



# RAPPORT DE STAGE 3A

**Analyste quantitative  
Crédit Agricole CIB**

Mai - Juillet 2023

Adil TABIB



# Table des matières

1	Modelisation des LGD des financements d'actifs . . . . .	3
1.1	Principe de l'approche de modélisation des prix . . . . .	3
1.2	Calibration de l'effet âge . . . . .	3
1.3	Calibration de l'effet économique . . . . .	4
2	L'optimisation bayésienne pour l'hyperparamètres tuning . . . . .	6
2.1	Définition . . . . .	6
2.2	Un modèle de fonction objectif : Les processus gaussiens . . . . .	6
2.3	Fonctions d'acquisition . . . . .	8
2.4	Analyse des résultats de l'optimisation bayésienne . . . . .	9
3	Modèle de Merton . . . . .	13
3.1	Définition . . . . .	13
3.2	Value-at-Risk . . . . .	14
3.3	Expected shortfall . . . . .	18

## Introduction

Depuis le 1er janvier 2018, les banques ont été mandatées à adopter une nouvelle directive comptable qui régit l'évaluation de la dépréciation des actifs. Cette directive, nommée IFRS 9 - Instruments financiers, prend la place de la précédente norme IAS 39, qui avait fait l'objet de vives critiques à la suite de la crise de 2008. L'objectif de cette directive est principalement d'aligner le niveau de risque porté par les actifs de la banque avec le montant des provisions au passif pour couvrir les pertes éventuelles sur le portefeuille.

En raison de la diversité des activités de la banque, ses actifs sont soumis à de nombreuses sources de risque. Cependant, nous nous focaliserons exclusivement sur le risque de crédit dans la suite. Dans ce cadre, nous définissons les pertes futures comme étant les engagements susceptibles de ne pas être recouvrés en cas de difficultés financières rencontrées par les contreparties. Plus spécifiquement, les modèles ont pour but de calculer les pertes futures moyennes liées au risque de crédit, également désignées sous le sigle ECL (pour Expected Credit Loss)

L'effort de modélisation pour répondre aux exigences de la norme IFRS 9 a donc consisté à mettre en place des modèles prospectifs des trois principales composantes de l'ECL : la courbe de probabilité de défaut (PDC) cumulée, la perte en cas de défaut (LGD) et l'exposition en cas de défaut (EAD).

Lors de ce stage de 4 mois au sein du département MQP (Quantitative Portfolio Modeling), il m'a été demandé de travailler sur plusieurs sujets concernant cette modélisation. Premièrement, il m'a été demandé de comprendre et d'affiner des modèles LGD et PD en travaillant à la fois sur l'automatisation du processus en langage R et sur une procédure d'optimisation des hyperparamètres utilisés dans ces modèles, permettant à la fois d'améliorer les performances du modèle et de réduire le temps de calcul.

Également, il m'a été demandé de travailler sur l'implémentation et l'adaptation aux portefeuilles CA-CIB du papier *Multi-factor adjustment - Pykhtin (2004)*, présentant une méthode analytique de calcul des *Value-at-Risk* (*VaR*) et des *Expected shortfall* (*ES*) d'un portefeuille multifactoriels de type Merton.

# 1 Modelisation des LGD des financements d'actifs

Une composante essentielle des activités de l'équipe MQP consiste à élaborer des modèles de modélisation de la perte en cas de défaut (LGD) ou de la probabilité de défaut (PD) des actifs. Ces actifs pouvant être de différents types, à savoir aéronautique, shipping, immobilier ou encore rail.

Dans la suite, un exemple de modèle LGD de la cas des actifs Aéronautique sera évoquées, les autres modèles se basant sur une approche similaire.

L'objectif est donc de modéliser des paramètres LGD forward looking sur l'horizon de la maturité résiduelle du produit, i.e., à tout instant entre la date de calcul et la date de maturité. Cette modélisation revenant à modéliser les valeurs de l'avion, le but est donc de déterminer la projection future du prix compte-tenu de la dépréciation de l'actif (l'avion) et des anticipations macro-économiques à partir d'un prix à une date donnée.

Durant les quatre mois de mon stage, mon objectif a consisté en une double démarche : premièrement, appréhender le fonctionnement du modèle ci-bas, puis apporter des améliorations ciblées, principalement en ce qui concerne la calibration de l'effet âge. En parallèle, j'ai pris en charge une partie de l'implémentation du code requis pour la projection des valeurs d'actifs (à savoir avion, shipping, immobilier et rail) en utilisant le langage R, tout en cherchant à automatiser au maximum ce processus pour les prochaines projections.

## 1.1 Principe de l'approche de modélisation des prix

Notons  $P_{(t, \hat{age}_t)}^s$  le prix de l'avion de modèle  $s$  (combinaison du constructeur, du type et du sous-type) à la date  $t$  et dont l'âge à cette date est  $\hat{age}_t$ . On propose la formule ci-dessous :

$$P_{(t, \hat{age}_t)}^s = \begin{cases} A_s * \exp(\lambda^s \cdot \hat{age}_t + Y_t^s) & \text{si } \hat{age}_t \leq 20 \\ Scrap_s & \text{sinon} \end{cases}$$

avec :

- $s$  : le modèle d'avion, combinaison du constructeur, du type et du sous-type
- $A_s$  : constante qui dépend du modèle  $s$ . Elle peut être interprétée comme le prix neuf de l'avion
- $\lambda^s$  : la constante de dépréciation de l'avion due à son âge. Elle ne dépend pas de la date  $t$ .
- $Y_t^s$  : le phénomène aléatoire dû à l'économie. Il dépend de la date  $t$  mais non de l'âge.

## 1.2 Calibration de l'effet âge

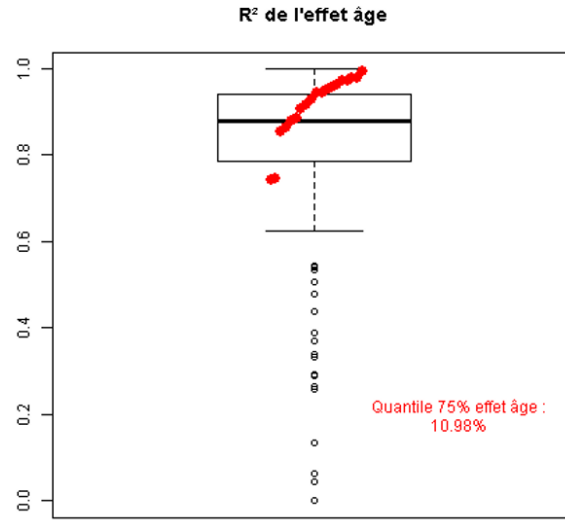
Lors de la proposition de formule ci-dessus, deux hypothèses sont proposées :

- Hypothèse 1 : La dépréciation due à l'effet âge adopte la forme exponentielle
- Hypothèse 2 :  $\lambda^s$  dépend uniquement du type d'actif et non de la valeur actuelle  $A_s$ , de la date ou de l'économie  $Y_t^s$

L'objectif de cette section est de justifier ces hypothèses et calibrer la valeur de  $\lambda^s$  pour chaque type d'avion.

Principe : A partir de la forme définie, on calibre l'effet de l'âge par la relation  $\ln(P_{t, age}^s) = \lambda^s \cdot age_t + Y_t^s + \ln(A_s)$ . Etant donné que  $Y_t^s$  est indépendant de  $age_t$ ,  $\lambda^s$  est le coefficient d'une régression linéaire de  $\ln(P_{t, age}^s)$  sur  $age_t$ .

Calibration : On calibre un modèle de régression linéaire  $\ln(P_{t, age}^s)$   $age_t$  pour chaque avion individuellement (combinaison de manufacturer + type + sous-type). Le pouvoir explicatif de la calibration permet de juger de la pertinence des hypothèses 1 et 2. On utilise le coefficient de détermination  $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$  comme indicateur du pouvoir explicatif, dont la distribution pour tous les types d'avions est illustrée dans la Figure 1. On constate que pour 75% des modèles d'avions, les  $R^2$  sont supérieurs à 80% ce qui justifie la fiabilité des hypothèses 1 et 2. Les points rouges ci-dessous indiquent les  $R^2$  des 19 avions modélisés.



