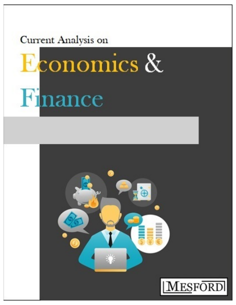

The Probability of Default Distribution of Heterogeneous Loan Portfolio

Biao Wu*

Credit Modeling and Methodology, Royal Bank of Canada, Toronto, ON, Canada.

Abstract:

Vasicek's homogeneous loan portfolio theory has been influential in both research and industry practice of credit risk management. However, big loan portfolio is usually heterogeneous in nature; meanwhile it can consist of a number of homogeneous sub-portfolios. Herein, a theoretical framework extending that of Vasicek's is proposed for heterogeneous loan portfolio. The probability of default distribution and its limit is studied. Similar to Vasicek's homogeneous loan portfolio percentage loss distribution, this limiting PD distribution has the desirable properties including symmetry, mean, variance and percentile properties which can promote its application in practice. As a key analytical component it can also help the estimation of the loss distribution of heterogeneous loan portfolio.

Publication History: Received: 04 March 2019 | Revised: 20 March 2019 | Accepted: 05 April 2019

Keywords:
Probability of Default, Asset Correlation, Homogeneous Portfolio, Heterogeneous Portfolio, Probability of Default Distribution, Loan Loss Distribution.

1. INTRODUCTION

Vasicek's framework of homogeneous loan portfolio is described in [7], and its limiting distribution of portfolio percentage loss as the portfolio size becomes very large is discussed in [8]. The desirable properties of this limiting distribution is well studied in [9].

Since their publications, Vasicek's famous original work of homogeneous loan portfolio has been influential in both research and industry practice of credit risk management. Thousands of research papers and other literatures are directly or indirectly related to Vasicek's papers. These results had been applied to KMV's *Portfolio Manager* or Moody's *Risk Frontier*. The Basel Committee on Banking Supervision (BCBS) actually employed Vasicek's homogeneous loan portfolio conditional distribution when it designed the fundamental Basel II regulatory capital formula, see [1].

However, homogeneous loan portfolio is mostly restricted to either small portfolio or portfolio with only few business lines, sectors and regions. But big bank loan portfolio usually consists of a number of business lines and sectors, and spreads many regions/countries and therefore is heterogeneous in nature. Hence, the heterogeneous portfolio also becomes a focus of research in academy. Vasicek (2002, [9]) considered a type of heterogeneous portfolios which can be approximated by homogeneous portfolios. Gordy (2003, [3]) studied the exposure weighted loss distribution of heterogeneous portfolio. Pykhtin (2004, [5]) proposed a multi-factor type model where all borrowers can be heterogeneous each other. Rosen and Saunders (2009, [6]) extended Pykhtin's model to static hedging of credit risk. Among others, Yang (2017, [10]) followed Vasicek's framework and set up a structure of heterogeneous portfolio to model long run PD over segments. Herein, motivated by these pioneering works, we also propose an extended theoretical framework of heterogeneous loan portfolio which can cover a number of homogeneous sub-portfolios with different expected default frequency. In particular, we study the limiting probability of default distribution under this framework and its properties.

Since BCBS published the guide to evaluate expected credit loss (ECL) following the IFRS 9 accounting standard (see [2]), IFRS 9 banks have been motivated to calculate ECL in both scientific and compliant way. To rigorously evaluate ECL, one of many challenges is to simultaneously model and predict PD term structure for portfolios with a few risk ratings / segments etc. These kinds of portfolios are certainly heterogeneous. To formulate the IFRS 9 PD term structure model, Yang proposed a unified formula of binomial type likelihood across all risk ratings which leads to maximum likelihood PD term structure estimate, see [10]. Similarly, the limiting probability of

*Address correspondence to this author at Credit Modeling and Methodology, Royal Bank of Canada, Toronto, ON, Canada; Tel: 1(416)-313-7239; E-mail: biao.wu@rbc.com

Mesford Publisher Inc
Office Address: Suite 2205, 350 Webb Drive, Mississauga, ON L5B3W4, Canada; T: +1 (647) 7109849 | E: caef@mesford.ca, contact@mesford.ca, https://mesford.ca/journals/caef/

default (PD) density function for heterogeneous loan portfolio can provide another choice of unified likelihood formula across all categories. This neat structure can promote formulating the likelihood of PD for heterogeneous portfolios.

This paper is organized as follows. In §2.1, we introduce Vasicek's homogeneous loan portfolio framework and its percentage loss limiting distribution and related properties. The heterogeneous portfolio research progresses are reviewed in §2.2. In §3, we describe our framework of heterogeneous loan portfolio. Its limiting PD distribution is presented in §4 and relevant properties are discussed in §5. The proofs or derivations for these properties are given in Appendix.

