coupon_15_20
		m. Lp_debt_consolidation
		n. Mo_9
		o. El_10+years
		p. Mo_8
		q. Mo_3
		r. Size_lte5
		s. Coupon_<10
		t. Cohort_2010

Empirical Analysis of Seasoning Effects

1. Default and Prepay over Time: Perhaps one of the most important predictors of a loan's prepayment and default rate is just the passage of time. The empirical exhibits of the conditional prepayment C-P probabilities over age (months on balance) by PeerIQ (2015) illustrate that monthly prepayment probabilities generally increase with loan seasoning.
2. Time Impact on the Delinquency: Conversely, monthly rolls to delinquency (and ultimately to default) generally fall as a loan seasons. Conditional on the fact that a loan has not defaulted to a certain point in time, it is more generally likely to prepay and not default in the future.
3. Actual vs. Fitted C-P Probabilities: PeerIQ (2015) find that the fitted C-P probabilities correspond quiet well to the actual C-P probabilities indicating that the age splines are quiet explanatory (however less so at later ages where there is some noise due to the thinning of the sample size). Nonetheless the general uptrend of prepayment by loan age is evident in both the actual and the fitted C-P probabilities.
4. Hump in the Delinquency Curve: Conversely, rolls into delinquency peak early in the life of a loan (within the first 15 months of age), followed by a gradual decline through the life of the loan. The result is particularly true for the 60 month loans which (due to their longer term) season meaningfully towards the back end.

Analysis of the Vintage/Cohort Effects

1. Vintage/Cohort Impact on Loan: Vintage and cohort effects are designed to capture the difference in prepayment and default experience of loans identical in all respects except for the fact that they were originated at different times.
2. Impact on Prepay/Delinquency Rolls: While the monotonicity of the prepayment rates are likely due to the seasoning effects described above (for e.g., the relative lower prepayment probability of the younger 2015 vintage loans), the peak in fitted delinquency rolls around the 2012 vintage for the 60 month loans and the2013 vintage for the loan 36 month loans is both interesting and consistent with the prepay/default exist in LC data seen before.

Analysis of Empirical Seasonality Effects

1. Usage of Origination DTI Metric: Once again the DTI metrics used in the model are at origination. Again, the results are quite intuitive and fit the data well.
2. DTI Impact on Prepay/Delinquency: Borrowers with higher DTI tend to prepay slower and tend to roll to delinquency at a higher rate, albeit not with the same effect as that observed on prepayment.

CPR And CDR Curve Estimation

1. Definition of CPR and CDR: The fit probabilities (and therefore the sparse transition matrices) obtained from the above are used to derive CPR and CDR curves. CPR and CDR curves are defined as

$$CDR(t) = 1 - \left[1 - \frac{Default(t)}{BeginBalance(t)}\right]^{12}$$

and

$$CPR(t) = 1 - \left[1 - \frac{Prepay(t)}{BeginBalance(t)}\right]^{12}$$

2. CPR/CDR for Loan Pools: CPR and CDR therefore represent an annualized measure of prepay and default for every dollar of principal outstanding at the start of a period. Clearly, in order to estimate the CPR and CDR, we need to project the cash flows for a given pool of loans.

3. Projecting Each Loan in Pool: A first step to projecting the cash flows is to project the status of each loan in the pool at all future points in time. This is achieved by using the transition matrix

$$P_{i,j}(t) = \begin{bmatrix} P_{c,c}(t) & \cdots & P_{c,d}(t) \\ \vdots & \ddots & \vdots \\ P_{6+,c}(t) & \cdots & P_{6+,d}(t) \end{bmatrix}$$

4. Using MNL Loan Transition Probabilities: Probabilities are estimated at all points in time, as per the earlier section. Suppose, for example, that at a given time t the loan status is current. Aw: The multinomial distribution for where the loan can transition to at time $t + 1$ is used, and it is simply by the first row of the transition matrix.
5. Loan Target Status Random Draw: Therefore a random draw is made from this distribution to project the status of the loan at time $t + 1$. If, in that draw, the loan ended up in a status of 30-day delinquent, a simulation is done using the second row of $P_{i,j}(t + 1)$ and so on, until the earliest of one of maturity, default, or prepay.
6. Projection of Loan Cash Flows: After having obtained the future status of all the loans in the pool, we can compute the cash flows appropriate to the product under consideration (in this case the 36 month vs. the 60 month fixed rate loans).
7. Loan State Dependent Cash Flow: For example, in the case of Lending Club, one can continue to apply the monthly fixed payment on the loan if the loan is in current status, and accrue interest if the loan becomes delinquent, or discontinue further payments altogether if the loan voluntarily prepays or defaults.
8. CPR/CDR from Cash Flows: Such logic allows the computation of the loan balances from period to period, and ultimately the CPR and the CDR above.
9. CPR and CDR Simulation: As a final example, we project the loan status, derive cash flows, and compute CPR and CDR for the entire population of outstanding Lending Club loans – outstanding as of January 2013 – and examine the results. DROP generates the results of one simulation on the portfolio containing 14,000 Lending Club loans as at 1 January 2013.
10. Actual vs. Simulated CPR/CDR: As evidenced from the above, the projection for the CPR and the CDR agrees with the realized values for much of the projection period. For CDR,

there is some volatility towards the latter end of the projection period where the sample thins out, and where the number of defaults (as a rare transition) can have a meaningful impact.

11. Multi-Path CPR/CDR Simulation: In addition to such ‘single path’ projections of CPR and CDR on a portfolio, because the calibration produces a *distribution* of transitions, one can generate multiple paths for the CPR and CDR on a loan portfolio. Using the model DROP simulated 10 paths for the LC portfolio resulting in a distribution of CPR and CDR for the portfolio.

Credit Model Future Enhancements

1. Inclusion of Macro-economic Regressors: For the purposes of keeping the analysis simple, DROP’s model excludes macro-economic regressors from the analysis. However economic variables can be integrated quiet easily into the structure of the model.
2. Stochastic Simulation of Market Variables: In future versions we plan to add in carefully selected market variables such as interest rates or unemployment, which can be used as scenario variables, or stochastically simulated via parameters calibrated from the prices of traded instruments (e.g., interest rate options).

References

- PeerIQ (2015): PeerIQ Analytics Credit Model Methodology, Version 2.0.
- Receiver Operating Characteristic (Wiki).