### 1.3 Calibration de l'effet économique

Cette section traite de l'effet économique formalisé par la variable  $Y_t^s$ . Il s'agit d'un modèle statistique de régression utilisant des variables explicatives macroéconomiques. Les bases de données utilisées par l'équipe MQP pour cette calibration sont :

- Prix d'avions selon le constructeur, le type, le sous-type, l'âge et la date
- Indices macro-économiques fournis par Bloomberg
- Statistiques de l'industrie aéronautique sur le site IATA economics

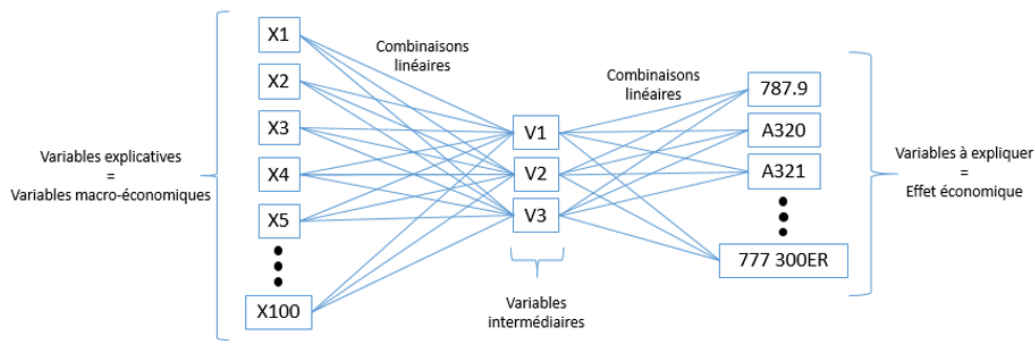
En raison du manque de données sur la variable  $Y_t^s$ , la variable  $r_t^s = Y_t^s - Y_{t-1}^s$  constitue la variable à expliquer car observable à partir de l'historique de prix en prenant la transformation

$$\ln \left( \frac{P_{t,age}^s}{P_{t-1,age}^s} \right) = Y_t^s - Y_{t-1}^s$$

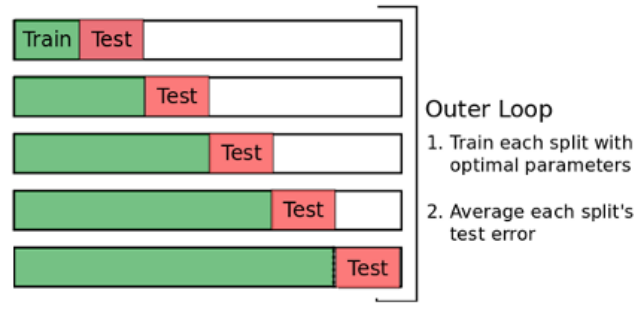
Cette variable est interprétée comme un rendement de prix. Une moyenne sur les différents âges est réalisée pour garder une observation par date.

Pour les sous-types d'avions les plus représentatifs, la modélisation consiste à établir une équation linéaire des variables explicatives macroéconomiques (PIB US, France, zone Euro et Chine, Consommation mondiale de pétrole, Indicateur prix du Brent (USD/Baril), Prix du pétrole, Chômage US, ...). Plusieurs approches existent pour réaliser cet objectif, la difficulté étant le faible nombre d'observations (entre 5 et 50 points selon le type d'avion) en contraste avec la grande variété de variables explicatives (une cinquantaine de variables).

On propose le schéma de régression suivant :

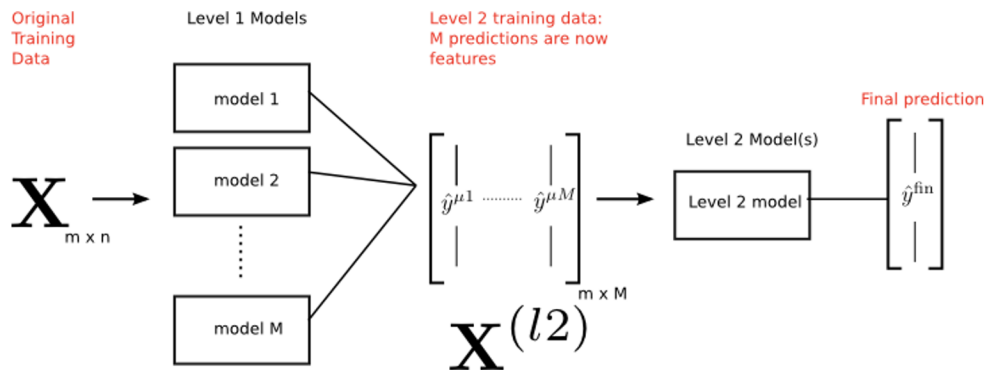


Le principe consiste à utiliser une couche de variables latentes comme condensatrice d'informations, de façon similaire à une analyse PCA. Ces variables latentes, non directement observables, sont inférées au cours de la calibration du modèle. L'objectif étant de trouver un modèle qui permettra de maximiser les performances prédictives, on met en place un processus de validation croisée pour mesurer les performances. Il doit prendre en compte le fait que les données modélisées sont des séries temporelles, c'est-à-dire que la chronologie est importante. Plutôt que de faire une validation croisée "classique", on utilise une validation croisée imbriquée



Le principe est de remonter dans le temps : la période d'apprentissage est bornée à l'année N, la vérification des performances est effectuée sur les années N+1 à N+3. Le choix de mesurer les performances sur une période de 3 ans est motivé par le fait que c'est l'horizon de projection utilisé couramment dans les exercices de production de LGD/PD. Cette validation croisée permet de tester et comparer différents jeux d'hyperparamètres.

Une fois les modèles satellites de régression linéaire retenus, ils sont agrégés par Stacking. Son principe est d'assigner à chaque modèle satellite un poids positif ou nul de telle sorte que la moyenne pondérée minimise l'erreur out-of-sample.



## 2 L'optimisation bayésienne pour l'hyperparamètres tuning

Le modèle précédent nécessite un certain nombre d'hyperparamètres tel que le nombre d'unités dans la couche intermédiaire ou encore la taille du batch size utilisée. Jusqu'à présent, en production, le test de jeux d'hyperparamètres est réalisé via une recherche aléatoire, une stratégie qui consiste à spécifier des distributions de probabilité pour chaque dimension de  $x$ , puis à échantillonner aléatoirement à partir de ces distributions. Cette technique, bien qu'elle soit simple d'utilisation présente quelques difficultés :

- Faire de nombreuses observations similaires qui fournissent des informations redondantes.
- Ne jamais échantillonner près du maximum global.

On propose ici une technique d'hyperparamètres tuning se basant sur l'optimisation bayésienne dont j'ai été chargé de poursuivre le travail d'implémentation d'un projet qui avait été lancé avant le début de mon stage.

### 2.1 Définition

La recherche d'hyperparamètres peut être vue comme un problème de minimisation de la loss, ou de manière équivalente de maximisation de la loss négative. Laissons donc pour le moment le cas spécifique de la recherche d'hyperparamètres et considérons l'optimisation bayésienne sous sa forme la plus générale. On considère pour cela un problème d'optimisation classique : Soit  $\mathcal{X}$  un ensemble fermé borné inclus dans  $\mathbb{R}^d$ . Soit  $f$  une fonction objectif :  $\mathcal{X} \mapsto \mathbb{R}$ , on cherche le point  $x^*$  qui atteint le maximum global de la fonction  $f$  de valeur  $f^*$  c'est-à-dire

$$x^* = \operatorname{argmax}_{x \in \mathcal{X}} f(x) \quad f^* = \max_{x \in \mathcal{X}} f(x) = f(x^*)$$

Remarquons dès à présent que notre choix de s'intéresser à un problème de maximisation est arbitraire, un problème de minimisation se traitera de manière équivalente (en considérant  $-f$ ). Souvent,  $f$  n'a pas d'expression close, ce qui rend le problème d'optimisation globale infaisable en l'état. On supposera cependant que la fonction  $f$  peut être évaluée en n'importe quel point  $x$  dans  $\mathcal{X}$ . Cette évaluation est coûteuse et produit un output  $y = f(x) + \varepsilon$  bruité. L'optimisation bayésienne est une technique d'optimisation basée sur la théorie de la probabilité bayésienne, le but est de s'approcher du maximum de  $f$  en minimisant le nombre d'évaluation de la fonction objectif. Cette optimisation bayésienne est basée sur une approche séquentielle. À chaque itération  $t$ , l'algorithme sélectionne un  $x_{t+1}$  et observe le  $y_{t+1}$  associé. Après  $T$  requêtes, l'algorithme fournit le point  $\tilde{x}$  qui maximise  $f$  parmi les points observés ainsi que le  $\tilde{y}$  associé.