2. HOMOGENEOUS AND HETEROGENEOUS LOAN PORTFOLIOS IN LITERATURES

2.1. Vasicek's System of Homogeneous Loan Portfolio

Before describing heterogeneous loan portfolio in details, we briefly summarize Vasicek's homogeneous loan portfolio. A homogeneous loan portfolio consists of n loans of equal or approximately equal dollar amount, see [7] and [9] . A loan defaults if its borrower's total asset value at the loan maturity falls below its contractual obligation. The asset value processes of these borrowers follow geometric Brownian motion, where the pairwise correlations of the underlying Brownian motions or Wiener processes are assumed to be same.

Specifically, let p be the common probability of default and $c = N^{-1}(p)$ the common default point of the homogeneous loans in the portfolio. Assume that the asset values of the borrowing companies are correlated with a coefficient ρ for any two companies, i.e.

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

where Y, Z_1, ..., Z_n are mutually independent standard normal variables and ρ is named asset correlation. $\sqrt{\rho}Y$ is often called systematic risk, and $\sqrt{1-\rho}Z_i$ idiosyncratic risk of the i-th loan.

Denote by L_i the default flag of the i-th loan, i.e., $L_i = 1$ if the i-th borrower defaults and $L_i = 0$ otherwise. L_i is the gross loss of the i-th loan. Let L be the portfolio percentage gross loss,

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

Conditional on systematic risk Y, the conditional probability of loss of single loan is

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

The probability density and cumulative distribution functions of portfolio percentage loss L as $n \to \infty$ are denoted by $f(x; p, \rho)$ and $F(x; p, \rho)$ respectively, due to their relation with respect to p and ρ:

$$f(x;p,\rho) = \sqrt{\frac{1-\rho}{\rho}}\exp\left(-\frac{1}{2\rho}(\sqrt{1-\rho}N^{-1}(x) - N^{-1}(p))^2 + \frac{1}{2}(N^{-1}(x))^2\right), \qquad (2.1)$$

$$F(x;p,\rho) = N\left(\frac{\sqrt{1-\rho}N^{-1}(x) - N^{-1}(p)}{\sqrt{\rho}}\right) \qquad (2.2)$$

For ease of comparison between homogeneous and heterogeneous loan portfolio limiting distributions, the desirable properties of L and its pdf $f(x; p, \rho)$ and cdf $F(x; p, \rho)$ are summarized in the following:

1) $F(x; p, \rho)$ possesses a **symmetry** property

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

2) $f(x; p, \rho)$ is unimodal. The **mode** of $f(x; p, \rho)$ is

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

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

3) The **mean** of L is

$$E[L] = p$$

4) The **variance** of L is

$$Var(L) = N_2(N^{-1}(p), N^{-1}(p), \rho) - p^2,$$

where N_2 is the bivariate cumulative normal distribution function.

5) The inverse of $F(x; p, \rho)$, i.e., the α-**percentile** value of L, is given by

$$L_\alpha = F(\alpha; 1-p, 1-\rho) = N\left(\frac{\sqrt{\rho}N^{-1}(\alpha) + N^{-1}(p)}{\sqrt{1-\rho}}\right)$$

Note that the portfolio percentage loss distribution given by (2.1) or (2.2) is actually PD distribution. Both percentage loss and PD mean the same things for the homogeneous loan portfolio when loan amounts are assumed to be same.

2.2. Research Progresses on Heterogeneous Loan Portfolios

Before we formally define the framework of our heterogeneous loan portfolio, we first review the relevant literatures. Since big loan portfolio is heterogeneous in nature, the heterogeneous portfolios have been the focus of research and modeling by many researchers and industry practitioners.

Vasicek (2002) considered a type of heterogeneous portfolio which can be approximated by one homogeneous portfolio, see [9]. Gordy (2003, [3]) studied the exposure weighted loss distribution of heterogeneous portfolio, and proved convergence of a few loss-related statistics.

Pykhtin (2004) proposed a multi-factor type model, see [5]. He assumed that asset returns depend linearly on a few normally distributed systematic risk factors representing industry, geography, global economy or any other relevant indexes.

Borrower i's standardized asset return is driven by both systematic factors Y_i and borrower-specific risk:

$$X_i = r_i Y_i + \sqrt{1-r_i^2}\xi_i$$

where $Y_i = \sum_{k=1}^{N} \alpha_{ik} Z_k$ satisfying $\sum_{k=1}^{N} \alpha_{ik}^2 = 1$, r_i is the factor loading measuring i-th borrower's sensitivity to the systematic risk, and ξ_i is the standardized normally distributed idiosyncratic shock. Under Pykhtin's model, all borrowers can be heterogeneous from each other.

Rosen and Saunders (2009, [6]) studied multi-factor model and applied it to static hedging of credit risk after they solved the parameters α_{ik} in Pykhtin's formulation by least-square method.

