Factor Models for Default Risk

November 5, 2025

1 Regulatory Background

Per regulatory protocol all sufficiently large banks must calculate Incremental Risk Charge (IRC) as part of their Market Risk RWA. Historically, the IRC model was part of the Basel 2.5 revision published in 2009-2010. The IRC model is complementary to Value-at-Risk (VaR) in the sense that VaR is designed to capture general and specific forms of Market Risk, but not idiosyncratic issuer risk. The IRC model applies specifically to debt positons in the Trading Book, e.g. corporate and sovereign debt instruments. IRC is usually modeled in a factor based Monte Carlo framework, wherein it is required that banks estimate the 99.9^{th} quantile of the IRC Loss distribution. The estimate in then converted into Risk Weighted Assets (RWA) via multiplication of a 12.5 risk weight, see (1) below, and serves as a key component of the total Market Risk RWA. This metric represents a 1 in 1,000 event wherein the market value of the applicable Trading Book positions are impacted by issuer default or rating migration.

$$\text{IRC-RWA} \; := 12.5 \cdot \text{IRC} \tag{1}$$

Similar to the IRC model is the Comprehensive Capital Analysis and Review (CCAR) Issuer Default Loss (IDL) model. The IDL model is a key component of the Global Market Shock aspect of the CCAR Program. As with the IRC model the IDL model is designed to capture the risk of issuer default to the Trading Book, but for a broader set of instruments, which include: Corporate Debt, Sovereign Debt, Equities, and Securitized Products. However, the CCAR IDL is usually reported at a less severe confidence level, e.g. 99^{th} quantile from the IDL Loss distribution.

The next round of internalional regulation for the Banking industry coincides the Basel IV Accords, which are likely to be finalized within the next few years (or sooner). As part of the Market Risk component of the Basel IV regulatoions, the Fundamental Review of the Trading Book (FRTB), a Default Risk Charge (DRC) model will be required for all banks of sufficient systemic importance. The DRC model offers yet another example of a factor formulated Monte Carlo based default risk model.

2 Mathematical Prescription

Factor Model Ingredients:

- Factors $F = (F_1, F_2, \cdots, F_m) \in \mathbb{R}^m$ with $\mathbb{E}[F] = 0$ and $Cov(F) = \Sigma \succ 0$
- Factor Loadings $\underline{\beta} = (\beta_1, \beta_2, \cdots, \beta_m) \in \mathbb{R}^m$
- Idiosyncratic Terms $\underline{\epsilon} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n) \in \mathbb{R}^n$

In the context of Default Risk Modeling, the m factors represent sources of common randomness by which to explain the joint dynamics of a collection of issuer specific credit events, e.g. default events, rating grade migrations, etc. To be realistic it is often assumed that the drivers of credit events are common across all firms, but effect them in bespoke ways, i.e. the loadings on each factor are specific to each firm. However, the behavorial characteristics of each firm give rise to a unique form of idiosyncratic risk that also needs to be taken into account. In this case represented by ϵ_i, which is assumed independent of the common factors.

Bringing these ideas together a firm specific 'latent factor' can be constructed as follows:

Issuer Specific Latent Factor Equation

$$X_i := \underline{\beta}_i^{'} F + \epsilon_i \;=\; \sum_{j=1}^{m} \beta_{i,j} F_j + \epsilon_i; \quad i = 1, 2, \cdots, n \tag{2}$$

$$(i)\ \mathbb{E}\left[X_i\right] = 0, \quad (ii)\ Var(X_i) = \underline{\beta}_i^{'} \Sigma \underline{\beta}_i + \sigma^2 = \sum_{j,k} \beta_{i,j} \sigma_{j,k} \beta_{i,k} + \sigma^2$$

or in vector form:

$$\underline{X} := B^{'} F + \epsilon; \quad \text{where} \quad B = [\underline{\beta}_1, \underline{\beta}_2, \cdots, \underline{\beta}_n] \in M_{m,n}(\mathbb{R}) \tag{3}$$

$$(i)\ \mathbb{E}[\underline{X}] = 0 \qquad (ii)\ Cov(\underline{X}) = B^{'} \Sigma B + \sigma^2 I$$

From here we can make some convenient assumptions about the distribution of F and $\underline{\epsilon}$ to make the model tractable. For example,

$$(i)\ F =_D N_m(0, \Sigma) \qquad (ii)\ \underline{\epsilon} =_D N_n(0, \sigma^2 I)$$

In this case it follows that $\underline{X} =_D N_n(0, B^{'} \Sigma B + \sigma^2)$, i.e. the latent factors follow a joint Gaussian distribution.