Décrivons à présent comment l'algorithme sélectionne à chaque itération  $t$ , le point  $x_{t+1}$ . Fondamentalement, l'optimisation Bayésienne comporte deux composantes principales :

Un modèle probabiliste de la fonction objectif : L'optimisation bayésienne utilise une approche probabiliste pour modéliser la fonction  $f$  (un prior décrivant notre croyance préalable sur la fonction) et affine séquentiellement son modèle au fur et à mesure que des points sont visités via une mise à jour à posteriori bayésienne (voir ci-bas).

Une fonction d'acquisition : Muni de ce modèle probabiliste, nous pouvons induire séquentiellement des fonctions d'acquisition  $\alpha_t : \mathcal{X} \mapsto \mathbb{R}$  faciles à maximiser et qui tirent parti de l'incertitude dans le postérieur pour guider l'exploration. Intuitivement, la fonction d'acquisition évalue l'utilité de points candidats pour la prochaine évaluation de  $f$  ; donc  $x_{t+1}$  est sélectionné en maximisant  $\alpha_t$ , où l'indice  $t$  indique la dépendance aux données actuellement disponibles. Dans la suite, nous examinons chacune des deux composantes principales (modèles probabilistes de la fonction objectif et les fonctions d'acquisition) à tour de rôle.

### ALGORITHME

### 2.2 Un modèle de fonction objectif : Les processus gaussiens

Il existe plusieurs façons de modéliser la fonction et son incertitude, mais l'approche la plus populaire consiste à utiliser des processus gaussiens (GP).

Un processus gaussien est une collection de variables aléatoires, où chaque collection finie de ces variables aléatoires suit une loi normale multidimensionnelle. Il est défini par

- Une fonction moyenne  $m(x)$
- Une fonction de covariance  $k(x, x')$  qui renvoie la similarité entre deux points.

Un exemple classique de choix de  $m$  et  $k$  est :

$$m(x) = k(x, x') = \exp\left(-\frac{1}{2}(x - x')^T(x - x')\right)$$

On remarque avec le choix de  $k$ , que la corrélation est de plus en plus faible à longueur que  $x$  et  $x'$  s'éloignent.

Étant données des observations  $\mathbf{f} = [f(x_1), f(x_2), \dots, f(x_t)]$  en  $t$  points, nous souhaitons faire une prédiction sur la valeur de la fonction en un nouveau point  $\tilde{x}$ . Cette nouvelle valeur de fonction  $\tilde{f} = f(\tilde{x})$  est conjointement normalement distribuée avec les observations  $\mathbf{f}$  de sorte que :

$$[\mathbf{f}, f(\tilde{x})] \sim \mathcal{N}\left(0, \begin{pmatrix} K(X, X) & K(X, \tilde{x}) \\ K(\tilde{x}, X) & k(\tilde{x}, \tilde{x}) \end{pmatrix}\right)$$

où  $K(X, X)$  est la  $t \times t$  matrice telle que  $K(X, X)_{i,j} = k(x_i, x_j)$ ,  $K(X, \tilde{x})$  est le vecteur de taille  $t$  tel que  $K(X, \tilde{x})_i = k(x_i, \tilde{x})$  et ainsi de suite.

D'après une propriété des vecteurs gaussiens, la distribution conditionnelle  $\tilde{f} | \mathbf{f}$  est également normale :

$$\tilde{f} | \mathbf{f} \sim \mathcal{N}(\mu(\tilde{x}), \sigma^2(\tilde{x}))$$

Où

$$\mu(\tilde{x}) = K(\tilde{x}, X) K(X, X)^{-1} \mathbf{f}$$

$$\sigma^2(\tilde{x}) = K(\tilde{x}, \tilde{x}) - K(\tilde{x}, X) K(X, X)^{-1} K(X, \tilde{x})$$

En utilisant cette formule, nous pouvons estimer la distribution de la fonction en tout nouveau point  $\tilde{x}$ . La meilleure estimation de la valeur de la fonction est donnée par la moyenne  $\mu(\tilde{x})$ , et l'incertitude est donnée par la variance  $\sigma^2(\tilde{x})$ .

### Incorporation de mesures bruyantes

Dans la partie précédente, nous avons supposé que la fonction que nous estimons est exempt de bruit et renvoie toujours la même valeur  $f(x)$  pour une entrée  $x$  donnée. À présent, nous n'observons plus directement les valeurs des fonctions  $f(x)$ , mais nous observons des outputs bruités  $y(x) = f(x) + \varepsilon$ .

Nous allons incorporer un terme de bruit supplémentaire en supposant à présent que

$$\mathbb{E}[(y(x) - m(x))(y(x') - m(x')))] = k(x, x') + \bar{\sigma}^2$$

Étant donné maintenant des observations bruitées  $\mathbf{y} = [y(x_1), y(x_2), \dots, y(x_t)]$  en  $t$  points, la nouvelle valeur de la fonction  $\tilde{f} = f(\tilde{x})$  est toujours conjointement normalement distribuée avec les observations  $\mathbf{y}$  de sorte que :

$$[\mathbf{y}, f(\tilde{x})] \sim \mathcal{N}\left(0, \begin{pmatrix} K(X, X) + \bar{\sigma}^2 I_t & K(X, \tilde{x}) \\ K(\tilde{x}, X) & k(\tilde{x}, \tilde{x}) \end{pmatrix}\right)$$

Où  $I_t$  est la matrice identité de taille  $t \times t$

Et la distribution conditionnelle  $\tilde{f} | \mathbf{y}$  vérifie maintenant :

$$\tilde{f} | \mathbf{y} \sim \mathcal{N}(\mu(\tilde{x}), \sigma^2(\tilde{x}))$$

Où



$$\mu(\tilde{x}) = K(\tilde{x}, X)(K(X, X) + \sigma^2 I_t)^{-1} \mathbf{f}$$

$$\sigma(\tilde{x}) = K(\tilde{x}, \tilde{x}) - K(\tilde{x}, X)(K(X, X) + \sigma^2 I_t)^{-1} K(X, \tilde{x})$$

## 2.3 Fonctions d'acquisition

Maintenant que nous avons un modèle probabiliste de la fonction objectif, nous allons définir nos fonctions d'acquisition afin de choisir le prochain point à échantillonner. Dans le cas d'un processus gaussien, la fonction d'acquisition prend la moyenne et la variance à chaque point  $x$  et calcule une valeur qui indique dans quelle mesure il est souhaitable de procéder au prochain échantillonnage à cette position. Une fonction d'acquisition doit faire un compromis entre l'exploration et l'exploitation.