Yang (2017, [10]) set up a structure mostly relevant to the framework described in §3, see [10]. Yang differentiated the asset correlations of different segments. Conditional on the systematic risk factor y_t, the i-th segmental PD at time t is given by

$$P_i(y_t) = N(c_i\sqrt{1+r_i^2} + r_i y_t)$$

where $r_i = \sqrt{\frac{\rho_i}{1-\rho_i}}$. To model the long run PD over segments, binomial distribution is applied to each segment, with $P_i(y_t)$ the probability of one occurrence of default, and the portfolio level likelihood is just the product of segment level binomial probability mass functions, based on the conditional independence between segments. This is an application of the non-limiting heterogeneous loan portfolio structure. This is a successful structure for stress testing purposes, and has been influential in the industry practice among the Canadian community.

3. OUR FRAMEWORK FOR HETEROGENEOUS LOAN PORTFOLIO

In this section, we delineate the theoretical framework of heterogeneous loan portfolio. Note that loan amounts of heterogeneous loan portfolios are usually not same. The portfolio PD distribution and percentage loss distribution is different subject under this framework. In this paper, we focus on the PD distribution only.

Assume a loan portfolio consists of K categories. The categories could be risk ratings, business lines, industry sectors, geographical regions etc. We denote these K categories as R_i with size n_i, $(1 \leq i \leq K)$. Therefore, the total number of loans is $n = n_1 + \dots + n_K$. We assume each sub-portfolio of a given category consists of homogeneous loans, but the loans from different categories are heterogeneous. Particularly, the loan amounts of different categories are assumed to be different. Herein, the defined portfolio is a heterogeneous loan portfolio.

For the i-th loan category or sub-portfolio, there are n_i loans of equal or approximately equal dollar amount. Use subscripts i and j for the j-th loan of the i-th category. The ij-th loan defaults if the ij-th borrower's total asset value at the loan maturity T falls below its contractual obligation B_{ij}. Denote by A_{ijt} the value of the ij-th borrower's assets at time t, where

$$dA_{ijt} = \mu_{ij} A_{ijt} dt + \sigma_{ij} A_{ijt} dx_{ijt},$$

and $\{x_{ijt}\}_{1\leq j\leq n_i, 1\leq i\leq K}$ are Wiener processes with

$$E\,[dx_{ijt}]^2 = dt, \qquad 1 \leq j \leq n_i, 1 \leq i \leq K,$$

$$E[dx_{ijt} dx_{i'j't}] = \sqrt{\rho_i \rho_{i'}}\,dt, 1 \leq j, j' \leq n_i, 1 \leq i \neq i' \leq K,$$

$$E[dx_{ijt} dx_{ij't}] = \rho_i dt, \quad 1 \leq j \neq j' \leq n_i, 1 \leq i \leq K.$$

Here ρ_i captures the specific features of the i-th category.

The asset value A_{ijT} at T can be represented as

$$\log A_{ijT} = \log A_{ij0} + \mu_{ij} T - \frac{1}{2}\sigma_{ij}^2 T + \sigma_{ij}\sqrt{T} X_{ij}$$

where X_{ij} is a standard normal variable. Denote by p_{ij} the ij-th loan's probability of default, i.e.,

$$p_{ij} = P[X_{ij} < c_{ij}] = N(c_{ij})$$

Where

$$c_{ij} = \frac{\log B_{ij} - \log A_{ij0} - \mu_{ij} T + \frac{1}{2}\sigma_{ij}^2 T}{\sigma_{ij}\sqrt{T}}$$

Let p_i be the common PD and $c_i = N^{-1}(p_i)$ the common default point of the loans in the i-th category. Assume further that the asset values of the borrowing companies from the i-th category can be decomposed by the correlation coefficient or asset correlation ρ_i as follows:

$$x_{ijt} = \sqrt{\rho_i} y_t + \sqrt{1-\rho_i} z_{ijt}$$

and

$$E[(dy_t)^2] = dt$$

$$E[(dz_{ijt})^2] = dt$$

$$E[(dy_t)(dz_{ijt})] = 0$$

$$E[(dz_{ijt})(dz_{i'j't})] = 0, (i,j) \neq (i',j')$$

where $\sqrt{\rho_i} y_t$ is the captured systematic risk, and $\sqrt{1-\rho_i} z_{ijt}$ the idiosyncratic risk of the ij-th loan.

4. THE LIMITING PD DISTRIBUTION OF HETEROGENEOUS LOAN PORTFOLIO

In this section, we derive the limiting PD distribution of the heterogeneous loan portfolio.