- $X_i =_D N(0, \underline{\beta}_i^{'} \Sigma \underline{\beta}_i + \sigma^2)$
- $Cov(X_i, X_j) = \underline{\beta}_i^{'} \Sigma \underline{\beta}_j$

Other alternatives include formulating the framework in a multi-variate Student's t distribution, which is desirable if more realistic tail-event clustering[1], is needed in the model. Lastly, there is a rich toolset associated with copulas that can be employed too.

Given a probability of default for the i^{th} issuer, say $p_i \in (0, 1)$ the factor model structure can then be used to simulate dependent, or correlated, credit risk events.

$$\mathbb{P}(\text{Default of Issuer i}) = \mathbb{P}(\Phi(X_i) \leq p_i) = \mathbb{P}(X_i \leq \Phi^{-1}(p_i)) = p_i \tag{4}$$

$$\mathbb{P}(\text{Default of Issuer i and Default of Issuer j}) = \mathbb{P}\left(\Phi(X_i) \leq p_i, \Phi(X_j) \leq p_j\right) \tag{5}$$

The time horizon associated with the model is inherently tied to the time horizon of the default probabilities (PDs). In practice, this depends on the particular modeling use case, e.g. for IRC (details below) the time horizon is 1-year and typically 1-year Through-the-Cycle (TTC) PDs are often utilized. Through-the-Cycle PDs represent long-term or average of the credit-cycle probabiltiies of default, whereas for example a Point-in-Time (PIT) PD is more reflective of default risk locally in the credit-cycle – see [2] for more details.

3 Examples

Vasicek (1987) – Single-factor Model – [4]

$$X_i := \omega Z + (1 - \omega^2)^{1/2} \epsilon_i; \qquad Z, \epsilon_i \overset{\perp}{=}_D N(0, 1) \;\; i = 1, 2, \cdots, n$$

In this case $X_i =_D N(0, 1)$ and $Corr(X_i, X_j) = \omega^2$. The model factors can be interpreted as follows: Z represents a global systemic risk factor, e.g. state of global macro-economy, and ϵ_i denotes issuer specific

[1] It is well known that the Normal distribution has tail dependency coefficients equal to zero, $\underline{\lambda}_{1,2} := \lim_{q \to 0^+} \mathbb{P}\left(X_1 \leq Q_{X_1}(q) | X_2 \leq Q_{X_2}(q)\right) = 0$ whenever the correlation coefficient $\rho \in [-1, 1)$ i.e. rare events are essentially independent.

idiosyncratic terms. This model is simple and easy to understand, but arguably only applies when all issuers are nearly homogenous.

Application: Suppose we have a portfolio of homogenous loans each with a Loss in Default (LID) of $\$100K$ and define the indicator of default of the i^{th} issues as follows:

$$\mathbf{1}_{D_i} := \begin{cases} 1; & \text{if issuer i defaults} \\ 0; & \text{otherwise} \end{cases}$$

and note that $\mathbb{E}[\mathbf{1}_{D_i}] = p_i$. We can then model the portfolio level default risk as:

$$Loss = \sum_{i=1}^{n} \mathbf{1}_{D_i} \$100K$$

In this case the first two moments of the loss distribution are straightforward to compute:

$$\mathbb{E}[Loss] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbf{1}_{D_i} \$100K\right] = \$100K \cdot \sum_{i=1}^{n} p_i$$

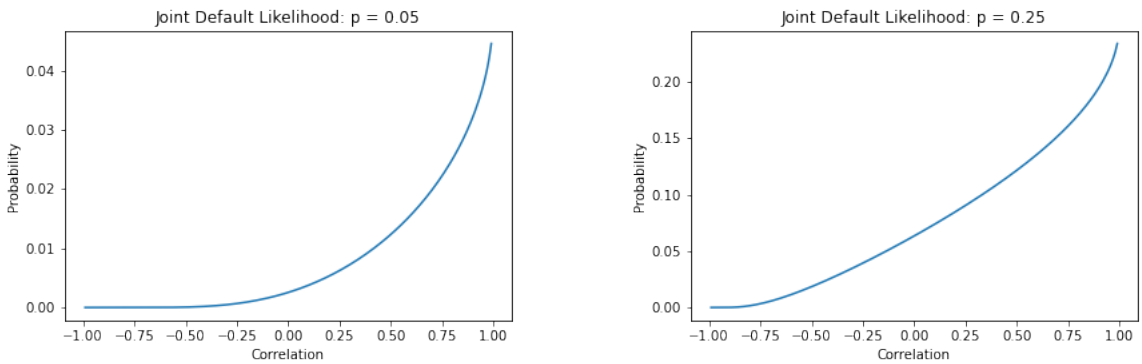
$$\mathbb{E}[Loss^2] = \mathbb{E}\left[\left|\sum_{i=1}^{n} \mathbf{1}_{D_i} \$100K\right|^2\right] = (\$100K)^2 \left(\sum_{i=1}^{n} p_i + \sum_{i \neq j} p_{ij}\right)$$