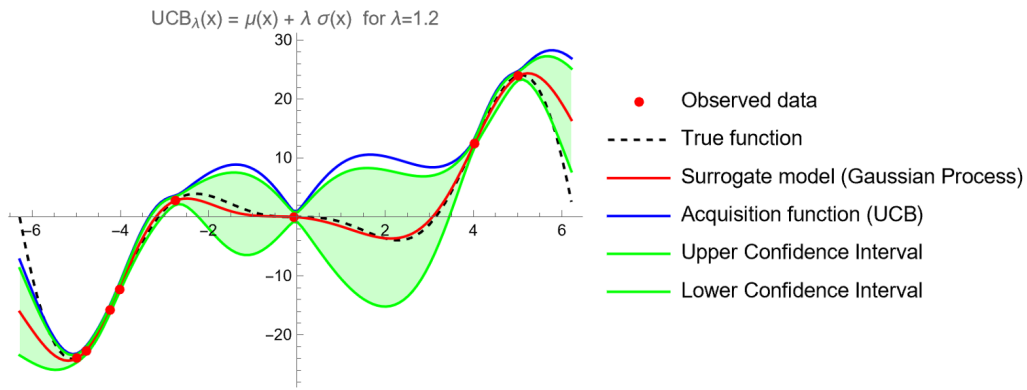
Présentons quatre exemples de fonctions d'acquisition usuelles :

- *Upper confidence bound* : Cette fonction d'acquisition est définie comme suit :

$$UCB(x) = \mu(x) + \lambda \sigma(x)$$

avec  $\lambda \geq 0$

UCB est une somme pondérée de la performance attendue capturée par  $\mu(x)$  du processus gaussien, et de l'incertitude  $\sigma(x)$ , capturée par l'écart type du *GP*. Lorsque  $\lambda$  est petit, l'optimisation bayésienne favorisera les solutions censées être performantes, c'est-à-dire ayant un  $\mu(x)$  élevé. Au contraire, lorsque  $\lambda$  est grand, elle récompense l'exploration de zones actuellement inexplorées dans l'espace de recherche.



- *Lower confidence bound* : Cette fonction d'acquisition est définie comme suit :

$$LCB(x) = \mu(x) - \lambda \sigma(x)$$

avec  $\lambda \geq 0$

La *lower confidence bound* accorde une importance plus grande à l'exploitation, plus l'incertitude  $\sigma$  est faible, plus la valeur d'acquisition est élevée, ce qui conduit à l'exploitation de points ayant une faible incertitude.

- Probability of Improvement (*PI*) :

Notons  $x^*$  la meilleure solution que nous ayons trouvée jusqu'à présent et définissons « l'amélioration »  $I(x) = \max(f(x) - f(x^*), 0)$ .

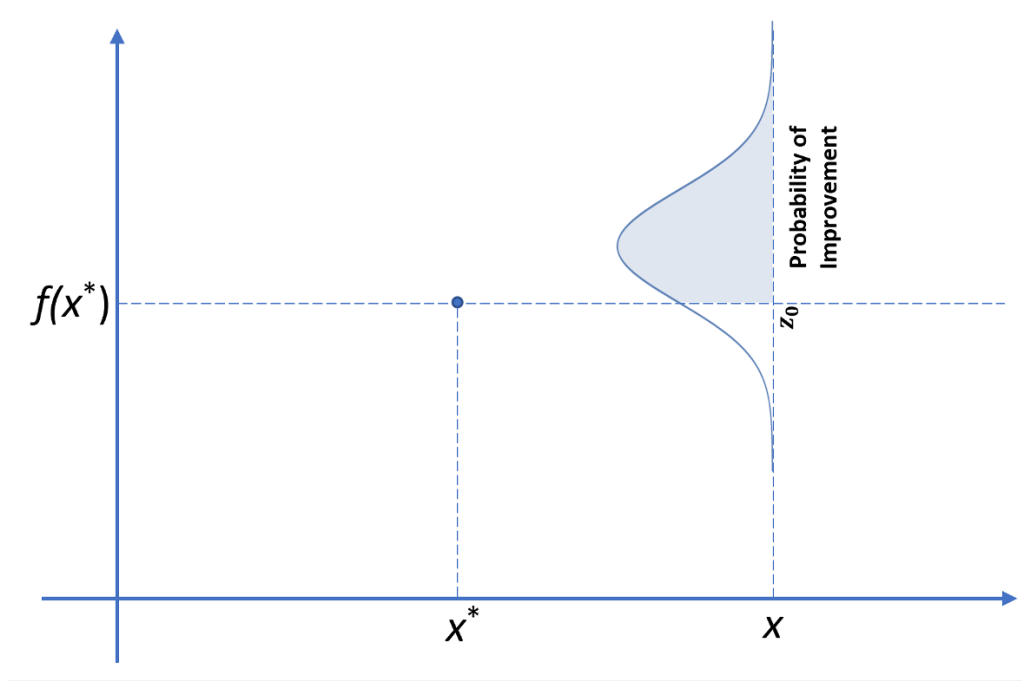
Comme dans le cas d'un processus gaussien on a  $f(x) \sim \mathcal{N}(\mu(x), \sigma^2(x))$ , notons  $f(x) \sim \mu(x) + \sigma(x)z$  où  $z \sim \mathcal{N}(0, 1)$  et on a  $I(x) = \max(\mu(x) + \sigma(x)z - f(x^*), 0)$

La « *probability of improvement* » est ainsi donnée par

$$PI(x) = \mathbb{P}(f(x) > f(x^*)) = \mathbb{P}(I(x) > 0) = \mathbb{P}(\mu(x) + \sigma(x)z - f(x^*) > 0)$$

$$= \mathbb{P} \left( z > \frac{f(x^*) - \mu(x)}{\sigma(x)} \right) = \Phi(-z_0)$$

Où  $\Phi(\bullet)$  est la fonction de répartition d'une loi normale standard,  $z_0 = \frac{f(x^*) - \mu(x)}{\sigma(x)}$



— Expected Improvement ( $EI$ ) :

Le principal inconvénient de la "probability of improvement" est qu'elle ne tient pas compte de l'ampleur de l'amélioration ; nous ne voulons pas favoriser les petites améliorations (même si elles sont très probables) au détriment des plus importantes. "L'Expected improvement" tient compte de cela :

$$\begin{aligned} EI(x) &= \mathbb{E}(I(x)) = \int_{-\infty}^{+\infty} \max(\mu(x) + \sigma(x)z - f(x^*), 0) \phi(z) dz \\ &= \int_{z_0}^{+\infty} (\mu(x) + \sigma(x)z - f(x^*)) \phi(z) dz \\ &= (\mu(x) - f(x^*)) \int_{z_0}^{+\infty} \phi(z) dz + \sigma(x) \int_{z_0}^{+\infty} z e^{-\frac{z^2}{2}} dz \\ &= (\mu(x) - f(x^*)) \Phi(-z_0) + \sigma(x) \phi(z_0) \end{aligned}$$

avec  $\phi(\bullet)$  la densité d'une loi normale standard.

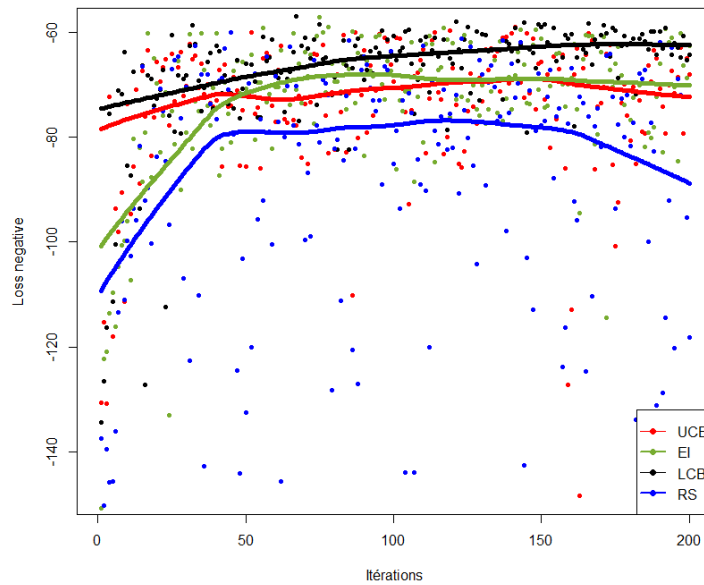
## 2.4 Analyse des résultats de l'optimisation bayésienne

### 1. Efficacité de l'optimisation bayésienne

Afin d'évaluer l'efficacité de l'optimisation bayésienne et de la comparer à la recherche naïve randomisée d'hyperparamètres, plusieurs *calibrations* du schéma linéaire ont été lancées. Ces calibrations peuvent être de 4 types différents : Un premier utilisant la recherche aléatoire (*RS*) pour l'exploration des hyperparamètres, les 3 autres types implémentant chacun une optimisation bayésienne avec un choix de fonction d'activation (*UCB*, *LCB*, *EI*). Notons que l'ensemble des autres paramètres nécessaires à la calibration sont égaux.