Denote by D_{ij} the default flag of the ij-th loan, i.e., $D_{ij} = 1$ if the ij-th borrower defaults and $D_{ij} = 0$ otherwise. Let *PD* be the portfolio probability of default,

$$PD = \frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i} D_{ij}$$

We wish to calculate the probability distribution of default rate or *PD* on the portfolio, that is

$$P_k = P[PD = \frac{k}{n}], k = 0, 1, \cdots, n$$

For a given k, the k defaults must arise from the K categories. Any combination of the k defaults must satisfy

$$k_1 + \cdots + k_K = k, 0 \le k_i \le n_i$$

When there are k_i defaults in the i-th category, the rest $n_i — k_i$ loans are performing. Therefore,

$$\begin{aligned}
&P[PD = \frac{k}{n}] \\
&= \sum_{\substack{k=k_1+\cdots+k_K,\\ 0\le k_i\le n_i, 1\le i\le K}} \Big[\prod_{i=1}^{K}\binom{n_i}{k_i}\Big] P\Big[\bigcap_{i=1}^{K}\Big\{\frac{z_{i1T}}{\sqrt{T}} < \Big(\frac{c_i - \frac{b_i y_T}{\sqrt{T}}}{a_i}\Big), \cdots, \frac{z_{ik_iT}}{\sqrt{T}} < \Big(\frac{c_i - \frac{b_i y_T}{\sqrt{T}}}{a_i}\Big), \\
&\qquad \frac{z_{i(k_i+1)T}}{\sqrt{T}} \ge \Big(\frac{c_i - \frac{b_i y_T}{\sqrt{T}}}{a_i}\Big), \cdots, \frac{z_{in_iT}}{\sqrt{T}} \ge \Big(\frac{c_i - \frac{b_i y_T}{\sqrt{T}}}{a_i}\Big)\Big\}\Big] \\
&= \sum_{\substack{k=k_1+\cdots+k_K,\\ 0\le k_i\le n_i, 1\le i\le K}} \Big[\prod_{i=1}^{K}\binom{n_i}{k_i}\Big] \int_{-\infty}^{\infty} P\Big[\bigcap_{i=1}^{K}\Big\{\frac{z_{i1T}}{\sqrt{T}} < \Big(\frac{c_i - b_i u}{a_i}\Big), \cdots, \frac{z_{ik_iT}}{\sqrt{T}} < \Big(\frac{c_i - b_i u}{a_i}\Big), \\
&\qquad \frac{z_{i(k_i+1)T}}{\sqrt{T}} \ge \Big(\frac{c_i - b_i u}{a_i}\Big), \cdots, \frac{z_{in_iT}}{\sqrt{T}} \ge \Big(\frac{c_i - b_i u}{a_i}\Big)\Big\}\Big| \frac{y_T}{\sqrt{T}} = u\Big] dN(u) \\
&= \sum_{\substack{k=k_1+\cdots+k_K,\\ 0\le k_i\le n_i, 1\le i\le K}} \Big[\prod_{i=1}^{K}\binom{n_i}{k_i}\Big] \int_{-\infty}^{\infty} \prod_{i=1}^{K}\Big(N(\frac{c_i - b_i u}{a_i})\Big)^{k_i}\Big(1 - N(\frac{c_i - b_i u}{a_i})\Big)^{n_i-k_i} dN(u) \\
&= \int_{-\infty}^{\infty} \sum_{\substack{k=k_1+\cdots+k_K,\\ 0\le k_i\le n_i, 1\le i\le K}} \prod_{i=1}^{K}\binom{n_i}{k_i}\Big(N(\frac{c_i - b_i u}{a_i})\Big)^{k_i}\Big(1 - N(\frac{c_i - b_i u}{a_i})\Big)^{n_i-k_i} dN(u), \qquad (4.1)
\end{aligned}$$

where $\alpha_i = \sqrt{1-\rho_i}$, $b_i = \sqrt{\rho_i}$, $p_i = N(c_i)$ or $c_i = N^{-1}(p_i)$.

To simplify calculations later on, we introduce simpler notations. Let $s_i(u) = N(\frac{c_i - b_i u}{a_i})$. Then $0 < s_i(u) < 1, \forall u \in (-\infty, \infty)$. By (4.1), we have

$$P_k = \int_{-\infty}^{\infty} \sum_{\substack{k=k_1+\cdots+k_K,\\ 0\le k \le n_i, 1\le i\le K}} \prod_{i=1}^{K}\binom{n_i}{k_i}(s_i(u))^{k_i}(1-s_i(u))^{n_i-k_i} dN(u) \qquad (4.2)$$

The cumulative probability that PD does not exceed $x \in (0,1)$ is

$$\begin{aligned}
F_n(x) &= \sum_{k=0}^{[nx]} P_k \\
&= \int_{-\infty}^{\infty} \sum_{k=0}^{[nx]} \sum_{\substack{k=k_1+\cdots+k_K,\\ 0\le k_i\le n_i, 1\le i\le K}} \prod_{i=1}^{K}\binom{n_i}{k_i}(s_i(u))^{k_i} \\
&\qquad (1 - s_i(u))^{n_i-k_i} dN(u) \qquad (4.3)
\end{aligned}$$