where above p_{ij} denotes the probability of joint default by the i^{th} and j^{th} issuer:

$$\begin{aligned} p_{ij} : &= \mathbb{E}[\mathbf{1}_{D_i} \cdot \mathbf{1}_{D_j}] = \mathbb{P}(\text{Default of Issuer i and Default of Issuer j}) \\ &= \int_{-\infty}^{\Phi^{-1}(p_i)} \int_{-\infty}^{\Phi^{-1}(p_j)} \phi(x, y; \omega) dx dy \\ &= \int_{-\infty}^{\Phi^{-1}(p_j)} \Phi\left(\frac{\Phi^{-1}(p_i) - \omega \cdot y}{(1 - \omega^2)^{1/2}}\right) \cdot \phi(y) dy \end{aligned}$$

The behavior of the joint default probability is illstrated below in Figure 1. Notice that as the correlation increases in both cases the likelihood of joint default increases significantly.

Figure 1: Joint Default Probability for Bi-variate Standard Gaussian



(a) Marginal Default Probabilities $p_1 = p_2 = 0.05$ (b) Marginal Default Probabilities $p_1 = p_2 = 0.25$

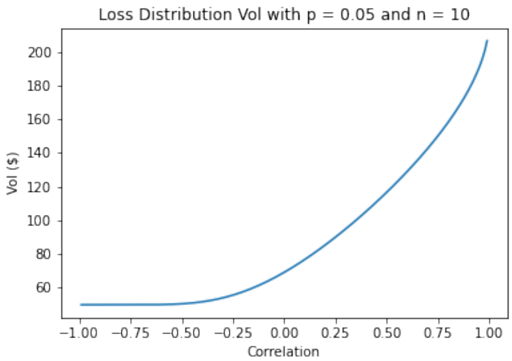
Putting these ideas together if follows that the Loss distribution volatility takes the form:

$$\sigma_{Loss,n} = \$100 \cdot \left(\sum_{i=1}^{n} p_i \cdot (1 - \sum_{i=1}^{n} p_i) + \sum_{i \neq j} p_{ij} \right)^{1/2} \tag{6}$$

See Figure 2 for an illustration of how the Loss volatility behaves as a function of the correlation parameter. Why is this?

- What value does (6) take above when $\omega = 0.0$ and $p_i \equiv p \in (0, 1)$?
- What classic probability distribution can you relate the portfolio Loss to when $\omega = 0.0$?

Figure 2: Joint Default Probability for Bi-variate Standard Gaussian



However, often times the tail-risk, e.g. Value-at-Risk(0.999) is of interest, but is not obtainable in closed form. Monte Carlo simulation techniques play a fundamental role in the practical application of these models.

Incremental Risk Charge (2012) – Multi-factor Model

It is an industry standard to formulate an IRC model using 3 types of common factors and 1 unique idiosyncratic factor period issuer. The 3 types of common factors vary from bank to bank, but often take the form:

- Common global systemic factor
- Industry factors, e.g. Technology, Healthcare, Agriculture, Manufacturing, etc.
- Region factors, e.g. North America, Europe, Middle East and Africa, Asia, etc.

Note, there are many ways this can be done. For example, the standard Global Industry Classfication Standard (GICS) [1] or the Bloomberg Industry Classification Standard (BICS) [3] can be used to create an industry / sector hierarchical structure within the model. Similarly, various region / country taxonomies are avaiable too. The latent factor equation often take the form:

$$X_i := \theta \cdot Z + \gamma_i \cdot I_{Ind(i)} + \beta_i \cdot R_{Reg(i)} + \left(1 - \theta^2 - \gamma_i^2 - \beta_i^2\right)^{1/2} \cdot \epsilon_i, \;\; i = 1, 2, \cdots, n \tag{7}$$

Here $Z, I_{Ind(i)}, R_{Reg(i)}, \epsilon_i \overset{\perp}{=}_D N(0, 1)$, where Z represents a global systemic factor, $I_{Ind(\cdot)}$ are industry specific factors, $R_{Reg(\cdot)}$ denote regional specific factors, and ϵ_i are issuer specific idiosyncratic factors.