De plus, ayant fait le choix d'exposer l'optimisation bayésienne dans un cadre de maximisation, nous allons présenter les résultats en maximisant la loss négative de notre modèle.

Une première visualisation consiste à afficher l'évolution de la loss négative en fonction des itérations pour 4 calibrations de différents types (*UCB*, *EI*, *LCB*, *RS*). Dans ce graphique le nombre d'itérations des calibrations a été fixé à 200. C'est à dire chaque calibration teste 200 combinaisons d'hyperparamètres.

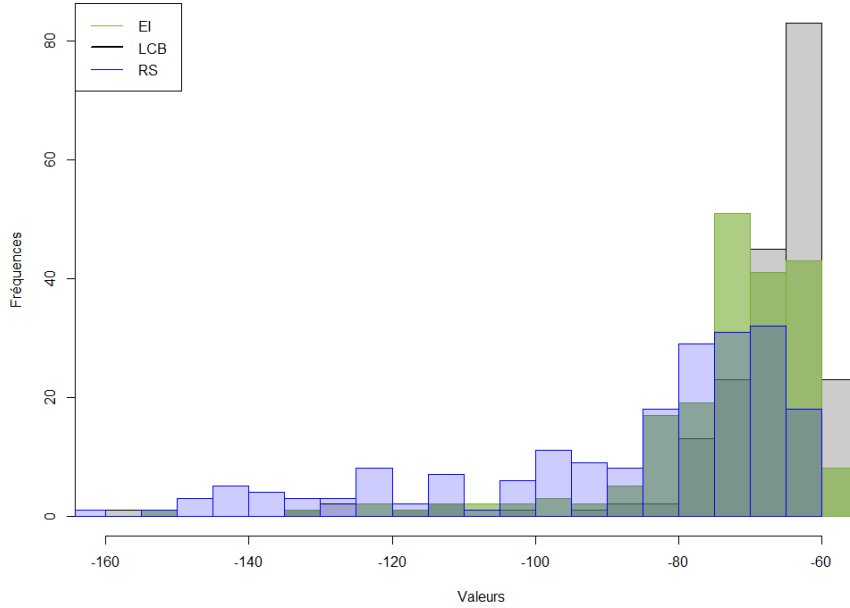


*Évolution de la loss négative en fonction des itérations pour 4 types de calibrations (UCB/EI/LCB/RS)*

Globalement, on s'aperçoit que l'optimisation bayésienne fournit des choix d'hyperparamètres permettant une meilleure maximisation de la loss négative.

Notons que si les valeurs spécifiques de la loss dépendent des 4 calibrations (avec une présence d'aléa), l'allure globale reste la même avec d'autres calibrations de même type et conforte nos observations.

Ce deuxième graphique permet d'établir une discrimination entre les fonctions d'activation :



Histogramme des valeurs des loss négatives (EI/LCB/RS)

On observe que dans notre cas, la fonction d'activation  $LCB$  semble fournir de meilleurs résultats avec une concentration plus importante pour les valeurs importantes de la loss négative. Remarquons que pour des raisons de lisibilité, la calibration  $UCB$ , fournissant des résultats très semblable à la calibration  $EI$ , a été retirée du graphique.

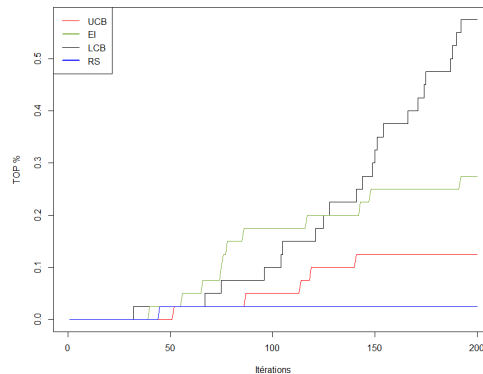
Afin de mieux quantifier ces deux observations, nous avons décidé de définir une métrique adaptée à ces comparaisons :

Considérons comme précédemment 4 calibrations de 4 types différents ( $UCB$ ,  $LCB$ ,  $EI$ ,  $RS$ ) réalisant chacune  $n$  itérations. (Dans les graphiques précédent  $n = 200$ ). Il y a donc  $4n$  combinaisons différentes d'hyperparamètres testées.

Parmi ces  $4n$  itérations, on conserve les 5% correspondant aux meilleures itérations, c'est à dire ayant la loss négative la plus élevée. Le concept de  $Top5\%$  pour une calibration est défini comme la proportion des résultats générés par cette calibration qui font partie de ces meilleures itérations.

Par exemple, si le  $Top5\%$  d'une calibration est de 0.8 , cela signifie que cette calibration a réussi à identifier 80% des combinaisons d'hyperparamètres ayant généré les 5% des loss négatives les plus élevées.

On généralise cette notion en définissant  $Top5\%(i)$  pour  $i$  entre 1 et 200 qui correspond à cette même notion en ne prenant en compte que les  $i$  premières itérations de chaque calibration.



Top% en fonction des itérations

Ainsi dans notre cas, la *LCB* a surpassé les autres fonctions d'acquisition qui ont été biaisées vers l'exploration ou l'exploitation excessive. On remarque que la *LCB* a un *Top5%* égale à 0.6 c'est à dire que 60% des combinaisons d'hyperparamètres ayant généré les 5% des loss négatives les plus élevées l'ont été par la calibration *LCB*. Le *random search* quant à lui ne participe qu'avec 3% et conforte nos observations précédentes.

## 2. Stabilité de l'optimisation bayésienne

Nous avons constaté que l'optimisation bayésienne fournit de bons résultats dans notre cas, en trouvant des combinaisons d'hyperparamètres efficaces qui entraînent une loss négative élevée. À présent, nous souhaitons aborder la question de la stabilité de l'optimisation bayésienne. Il est souhaitable que deux calibrations effectuées avec les mêmes paramètres produisent des estimations similaires des variables à expliquer.

Notons dans la suite  $x$  un vecteur d'observation des variables explicatives de test qu'on suppose centré réduit et  $y$  le vecteur des variables à expliquer associé.

Le résultat d'une calibration peut être vu comme la donnée d'une matrice de coefficients  $C$  et d'un vecteur de biais  $b$  tels que l'estimation de  $y$  vaut  $Cx + b$ .

Comme le biais est (théoriquement) égal à la moyenne de l'output  $y$ , l'étude de la stabilité de l'optimisation bayésienne revient donc à comparer l'output  $C_i x$  et  $C_j x$  pour  $C_i, C_j$  2 matrices issues de deux calibrations aux paramètres identiques.

Plus précisément et afin d'obtenir des résultats plus lisses on compare  $C_i X$  et  $C_j X$  pour  $X$  une matrice correspondant à  $N$  tirages de vecteurs gaussiens avec la structure de covariance de  $x$ .

On calcule par la suite les coefficients de disparité normalisé :

$$d_{i,j} = \frac{\text{colmeans}((C_i X - C_j X)^2)}{\text{var}(y)}$$

Où  $\text{colmeans}(M)$  est la moyenne sur les colonnes de  $M$ .

On présente ici les résultats dans le cadre d'une calibration *LCD*, fournissant dans notre cas et comme vu précédemment, de meilleurs résultats.