Proposition 1. Assume a heterogeneous loan portfolio consists of K homogeneous loan categories with category size n_i $(1 \le i \le K)$. Let $n = n_1 + \cdots + n_K$. Assume that p_i is the probability of default of any borrower, and ρ_i the correlation coefficient between any pair of borrowers in the i-th category. Assume also that $\lim_{n\to\infty}\frac{n_i}{n} = w_i > 0$ for each $1 \le i \le K$. Then the limiting cumulative distribution function $F(x; p_1, \ldots, p_K; \rho_1, \ldots, \rho_K)$ and probability density function $f(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)$ of *PD* are given by

$$F(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K) = N\Big(g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)\Big), \qquad (4.4)$$

$$f(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K) = \frac{n\Big(g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)\Big)}{\sum_{i=1}^{K} w_i\sqrt{\frac{\rho_i}{1-\rho_i}}\, n\Big(\frac{N^{-1}(p_i) + \sqrt{\rho_i}\, g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)}{\sqrt{1-\rho_i}}\Big)},$$

where (4.5)

$$g(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K) \equiv \sum_{i=1}^{K} w_i N\Big(\frac{N^{-1}(p_i) + \sqrt{\rho_i}\, x}{\sqrt{1-\rho_i}}\Big). \qquad (4.6)$$

Proof. Conditional on $u \in (-\infty, \infty)$, we construct Bernoulli sequences $\{X_{i1}, X_{i2}, \cdots, X_{in_i}\}_{i=1}^{K}$ on some probability space $(\Omega, \mathcal{F}, P)$ in a way such that X_{ij} $(1 \le j \le n_i)$ are identically distributed with

$$P(X_{ij} = 1|u) = s_i(u) = N(\frac{c_i - b_i u}{a_i}),$$

$$P(X_{ij} = 0|u) = 1 - s_i(u) = 1 - N(\frac{c_i - b_i u}{a_i}),$$

and $\{X_{i1}, X_{i2}, \cdots, X_{in_i}\}_{i=1}^{K}$ are conditionally independent. Next, let

$$S_{n_i}^{(i)} = \sum_{j=1}^{n_i} X_{ij} \text{ and } S_n = \sum_{i=1}^{K} S_{n_i}^{(i)}.$$

It follows that

$$P(S_n \le [nx]|u) = \sum_{k=0}^{[nx]} P(S_n = k|u)$$

$$=\sum_{k=0}^{[nx]}\sum_{\substack{k=k_1+\cdots+k_K\\0\le k_i\le n_i,1\le i\le K}}\left[\prod_{i=1}^{K}\binom{n_i}{k_i}\right]P(S_{n_1}^{(1)}=k_1,\cdots,S_{n_K}^{(K)}=k_K|u)$$

$$=\sum_{k=0}^{[nx]}\sum_{\substack{k=k_1+\cdots+k_K\\0\le k_i\le n_i,1\le i\le K}}\left[\prod_{i=1}^{K}\binom{n_i}{k_i}\right]\prod_{i=1}^{K}P(S_{n_i}^{(i)}=k_i|u)$$

$$=\sum_{k=0}^{[nx]}\sum_{\substack{k=k_1+\cdots+k_K\\0\le k_i\le n_i,1\le i\le K}}\prod_{i=1}^{K}\binom{n_i}{k_i}(s_i(u))^{k_i}(1-s_i(u))^{n_i-k_i} \quad (4.7)$$

By the law of large numbers, we have $\lim_{n\to\infty}\frac{S_{n_i}^{(i)}}{n_i}=s_i(u)$, and by the assumption $\lim_{n\to\infty}\frac{n_i}{n}=w_i$, we get

$$\lim_{n\to\infty}\frac{S_n}{n}=\lim_{n\to\infty}\sum_{i=1}^{K}\frac{n_i}{n}\frac{S_{n_i}^{(i)}}{n_i}\overset{a.s}{=}\sum_{i=1}^{K}w_is_i(u)$$

$$=\sum_{i=1}^{K}w_iN\left(\frac{c_i-b_iu}{a_i}\right)$$

Then, we get

$$\lim_{n\to\infty}P(S_n\le[nx]|u)=$$

$$\begin{cases}1, & \text{if } \sum_{i=1}^{K}w_iN\left(\frac{c_i-b_iu}{a_i}\right)\le x\\ 0, & \text{otherwise.}\end{cases} \quad (4.8)$$

Therefore, combining (4.3), (4.7) and (4.8), we obtain

$$\lim_{n\to\infty}F_n(x) = \lim_{n\to\infty}\int_{-\infty}^{\infty}P(S_n\le[nx]|u)dN(u)$$

$$= \int_{\sum_{i=1}^{K}w_iN\left(\frac{c_i-b_iu}{a_i}\right)\in[0,x]}dN(u). \quad (4.9)$$

By the relations of a_i, b_i and c_i with respect to ρ_i and p_i, define function g in the following:

$$g(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)=\sum_{i=1}^{K}w_iN\left(\frac{c_i+b_ix}{a_i}\right)$$

g is a strictly increasing function of x on its domain $(-\infty,\infty)$, for any given $p_i\in(0,1)$, $\rho_i\in(-1,1)$, $1\le i\le K$. Therefore, g-1 exists, although its closed form is complex, except the case $K=1$ (see Remark 1).