- $Ind(\cdot)$ denotes an issuer to industry mapping, e.g. $Ind : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, m\}$
- $Reg(\cdot)$ represnts an issuer to region mapping, e.g. $Reg : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, k\}$

Here each issuer is mapped to one of m possible industries and k possible regions, e.g. Alphabet (GOOG) would map to the Tech industry and the North American region. Under this formulation each issuer latent factor is such that $X_i =_D N(0, 1)$. This is an extension of the classic Vasicek model in many ways that offers a more realistic and richer dependency structure.

Dependency Structure:

$$
Corr(X_i, X_j) = \begin{cases} \theta^2; & \text{if Reg(i)} \neq \text{Reg(j) and Ind(i)} \neq \text{Ind(j)} \\ \theta^2 + \gamma_i \cdot \gamma_j; & \text{if Reg(i)} \neq \text{Reg(j) and Ind(i)} = \text{Ind(j)} \\ \theta^2 + \beta_i \cdot \beta_j; & \text{if Reg(i)} = \text{Reg(j) and Ind(i)} \neq \text{Ind(j)} \\ \theta^2 + \gamma_i \cdot \gamma_j + \beta_i \cdot \beta_j; & \text{if Reg(i)} = \text{Reg(j) and Ind(i)} = \text{Ind(j)} \end{cases}
$$

Default Risk Charge (2023) – Basel III End Game

Thus far default has been treated as a binary event with no regard to the time of default. However, in the real-world this notion is important since there is an interplay between a portfolio of positions and the associated hedges. In order to hedge efficiently, from a financial cost perspective, liquid index products are typically utilized. Furthermore, shorter maturity based products, which are often more liquid and thus can be purchased with lower transaction costs, are typically preferred to longer dated alternatives. To be more realistic the default risk model should take timing of defaults into account accordingly. For example, if a position defaults before a relevant hedge has expired, then the loss will be reduced relative to the case when it defaults after the hedge has expired. In reality, the hedge would likely have been rolled, but in a model designed to capture extreme idiosyncratic issuer default risk, or for stress testing applications, this mismatch in liquidity and expiry should be taken into account.

In order to extend the model to capture this effect many different approaches are available. However, for illustrative purposes we will present one relatively simplistic choice. The key idea is to transform the issuer specific latent factors from (3) into a system of Exponential random variables. For example, utilizing the IRC structure from (7) this can be easily done via the quantile transform method[2]

$$
\tau_i := Exp^{-1}(\Phi(X_i); \lambda_i) \;\; i = 1, 2, \cdots, n \tag{8}
$$

where above $Exp^{-1}(\cdot; \lambda)$ denotes the quantile function of the Exponential distribution with rate parameter $\lambda \in (0, \infty)$.

$$
Exp^{-1}(u; \lambda) = -\frac{log(1-u)}{\lambda}; \;\; u \in (0,1)
$$

In (8) $(\tau_i)_{i=1}^n$ represents a collection of dependent default times, each of which follows an Exponential distribution marginally, i.e.

$$
\tau_i =_D Exp(\lambda_i)
$$

Furthermore, since the Exponential distribution has a constant hazard rate, given a 1-year PD, e.g. p_i, we can compute the equivalent default rate via the following transform:

$$
p_i = \mathbb{P}(X_i \leq \Phi^{-1}(p_i)) = \mathbb{P}(\tau_i \leq \text{1-year}) = 1 - e^{-\lambda_i} \;\; \Rightarrow \;\; \lambda_i = -ln(1-p_i),
$$

from which it follows that for any given time $t \in (0, \infty)$

$$
\mathbb{P}(\tau_i \leq t) = 1 - e^{-\lambda_i \cdot t}.
$$

Note, in general the map $t \mapsto \mathbb{P}(\tau \leq t)$ is known as the term structure of default probabilities and would likely be estimated in a semi-parametric manner.

[2]Recall that for any continuous random variable X with distribution F, $F(X) =_D U(0,1)$

References

[1] Msci: The global industry classification standard, 2025.

[2] Lawrence; Wong Elaine; Diaz-Ledezma Diana Aguais, Scot; Forest. Point-in-time versus through-the-cycle ratings.

[3] Bloomberg. Bloomberg industry classification standard, 2025.

[4] O.A. Vasicek. Probability of loss on loan portfolio. *Finance, Economics, and Mathematics*, pages 143–146, 1987.