Considérons ainsi  $n = 5$  calibrations *LCD* dans le cadre d'un schéma de modélisation *PD*. Ces  $n$  calibrations fournissent des matrices  $(C_1, \dots, C_n)$  telles que les moyennes des coefficients de disparité  $\frac{1}{\binom{n-1}{2}} \sum_{1 \leq i < j \leq n} d_{i,j}$  sont données par :

### TABLEAU1

En conclusion, on remarque que dans ce cas, les outputs sont plutôt semblables avec une moyenne de similarité (= 1 – disparité) égale à 89%. Le dernier output correspondant à la variable à expliquer *ZMig\_IF\_Sov* semble être le plus volatil avec un coefficient de similarité égal à seulement 77%.

À titre de comparaison, l'étude de cette même stabilité dans le cadre de la recherche aléatoire mène aux résultats suivants :

### TABLEAU2

On remarque ainsi que l'optimisation bayésienne n'influe pas sur la stabilité des résultats. Cette approche semble donc être avantageuse dans notre cas.

### 3 Modèle de Merton

Une autre composante essentielle des travaux accomplis par l'équipe MQP réside dans l'évaluation des pertes associées à son portefeuille. Cela inclut la détermination des Value-at-Risk (VaR) et de l'Expected Shortfall (ES) par le biais de méthodes d'estimation.

Jusqu'à présent des simulations de Monte Carlo ont été employées pour estimer ces quantités. Une partie de mon stage s'est focalisé sur l'adaptation d'une approche analytique basée sur les travaux de Michael Pykhtin (voir [?]). L'objectif étant de mettre en œuvre cette méthode afin d'évaluer les Value-at-Risk et Expected Shortfall des portefeuilles MQP, en s'affranchissant de la nécessité de recourir aux simulations de Monte Carlo, qui sont particulièrement chronophage. De plus, j'ai dû adapter cette méthode aux particularités propres du portefeuille MQP, notamment la gestion de plusieurs prêts pour un même emprunteur.

#### 3.1 Définition

Commençons par mettre en place un modèle de Merton multifactoriel. Pour cela, considérons un portefeuille de prêts à  $M$  emprunteurs distincts. Chaque emprunteur  $i$  a  $N_i$  prêts d'expositions  $w_{ij}$  et de maturités  $m_{i,j}$  (avec  $m_{i,j} \leq m_{i,j+1}$ ). L'emprunteur  $i$  fera défaut pour son  $j$ -ème prêt avec une probabilité  $p_{ij}$ , on suppose pour simplifier que  $p_{ij} = 1 - (1 - p_i)^{m_{ij}}$  où  $p_i$  est une probabilité globale que l'emprunteur  $i$  fasse défaut dans un délai choisi (typiquement, un an).

Il y a défaut lorsqu'une variable continue  $X_i$  décrivant la situation financière de l'emprunteur  $i$  à l'horizon tombe en dessous d'un seuil. Nous supposons que les variables  $X_i$  (qui peuvent être interprétées comme les rendements standardisés des actifs) ont une distribution normale standard. Le seuil de défaillance de l'emprunteur  $i$  pour son prêt  $j$  est donné par  $N^{-1}(p_{ij})$ , où  $N^{-1}(\bullet)$  est l'inverse de la fonction de répartition de la loi normale.

Nous supposons que les rendements des actifs dépendent linéairement de  $N$  facteurs de risque systémiques normalement distribués avec une matrice de corrélation de rang plein. Ces facteurs systémiques représentent le secteur d'activité, la géographie, l'économie mondiale ou tout autre indice pertinent susceptible d'affecter les défaillances des emprunteurs de manière systémique.

Le rendement standardisé de l'actif de l'emprunteur  $i$  est déterminé par une certaine combinaison spécifique à l'emprunteur de ces facteurs systémiques  $Y_i$  (appelée facteur composite) :

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

Où  $\xi_i$  est le choc idiosyncratique de loi normale standard.  $r_i$  mesure la sensibilité de l'emprunteur  $i$  au risque systémique.

Comme il est plus commode de travailler avec des facteurs indépendants, nous supposons que les  $N$  facteurs systémiques corrélés d'origine sont décomposés en  $N$  facteurs systémiques normaux indépendants  $\{Z_k\}$ . La relation entre les facteurs  $\{Z_k\}$  et les facteurs composites est donnée par :

$$(2) \quad Y_i = \sum_{k=1}^N \alpha_{ik} Z_k$$

où les coefficients  $\alpha_{ik}$  vérifient la condition  $\sum_{k=1}^N \alpha_{ik}^2 = 1$ .

La corrélation entre deux emprunteurs distincts  $i$  et  $s$  est donnée par  $\rho_{i,s} = r_i r_s \sum_{k=1}^N \alpha_{ik} \alpha_{sk}$

Si l'emprunteur  $i$  fait défaut pour son prêt  $j$ , le montant de la perte est déterminé par sa *loss given default* stochastique  $Q_{ij}$ . On suppose que les  $\{Q_{ij}\}_j$  sont *i.i.d* de moyenne  $\mu_i$  et d'écart-type  $\sigma_i$ .

On suppose également que ces variables LGD sont indépendantes entre elles ainsi que de toutes les autres variables du modèle. Nous n'émettons aucune hypothèse spécifique sur la distribution de probabilité des variables LGD.

Enfin, la perte du portefeuille (*portfolio loss*)  $L$  peut s'écrire comme la moyenne des pertes individuelles :

$$(3) \quad L = \sum_{i=1}^M L_i = \sum_{i=1}^M \sum_{j=1}^{N_i} w_{ij} 1_{\{X_i < N^{-1}(p_{ij})\}} Q_{ij}$$

Une approche traditionnelle pour estimer les quantiles de la distribution des pertes du portefeuille est la simulation de Monte Carlo. Si le portefeuille est suffisamment important pour être considéré comme à *grain fin*, la majeure partie du risque idiosyncrasique du portefeuille est diversifiée et les pertes du portefeuille sont principalement dues aux facteurs systémiques. Dans ce cas, l'équation (3) peut être remplacée par la distribution de la perte limite d'un portefeuille à granularité infinie. La perte limite est donnée par la perte attendue conditionnellement au facteur de risque systémique, comme on peut le montrer en appliquant la loi des grands nombres conditionnellement au facteur.

$$(4) \quad L^\infty = E(L|Y) = \sum_{i=1}^M \sum_{j=1}^{N_i} w_{ij} \mu_i N \left[ \frac{N^{-1}(p_{ij}) - r_i \sum_{k=1}^N \alpha_{ik} Z_k}{\sqrt{1 - r_i^2}} \right]$$

Bien que l'équation (4) soit beaucoup plus simple que l'équation (3), on ne sait pas exactement quelle doit être la taille du portefeuille pour que l'équation (3) devienne exacte.

Dans ce qui suit, nous concevons une méthode analytique pour calculer les quantiles de queue et les espérances de queue de la perte de portefeuille  $L$  donnée par l'équation (3).

### 3.2 Value-at-Risk

Nous souhaitons calculer le quantile à un niveau de confiance  $q$  de la perte du portefeuille  $L$ . Nous noterons ce quantile comme  $t_q(L)$ . Supposons que nous ayons construit une variable aléatoire  $\bar{L}$  telle que son quantile au niveau  $q$ ,  $t_q(\bar{L})$ , peut être calculée analytiquement et est suffisamment proche de  $t_q(L)$ . Nous considérerons la perte du portefeuille comme la variable  $\bar{L}$  plus la perturbation  $U$  définie comme  $U = L - \bar{L}$ . Pour décrire l'échelle de la perturbation, introduisons également une variable perturbée  $L_\varepsilon = \bar{L} + \varepsilon U$ . Pour des niveaux de confiance suffisamment élevés  $q$ ,  $t_q(L_\varepsilon)$  peut être calculée par un développement en  $\varepsilon$  autour de  $t_q(\bar{L})$ .

