

**Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 14/03/2022**

Par : **Pidio Amos HOUNGBEDJI**

Titre : **Analyse des facteurs explicatifs de l'octroi de crédit
sur les financements structurés**

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de la filière

Entreprise : **SOCIETE GENERALE**

Nom : **Hillairet Caroline**



Signature :

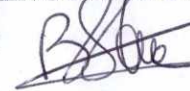
Membres présents du jury de l'Institut
des Actuaires

Directeur du mémoire en entreprise :

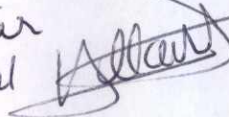
Nicolas Zec

Nom : **M. BENGUESSOUM Stanislas**

Signature :



NEPRI François Xavier
(visio) p/o C. Hillairet



**Autorisation de publication et de
mise en ligne sur un site de
diffusion de documents actuariels**

*(après expiration de l'éventuel délai de
confidentialité)*

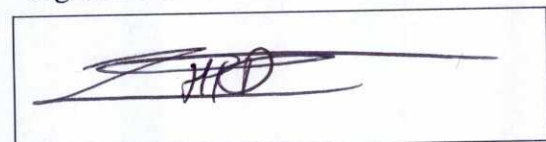
Secrétariat :

Signature du responsable entreprise



Bibliothèque :

Signature du candidat



Résumé

Le processus d'approbation du crédit dans le cadre des financements structurés est riche, requiert de nombreuses expertises et produit une grande quantité d'informations. Les outils de type raro(r)c (risk adjusted return on regulatory capital) constituent des métriques d'évaluation de la performance ajustée du risque de crédit et sont indispensables dans le processus d'octroi de financements structurés. En partant des informations générées dans le cadre de l'octroi de crédit sur les financements structurés, cette étude permet d'analyser le poids de la rentabilité et des autres facteurs dans la décision d'approbation d'une demande de financement. Il en ressort que les niveaux de rentabilité sont différents en moyenne et en volatilité selon les types de financement. Il est de ce fait inapproprié de définir un seuil d'approbation unique pour valider toutes les transactions. Cette étude permet également d'établir que l'approbation de ces financements ne peut être de façon exclusive associée à la bonne rentabilité des transactions. L'environnement concurrentiel, le besoin de formuler assez souvent des offres commerciales, la particularité du marché et des pratiques opérationnelles dans certains pays ou encore des avantages liés à certaines formes de syndication sont autant de facteurs qui complètent de façon significative l'analyse de la rentabilité ajustée au risque. La prise en compte de ces facteurs ainsi que des interactions entre ces facteurs peut conduire à la possibilité d'un système de scoring pour l'approbation des financements structurés.

Mots-clés : financements structurés, risque de crédit, raro(r)c, seuil d'approbation

Abstract

The credit approval process for structured finance is rich, requires a lot of expertise and produces a lot of information. Rarorc (risk adjusted return on regulatory capital) tools are metrics for assessing credit risk adjusted performance and are indispensable in the structured finance approval process. Based on information generated in the context of the credit granting process for structured finance, this study analyzes the weight of profitability and other factors in the decision to approve a financing request. The study shows that the profitability levels assessed from the rarorc are variable depending on the type of financing. It is therefore inappropriate to define a single approval threshold to validate all transactions. This study also shows that the approbation of these financings cannot be exclusively associated with the good profitability of the transactions. The competitive environment, the need to formulate commercial offers often enough, the particularity of the market and operational practices in certain countries or the advantages linked to certain forms of syndication are all factors that significantly complete the analysis of risk-adjusted profitability. Taking these factors and the interactions between them into account can lead to the possibility of a scoring system for the approval of structured finance.

Keywords : structured finance, credit risk, rarorc, hurdle rate,

Remerciements

Je tiens tout d'abord à remercier M. BENGUESSOUM Stanislas, M. PETER Eric et M. SEVIN Benjamin pour leur confiance et pour m'avoir donné l'opportunité d'effectuer mon stage de fin de formation au sein de l'équipe GLBA/CRE/MOD de la SOCIÉTÉ GÉNÉRALE. Je les remercie pour m'avoir guidé tout au long de cette étude. Leur aide a été précieuse dans la compréhension du sujet et sur les réflexions menées.

Un grand merci à tous les membres de cette équipe. Leur accueil et leur bonne humeur m'ont permis de m'intégrer rapidement et de réaliser ce stage dans les meilleures conditions possibles. Je remercie tous ceux qui ont de près ou de loin contribué à la réalisation de ce mémoire notamment à travers leurs conseils. Ils ont été à l'écoute en répondant à mes questions et en m'aidant lorsque j'avais des zones d'ombres.

Table des matières

| | |
|---|-----------|
| Résumé | * |
| Abstract | * |
| Introduction | 1 |
| I La modélisation du risque de crédit | 3 |
| I.1 Définition et enjeux | 3 |
| I.2 Les paramètres du risque de crédit | 4 |
| I.3 Le risque de crédit sur un portefeuille de financement | 6 |
| II Présentation du portefeuille de financements structurés | 12 |
| II.1 Définition | 12 |
| II.2 L'échantillon de financements étudiés | 13 |
| II.3 Caractéristiques en terme de réduction des risques | 13 |
| II.4 Typologie de financement | 15 |
| III Présentation de la métrique de rentabilité | 18 |
| III.1 Expression des indicateurs RARO(R)C | 18 |
| III.2 Les indicateurs intermédiaires | 19 |
| III.3 Distribution du RARORC par type de financement | 20 |
| III.4 Distribution du rarorc par type de syndication | 22 |
| III.5 Pouvoir discriminant du RARORC sur le stock de financements étudiés | 24 |
| III.6 Conclusion partielle | 29 |
| IV Modélisation de la décision d'approbation | 31 |
| IV.1 Théorie du modèle linéaire généralisé. | 31 |
| IV.2 Approche par les méthodes d'apprentissage automatique | 36 |
| IV.3 Présentation de l'algorithme du Gradient Boosting | 40 |
| IV.4 Mesure d'influence relative des variables | 45 |
| IV.5 Application pratique | 49 |
| IV.6 Conclusion partielle | 59 |

| | |
|--|-----------|
| Conclusion | 60 |
| A Statistiques pour la présélection des variables | b |
| B Résultats complémentaires sur l'influence des variables | f |