It follows that

$$F(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K) = \lim_{n\to\infty}F_n(x)$$

$$= \int_{-g^{-1}(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)}^{\infty}dN(u)$$

$$= 1-N(-g^{-1}(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K))$$

$$= N\left(g^{-1}(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)\right),$$

where g is given by (4.6).

The probability density function $f(x;p_1,\ldots,p_K;\rho_1,\ldots,\rho_K)$ in (4.5) can be derived by taking derivative with respect to x on $F(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)$.

Remark 1. If $c_i=c(p=p_i)$ and $\rho_i=\rho$, $\forall 1\le i\le K$, then the K categories collapse to one. The functions g and g^{-1} are

$$g(x;p,\rho)=N\left(\frac{N^{-1}(p)+\sqrt{\rho}x}{\sqrt{1-\rho}}\right),$$

$$g^{-1}(x;p,\rho)=\frac{\sqrt{1-\rho}N^{-1}(x)-N^{-1}(p)}{\sqrt{\rho}}$$

Therefore,

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

which is exactly Vasicek's homogeneous loan portfolio percentage loss distribution function (2.2).

5. PROPERTIES OF LIMITING PD DISTRIBUTION OF HETEROGENEOUS LOAN PORTFOLIO

In this section, we discuss properties of the limiting distribution (4.4). The heterogeneous loan portfolio PD with pdf $f(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)$ and cdf $F(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)$ have the following properties:

1) $F(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)$ possesses a **symmetry** property

$$F(x;p_1,\cdots,p_K;\rho_1,\cdots,\rho_K)=1-F(1-x;1-p_1,\cdots,1-p_K;\rho_1,\cdots,\rho_K) \quad (5.1)$$

2) There is no simple formula for **mode** of $f(x;p_1,\ldots,p_K;\rho_1,\ldots,\rho_K)$ due to the complexity of (4.5), except K= 1.

3) The **mean** of PD is

$$E[PD]=\sum_{i=1}^{K}w_ip_i \quad (5.2)$$

4) The **variance** of PD is

$$Var[PD] = \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j [N_2(N^{-1}(p_i), N^{-1}(p_j) ,\sqrt{\rho_i\rho_j}) - p_i p_j] \quad (5.3)$$

where N_2 is the bivariate cumulative normal distribution function.

5) The inverse of $F\ (x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)$, i.e., the α-**percentile** value of *PD,* is given by

$$PD_\alpha = \sum_{i=1}^{K} w_i L_\alpha^{(i)}, \quad (5.4)$$

where $L_\alpha^{(i)}$ is the α-percentile of the *i*-th sub-portfolio, i.e.,

$$L_\alpha^{(i)} = F(\alpha; 1-p_i, 1-\rho_i) = N\left(\frac{\sqrt{\rho_i}N^{-1}(\alpha) + N^{-1}(p_i)}{\sqrt{1-\rho_i}}\right)$$

Remark 2. We have seen some features of mixture-type property in (4.6), (5.2), and (5.4). These features originate from the assumption that the heterogeneous loan portfolio consists of homogeneous sub-portfolios. However, the limiting distribution of *PD* is neither of real compounding type nor of strict mixture type. The main reason is that the inverse of *g* appears in the cumulative distribution function *N* (.) of standard normal random variable, see (4.4) for details.

However, the properties listed above can promote applications using the heterogeneous loan portfolio *PD* distribution, due to their straightforward relation with respect to homogeneous loan portfolios which are well studied and utilized in industry.

CONCLUSION

The limiting PD distribution can be applied to estimate the percentiles of heterogeneous portfolio PD, which are useful for PD quantification. It can be used to estimate long run PD for stress testing purposes, using its limiting probability density function. More importantly, if we define portfolio loss ratio as the exposure weighted loss, same way as Gordy (see [3]), the limiting PD distribution can help the estimation of loss distribution of heterogeneous portfolio as a key analytical component, which is of high values.

ACKNOWLEDGEMENT

The author would like to thank both referees for the suggestion of enhancing the introduction and reference section, and in particular for the advice of extending the overall asset correlation to segment-wise asset correlations. Special thanks to my colleague Bill Yang for his constructive omments.