En faisant un développement de Taylor d'ordre 2 et en fixant  $\varepsilon = 1$ , nous pouvons espérer calculer le quantile de perte du portefeuille comme suit :

$$(5) \quad t_q(L) = t_q(\bar{L}) + \left. \frac{dt_q(L_\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} + \frac{1}{2} \left. \frac{d^2 t_q(L_\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0}$$

La dérivée première est donnée par l'espérance de la perturbation conditionnelle à  $\bar{L} = t_q(\bar{L})$  :

$$(6) \quad \left. \frac{dt_q(L_\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = E(U | \bar{L} = t_q(\bar{L}))$$

tandis que la dérivée seconde est donnée par :

$$(7) \quad \left. \frac{d^2 t_q(L_\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0} = - \left. \frac{1}{f_{\bar{L}}(l)} \frac{d}{dl} (f_{\bar{L}}(l) \text{ var}[U | \bar{L} = l]) \right|_{l=t_q(\bar{L})}$$

où  $f_{\bar{L}}(\bullet)$  la fonction de densité de  $\bar{L}$ .

Dans la suite,  $\bar{L}$  sera définie comme la distribution limite des pertes :

$$(8) \quad \bar{L} = l(\bar{Y}) = \sum_{i=1}^M \mu_i \sum_{j=1}^{N_i} w_{ij} \hat{p}_{ij}(\bar{Y})$$

où  $\bar{Y}$  est le facteur de risque systémique unique ayant une distribution normale standard, et  $\hat{p}_{ij}(\bar{Y})$  est la probabilité de défaut de l'emprunteur  $i$  concernant le prêt  $j$ , conditionnellement à  $\bar{Y} = y$ , qui est donnée par :

$$(9) \quad \hat{p}_{ij}(\bar{Y}) = N \left[ \frac{N^{-1}(p_{ij}) - a_i \bar{Y}}{\sqrt{1 - a_i^2}} \right]$$

où  $a_i$  est le *effective factor loading* pour l'emprunteur  $i$ , son expression sera détaillée dans la suite.

Comme  $\bar{L}$  est une fonction déterministe décroissante de  $\bar{Y}$ , le quantile de  $\bar{L}$  au niveau  $q$  peut être calculé analytiquement simplement comme la valeur de la fonction en  $Y = N^{-1}(1 - q)$  :

$$(10) \quad t_q(\bar{L}) = l(N^{-1}(1 - q))$$

Et peut donc être utilisé comme l'approximation d'ordre zéro de  $t_q(L)$ .

Notons que les dérivées de la VaR dans les équations (6) and (7) sont données par des expressions conditionnelles à  $\bar{L} = t_q(\bar{L})$ . Comme  $\bar{L}$  est une fonction déterministe décroissante de  $\bar{Y}$ , il est équivalent de conditionner en  $\bar{Y} = N^{-1}(1 - q)$ .

Les dérivées première et seconde de la VaR peuvent maintenant être énoncées comme suit :

$$(11) \quad \left. \frac{dt_q(L_\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = E(U | \bar{Y} = N^{-1}(1 - q)) = 0$$

$$(12) \quad \left. \frac{d^2 t_q(L_\varepsilon)}{d\varepsilon^2} \right|_{\varepsilon=0} = - \frac{1}{n(y)} \frac{d}{dy} \left( n(y) \frac{v(y)}{l'(y)} \right) \Big|_{y=N^{-1}(1-q)}$$

$$= \frac{1}{l'(y)} \left[ v'(y) - v(y) \left( \frac{l''(y)}{l'(y)} + y \right) \right] \Big|_{y=N^{-1}(1-q)}$$

où  $v(\bullet)$  est la variance conditionnelle de  $U$  définie par  $v(y) = \text{var}[U | \bar{Y} = y]$ ,  $v'(\bullet)$  sa dérivée.

$l'(\bullet)$  et  $l''(\bullet)$ , respectivement la dérivée, première et seconde de  $l(\bullet)$  et  $n(\bullet)$  est la densité de la loi normale standard (qui apparaît ici comme la fonction de densité de  $\bar{Y}$ ).

En utilisant l'équation (8) :

$$(13) \quad l'(y) = \sum_{i=1}^M \mu_i \sum_{j=1}^{N_i} w_{ij} \hat{p}'_{ij}(y)$$

avec :

$$(14) \quad \hat{p}'_{ij}(y) = - \frac{a_i}{\sqrt{1 - a_i^2}} n \left[ \frac{N^{-1}(p_{ij}) - a_i y}{\sqrt{1 - a_i^2}} \right]$$

De même,

$$(15) \quad l''(y) = \sum_{i=1}^M \mu_i \sum_{j=1}^{N_i} w_{ij} \hat{p}''_{ij}(y)$$

avec :

$$(16) \quad \hat{p}''_{ij}(y) = - \frac{a_i^2}{1 - a_i^2} \frac{N^{-1}(p_{ij}) - a_i y}{\sqrt{1 - a_i^2}} n \left[ \frac{N^{-1}(p_{ij}) - a_i y}{\sqrt{1 - a_i^2}} \right]$$

Ainsi, on a

$$(17) \quad \mathbf{t}_q(\mathbf{L}) = \mathbf{l}(N^{-1}(1 - q)) + \frac{1}{2} \frac{1}{l'(y)} \left[ \mathbf{v}'(y) - \mathbf{v}(y) \left( \frac{l''(y)}{l'(y)} + y \right) \right] \Big|_{y=N^{-1}(1-q)}$$

À présent, nous allons établir une formule pour les  $a_i$  :



Supposons qu'il existe une relation linéaire entre le facteur de risque systémique  $\bar{Y}$  et les facteurs systémiques originaux  $\{Z_k\}$  :

$$(18) \quad \bar{Y} = \sum_{k=1}^N b_k Z_k$$

où les coefficients  $b_k$  vérifient la condition  $\sum_{k=1}^N b_k^2 = 1$ .

Une condition importante pour déterminer les  $a_i$  est d'exiger que  $\bar{L}$  vérifie  $\bar{L} = E(L|\bar{Y})$  pour toutes compositions d'un portefeuille.

Pour calculer  $E(L|\bar{Y})$  écrivons  $Y_i = \bar{\rho}_i \bar{Y} + \sqrt{1 - \bar{\rho}_i^2} \eta_i$ , avec  $\eta_i$  une loi normale standard indépendante de  $\bar{Y}$  et  $\bar{\rho}_i$  est la corrélation entre  $Y_i$  et  $\bar{Y}$  :

$$(19) \quad \bar{\rho}_i = \text{cov}(Y_i, \bar{Y}) = \sum_{k=1}^N \alpha_{ik} b_k$$

En utilisant ces notations, on a également  $X_i = r_i \bar{\rho}_i \bar{Y} + \sqrt{1 - (r_i \bar{\rho}_i)^2} \zeta_i$  où  $\zeta_i$  est une loi normale standard indépendante de  $\bar{Y}$ .

Ainsi, on a

$$(20) \quad E(L|\bar{Y}) = \sum_{i=1}^M \mu_i \sum_{j=1}^{N_i} w_{ij} N \left[ \frac{N^{-1} (p_{ij}) - r_i \bar{\rho}_i \bar{Y}}{\sqrt{1 - r_i \bar{\rho}_i^2}} \right]$$

En comparant les équations (19) et (8), on a donc par identification :

$$(21) \quad a_i = r_i \bar{\rho}_i = r_i \sum_{k=1}^N \alpha_{ik} b_k$$

Le choix des  $\{b_k\}$  n'est pas unique et cette méthode fonctionne pour différents ensembles  $\{b_k\}$ . Idéalement, il faut choisir  $\{b_k\}$  tel que  $t_q(L)$  et  $t_q(\bar{L})$  sont proches mais trouver un tel ensemble n'est pas facile.

On peut choisir les  $\{b_k\}$  de manière à ce que le facteur de risque effectif unique optimal  $\bar{Y}$  ait la plus grande corrélation possible avec les facteurs de risque composites  $\{Y_i\}$ , mathématiquement, on doit résoudre l'optimisation sous contraintes suivant :

$$(22) \quad \max_{\{b_k\}} \sum_{i=1}^M c_i \text{cor}(\bar{Y}, Y_i)$$

tel que  $\sum_{k=1}^N b_k^2 = 1$

En prenant en compte l'équation (19) on trouve

$$(23) \quad b_k = \sum_{i=1}^M \frac{c_i}{\lambda} \alpha_{ik}$$

Où  $\lambda$  est une constante choisie telle que  $\{b_k\}$  vérifie la condition  $\sum_{k=1}^N b_k^2 = 1$ .