A. APPENDIX

A.1. Proof of the symmetry (5.1)

Proof. Now we prove the symmetry property of the cumulative distribution function of the heterogeneous loan portfolio *PD*, i.e.

$$F(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K) = 1 - F(1-x; 1-p_1, \cdots, 1-p_K; \rho_1, \cdots, \rho_K).$$

It follows from (4.4) that

$$F(1-x; 1-p_1, \cdots, 1-p_K; \rho_1, \cdots, \rho_K) = N\left(g^{-1}(1-x, 1-p_1, \cdots, 1-p_K; \rho_1, \cdots, \rho_K)\right)$$

By (4.6),

$$1 - g(x; 1-p_1, \cdots, 1-p_K; \rho_1, \cdots, \rho_K) = 1 - \sum_{i=1}^{K} w_i N\left(\frac{N^{-1}(1-p_i) + \sqrt{\rho_i}x}{\sqrt{1-\rho_i}}\right)$$

$$= \sum_{i=1}^{K} w_i \left[1 - N\left(\frac{-N^{-1}(p_i) + \sqrt{\rho_i}x}{\sqrt{1-\rho_i}}\right)\right]$$

$$= \sum_{i=1}^{K} w_i N\left(\frac{N^{-1}(p_i) - \sqrt{\rho_i}x}{\sqrt{1-\rho_i}}\right)$$

$$= g(-x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K). \quad (A.1)$$

It follows from (A.1) that

$$g^{-1}(1-x; 1-p_1, \cdots, 1-p_K; \rho_1, \cdots, \rho_K) = -g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K).$$

Therefore,

$$F(1-x; 1-p_1, \cdots, 1-p_K; \rho_1, \cdots, \rho_K) = = N\left(-g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)\right)$$

$$= 1 - N\left(g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)\right)$$

$$= 1 = F(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)$$

A.2. Proof of the mean (5.2)

Before giving the proof for the mean property, we introduce the following lemma which is used in the proof:

Lemma 1. Assume that S is a normal random variable with mean μ and standard deviation σ, i.e. $S \sim N(\mu, \sigma^2)$, then the following equality holds:

$$E_S[N(a_0 + a_1 S)] = N\left(\frac{a_0 + a_1\mu}{\sqrt{1 + a_1^2\sigma^2}}\right) \quad (A.2)$$

Proof of the **mean:** We compute the expected value of the limit distribution given by (4.4): let $s = g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots \rho_K)$, then,

$$\begin{aligned} E[PD] &= \int_0^1 x dF(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K) \\ &= \int_0^1 x dN\left(g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)\right) \\ &= \int_{-\infty}^{\infty} g(s; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K) dN(s) \\ &= \sum_{i=1}^{K} w_i \int_{-\infty}^{\infty} N\left(\frac{N^{-1}(p_i) + \sqrt{\rho_i}s}{\sqrt{1-\rho_i}}\right) dN(s) \\ &= \sum_{i=1}^{K} w_i p_i, \end{aligned}$$

where the last equality applies Lemma 1 with $a_0 = \frac{N^{-1}(p_i)}{\sqrt{1-\rho_i}}$, $a_1 = \sqrt{\frac{\rho_i}{1-\rho_i}}$, and $S \sim N(0, 1)$.

A.3. Proof of the variance (5.3)

Before proving the variance formula (5.3), we introduce another lemma about normal random variables, then compute the second moments of the limit distribution (4.4).

Lemma 2. Assume that $X_1 \sim N(\mu_{X1}, \sigma^2_{X1})$ and $X_2 \sim N(\mu_{X_2}, \sigma^2_{X2})$, $S \sim N(\mu_S, \sigma^2_S)$. X_1, X_2 and S are independent. Then,

$$\begin{aligned} & E_S[F_{X_1}(a_0 + a_1 S) F_{X_2}(b_0 + b_1 S)] \\ = & N_2\left(\frac{a_0 - \mu_{X_1} + a_1\mu_S}{\sqrt{\sigma_{X_1}^2 + a_1^2\sigma_S^2}}, \frac{b_0 - \mu_{X_2} + b_1\mu_S}{\sqrt{\sigma_{X_2}^2 + b_1^2\sigma_S^2}}, \frac{a_1 b_1 \sigma_S^2}{\sqrt{(\sigma_{X_1}^2 + a_1^2\sigma_S^2)(\sigma_{X_2}^2 + b_1^2\sigma_S^2)}}\right), \end{aligned} \quad (A.3)$$

where F_{Xi} is the cumulative distribution function of X_i (i = 1,2) and N_2 is bivariate normal cumulative distribution function.

Proof. Since X_1, X_2 and S are independent, then $X_1 - a_1 S \sim N(\mu_{X1} - a_1\mu_S, \sigma^2_{X1} + a_1^2\sigma_S^2)$, and

$X_2 - b_1 S \sim N(\mu_{X_2} - b_1\mu_S, \sigma_{X_2}^2 + b_1^2\sigma_S^2)$. It follows that

$$\begin{aligned} & E_S[F_{X_1}(a_0 + a_1 S) F_{X_2}(b_0 + b_1 S)] \\ = & E_S[E_{X_1}[I_{\{X_1 - a_1 S \le a_0\}}] E_{X_2}[I_{\{X_2 - b_1 S \le b_0\}}]] \\ = & E_S[E_{X_1 X_2}[I_{\{X_1 - a_1 S \le a_0, X_2 - b_1 S \le b_0\}}]] \\ = & P(X_1 - a_1 S \le a_0, X_2 - b_1 S \le b_0) \\ = & N_2\left(\frac{a_0 - \mu_{X_1} + a_1\mu_S}{\sqrt{\sigma_{X_1}^2 + a_1^2\sigma_S^2}}, \frac{b_0 - \mu_{X_2} + b_1\mu_S}{\sqrt{\sigma_{X_2}^2 + b_1^2\sigma_S^2}}, \frac{a_1 b_1 \sigma_S^2}{\sqrt{(\sigma_{X_1}^2 + a_1^2\sigma_S^2)(\sigma_{X_2}^2 + b_1^2\sigma_S^2)}}\right) \end{aligned} \quad (A.4)$$

The last equality holds because the correlation between $X_1 - a_1 S$ and $X_2 - b_1 S$ is $\frac{a_1 b_1 \sigma_S^2}{\sqrt{(\sigma_{X_1}^2 + a_1^2\sigma_S^2)(\sigma_{X_2}^2 + b_1^2\sigma_S^2)}}$.

Proof of the **variance**: Now we compute the second moment of the limit distribution (4.4) and then prove the variance formula (5.3). Let $s = g^{-1}(x, p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)$, then $x = g(s; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)$. It follows that

$$\begin{aligned} E[PD^2] &= \int_0^1 x^2 dF(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K) \\ &= \int_0^1 x^2 dN\left(g^{-1}(x; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)\right) \\ &= \int_{-\infty}^{\infty} g^2\left(s; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K\right) dN(s) \\ &= \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j \int_{-\infty}^{\infty} N\left(\frac{N^{-1}(p_i) + \sqrt{\rho_i}s}{\sqrt{1-\rho_i}}\right) N\left(\frac{N^{-1}(p_j) + \sqrt{\rho_j}s}{\sqrt{1-\rho_j}}\right) dN(s) \\ &= \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j N_2(N^{-1}(p_i), N^{-1}(p_j), \sqrt{\rho_i\rho_j}), \end{aligned} \quad (A.5)$$

The last equality follows from Lemma 2.

It follows from (5.2) and (A.5) that

$$Var[PD] = \sum_{i=1}^{K}\sum_{j=1}^{K} w_i w_j [N_2(N^{-1}(p_i), N^{-1}(p_j), \sqrt{\rho_i\rho_j}) - p_i p_j]$$

A.4. Proof of the α-percentile (5.4)

Proof. Now we derive the inverse function or α-percentile value PD_α of $F(x; p_1, \ldots, p_K; \rho_1, \cdots, \rho_K)$.

It is implied by (4.4) that

$$\alpha = N\left(g^{-1}(PD_\alpha; p_1, \cdots, p_K; \rho_1, \cdots, \rho_K)\right)$$

Then by (4.6) that

$$
\begin{aligned}
PD_\alpha &= g\Big(N^{-1}(\alpha); p_1, \cdots, p_K; \rho_1, \cdots, \rho_K\Big) \\
&= \sum_{i=1}^{K} w_i N\left(\frac{N^{-1}(p_i) + \sqrt{\rho_i}N^{-1}(\alpha)}{\sqrt{1-\rho_i}}\right) \\
&= \sum_{i=1}^{K} w_i L_\alpha^{(i)}, \qquad \text{(A.6)}
\end{aligned}
$$

where $L_\alpha^{(i)}$ is the α-percentile of the i-th sub-portfolio, i.e.,

$$
\begin{aligned}
L_\alpha^{(i)} &= F(\alpha; 1-p_i, 1-\rho_i) \\
&= N\left(\frac{\sqrt{\rho_i}N^{-1}(\alpha) + N^{-1}(p_i)}{\sqrt{1-\rho_i}}\right)
\end{aligned}
$$

REFERENCES

[1] Basel committee on banking supervision. An explanatory note on the Basel II IRB risk weight functions. July 2005.

[2] Basel committee on banking supervision. Guidance on credit risk and accounting for expected credit loss. December 2015.

[3] Gordy M. A risk-factor model foundation for ratings-based bank capital rules. J Financial Intermediation 2003; 12: 199-232.

[4] Merton R. On the pricing of corporate debt: the risk structure of interest rates. J Finance 1974; 29: 449-70.

[5] Pykhtin M. Multi-factor adjustment. Risk 2004; 85-90.

[6] Rosen D, Saunders D. Analytical methods for hedging systematic credit risk with linear factor portfolios. J Econ Dyn Control 2009; 33: 37-52.

[7] Vasicek O. Probability of Loss on Loan Portfolio. KMV Corporation (1987).

[8] Vasicek O. Limiting Loan Loss Probability Distribution. KMV Corporation (1991).

[9] Vasicek O. Loan portfolio value. Risk 2002; 15: 160-2.

[10] Yang B. Point-in-time PD term structure models for multi-period scenario loss projection: methodologies and implementations for IFRS 9 ECL and CCAR stress testing. J Risk Model Valid 2017; 11.