Le choix des coefficients  $\{c_i\}$  n'est pas clair, un choix possible est donné par :

$$(24) \quad c_i = \mu_i \sum_{j=1}^{N_i} w_{ij} N [N^{-1} (p_{ij})]$$

Nous devons également établir une formule de  $v(\bullet)$  :

Pour simplifier les calculs et les notations, on suppose dans cette partie que chaque emprunteur a un seul prêt. Cette supposition pour le calcul de variance n'est pas très restrictive et le framework précédent s'adapte bien :

On notera dans cette partie  $w_i = \sum_{j=1}^{N_i} w_{ij}$  le principal du prêt de l'emprunteur  $i$ ,  $\tilde{p}_i = \frac{\sum_{j=1}^{N_i} w_{ij} p_{ij}}{\sum_{j=1}^{N_i} w_{ij}}$  sa probabilité de défaut globale (pondérée par les EAD),  $Q_i$  sa *loss given default* stochastique de moyenne  $\tilde{\mu}_i = \mu_i$  et de variance  $\tilde{\sigma}_i^2 = \frac{\sigma_i^2 \sum_{j=1}^{N_i} w_{ij}^2}{(\sum_{j=1}^{N_i} w_{ij})^2}$ .

Définissons également la matrice  $(\rho^Y)_{(i,j)}$  telle que  $\rho_{i,j}^Y$  est égal à la corrélation conditionnelle des emprunteurs  $i$  et  $j$  qui est donnée par :

$$(25) \quad \rho_{i,j}^Y = \frac{r_i r_j \sum_{k=1}^N \alpha_{ik} \alpha_{jk} - a_i a_j}{\sqrt{1 - a_i^2} \sqrt{1 - a_j^2}}$$

Bien que  $\rho_{i,j}^Y$  n'a le sens de corrélation conditionnelle des actifs que pour les emprunteurs distincts  $i$  et  $j$ , nous étendons l'équation (25) pour inclure le cas  $i = j$ .

De plus, en prenant en compte que  $\bar{L}$  est une fonction déterministe de  $\bar{Y}$  et en conditionnant par  $\{Z_k\}$  on a :

$$(26) \quad v(y) = \text{var} [U|\bar{Y} = y] = \text{var} [L|\bar{Y} = y] = \text{var} [E(L|\{Z_k\})|\bar{Y} = y] + E(\text{var} [L|\{Z_k\})|\bar{Y} = y)$$

Notons  $v_\infty(y) = \text{var} [E(L|\{Z_k\})|\bar{Y} = y]$  et  $v_{GA}(y) = E(\text{var} [L|\{Z_k\})|\bar{Y} = y)$

Un calcul permet d'obtenir

$$(27) \quad v_\infty(y) = \sum_{i=1}^M \sum_{j=1}^M w_i w_j \mu_i \mu_j N_2(N^{-1}[\hat{p}_i(y)], N^{-1}[\hat{p}_j(y)], \rho_{i,j}^Y) - \hat{p}_i(y) \hat{p}_j(y)$$

$$(28) \quad v_{GA}(y) = \sum_{i=1}^M w_i^2 (\mu_i^2 [N_2(N^{-1}[\hat{p}_i(y)], N^{-1}[\hat{p}_i(y)], \rho_{i,i}^Y)] + \tilde{\sigma}_i^2 \hat{p}_i(y))$$

Avec pour dérivée

$$(29) \quad v'_\infty(y) = 2 \sum_{i=1}^M \sum_{j=1}^M w_i w_j \mu_i \mu_j \hat{p}'_i(y) \left( N \left( \frac{N^{-1}[\hat{p}_j(y)] - \rho_{i,j}^Y N^{-1}[\hat{p}_i(y)]}{\sqrt{1 - (\rho_{i,j}^Y)^2}} \right) - \hat{p}_j(y) \right)$$

$$(30) \quad v'_{GA}(y) = 2 \sum_{i=1}^M w_i^2 \hat{p}'_i(y) \left( \mu_i^2 \left( 1 - 2 N \left( \frac{N^{-1}[\hat{p}_i(y)] - \rho_{i,i}^Y N^{-1}[\hat{p}_i(y)]}{\sqrt{1 - (\rho_{i,i}^Y)^2}} \right) \right) - \tilde{\sigma}_i^2 \right)$$

où  $N_2(\bullet, \bullet, \bullet)$  est la fonction de répartition de la loi normale bivariee :

$$(31) \quad N_2(x, z, r) = \int_{-\infty}^z n(y) N \left[ \frac{x - ry}{\sqrt{1 - r^2}} \right] dy$$

$\hat{p}_i(y)$  définie de la même manière que l'équation (9) :

$$(32) \quad \hat{p}_i(y) = N \left[ \frac{N^{-1}(\tilde{p}_i) - a_i y}{\sqrt{1 - a_i^2}} \right]$$

### 3.3 Expected shortfall

ES à un niveau de confiance  $q$  pour  $L$  est définie comme la perte attendue au-dessus du quantile  $q$  :

$$(33) \quad ES_q(L) = E(L | L \geq t_q(L))$$

On peut réécrire l'équation (33) comme :

$$(34) \quad ES_q(L) = \frac{1}{1-q} \int_q^1 t_s(L) ds$$

Réécrivons l'équation (5) comme :

$$(35) \quad t_s(L) = l(N^{-1}(1-s)) - \frac{1}{2n(y)} \frac{d}{dy} \left( n(y) \frac{v(y)}{l'(y)} \right) \Big|_{y=N^{-1}(1-s)}$$

L'ES peut donc être exprimé par :

$$(36) \quad ES_q(L) = \frac{1}{1-q} \int_{-\infty}^{N^{-1}(1-q)} n(y) l(y) dy - \frac{1}{2} \int_{-\infty}^{N^{-1}(1-q)} \frac{d}{dy} \left( n(y) \frac{v(y)}{l'(y)} \right) dy$$

En substituant  $l(\bar{Y})$  de l'équation(8) dans l'équation (36) et en évaluant l'intégrale, on obtient :

$$(37) \quad ES_q(L) = \frac{1}{1-q} \sum_{i=1}^M \sum_{j=1}^{N_i} w_{ij} \mu_i N_2 [N^{-1}(\mathbf{p}_{ij}), N^{-1}(1-q), \mathbf{a}_i] - \frac{\mathbf{n}[N^{-1}(1-q)]}{2(1-q)} \frac{\mathbf{v}[N^{-1}(1-q)]}{l'[N^{-1}(1-q)]}$$

# Bibliographie

- [1] Pablo ACOSTA. “Promotion dynamics the Peter Principle : Incumbents vs. external hires”. In : *Labour Economics* 17.6 (2010), p. 975-986.
- [2] David L DICKINSON et Marie Claire VILLEVAL. “The Peter Principle : An Experiment”. In : (2007).
- [3] Julius KANE. “Dynamics of the Peter principle”. In : *Management Science* 16.12 (1970), B-800.
- [4] Edward P LAZEAR. *The Peter Principle : promotions and declining productivity*. 2001.
- [5] Erik J OLSSON et Carlo PROIETTI. “Promoting the best as an incentive : reply to Pluchino et al. on the Peter Principle”. In : *DOU* : <https://doi.org/10.13140/2.1> 2201 (2015).
- [6] Alan L PATZ. “The development of managerial ability : a mathematical analysis”. In : *Interfaces* 5.3 (1975), p. 29-35.
- [7] Laurence J PETER, Raymond HULL et al. *The peter principle*. 04 ; RMD, PN6231. M2 P4. Souvenir Press London, 1969.
- [8] Pawel SOBKOWICZ. “Dilbert-Peter model of organization effectiveness : computer simulations”. In : *arXiv preprint arXiv :1001.4235* (2010).