Introduction

L'activité bancaire consiste avant tout à maîtriser les risques financiers qui peuvent être de différentes natures. Dans le cadre des opérations de financement, le risque de crédit est prépondérant. Le risque de crédit mesure la perte potentielle (donc probabilisée) à laquelle s'expose la banque du fait de l'incapacité de l'emprunteur à la rembourser. La gestion de ce risque est donc inhérente à l'activité de financement et d'investissement. A cet effet, les banques ont développé des outils permettant d'estimer le montant de fonds propre qu'elles doivent allouer à chaque transaction. Les intérêts facturés par la banque visent en partie à contrebalancer ce risque. Plus le risque est élevé, plus les intérêts correspondants le seront afin d'assurer la rentabilité de la transaction. La mesure de rentabilité communément utilisée est le RARORC (Risk Adjusted Return On Regulatory Capital qui fournit une estimation de la rentabilité d'une opération ramenée au capital immobilisé, en tenant compte de l'incertitude des cash flows, des coûts de transaction (notamment les coûts de liquidité) et des pertes potentielles. Les banques peuvent alors définir un seuil minimal de RARORC à partir duquel l'opération générera de la valeur pour les actionnaires.

Contrairement à la clientèle particulière où la décision d'octroi est systématisée en fonction de paramètres objectifs, la décision de crédit sur les financements structurés se base sur une analyse individuelle des dossiers. En effet, L'octroi de crédit sur les financements structurés est un processus riche qui regroupe de nombreuses expertises (économiques, industriels, financières, commerciales, etc.) et produit un grand nombre d'information. Dès lors, le RARORC n'apparaît plus comme le seul critère pris en compte pour l'octroi de crédit.

L'objectif de cette étude est donc d'évaluer les déterminants de l'octroi de crédit sur les financements structurés et d'identifier les informations complétant le RARORC dans la décision d'octroi.

Dans un premier chapitre, les paramètres permettant d'évaluer le risque de crédit et intervenant dans la construction de la métrique RARORC sont présentés. Cette métrique est en effet calculée à partir d'indicateurs de risque de crédit qui sont modélisés dans le cadre de la gestion de ce risque. Dans un deuxième chapitre, la définition et l'échantillon des financements étudiés sont présentés. Ensuite, un inventaire des instruments financiers et des techniques de réduction de risques utilisées dans ces transactions permet de mettre en évidence leur complexité. Le troisième chapitre permet de présenter la métrique de rentabilité et ses composantes. Ce

chapitre permettra également d'évaluer la politique d'octroi à partir d'un unique seuil fixé pour la rentabilité ajustée au risque de crédit. Le quatrième chapitre de ce document permet de modéliser la décision d'approbation des transactions afin d'en ressortir les facteurs importants d'une part et, d'autre part, d'étudier les interactions entre facteurs conduisant à des utilisations différentes de la métrique RARORC.

Chapitre I

La modélisation du risque de crédit

Le risque de crédit est la principale source de perte sur un portefeuille de financement. A juste titre, l'évaluation de ce risque est une étape incontournable dans le processus d'octroi. Les établissements de crédit ont développé des techniques permettant de quantifier ce risque afin d'orienter leur décision. Dans ce chapitre nous présentons la définition et les enjeux de la modélisation du risque de crédit ainsi que les approches de calcul de l'exposition en cas de défaut, de la probabilité de défaut et de la perte en cas de défaut. Dans une deuxième partie, il sera présenté le modèle monofactoriel à l'origine du calcul du capital réglementaire. Ces outils d'évaluation du risque de crédit interviennent également dans le calcul de la rentabilité des opérations de financements structurés.

I.1 Définition et enjeux

Le risque de crédit est défini comme le risque de perte liée à l'évolution de la qualité de signature d'un émetteur. Ce risque peut se matérialiser sous la forme — du changement de sa note (upgrade ou downgrade) telle celle émise par les grandes agences de notations, — d'une variation de son spread de crédit, — d'un événement de crédit tel le défaut de paiement ou la restructuration de sa dette. Ces trois risques sont corrélés. Une augmentation brutale du niveau du spread émetteur augmente la probabilité d'un événement de crédit. De la même manière, un changement de notation influe fortement sur la probabilité de défaut d'un émetteur. Toutes les institutions financières accumulent une grande quantité de risque de crédit : soit directement par l'intermédiaire de leurs portefeuilles de placement, soit indirectement sous la forme de risques de contrepartie dans leurs portefeuilles d'actifs et de produits dérivés. Le risque principal inhérent au portefeuille de placement (les actifs étant destinés à être conservés jusqu'à l'échéance) est celui de non-remboursement (du capital et/ou des intérêts) dû au défaut (c'est-à-dire à l'insolvabilité) du débiteur.

L'enjeu que représente la modélisation de ce risque est donc très important. Il est, en effet, indispensable de pouvoir contrôler l'exposition au risque de crédit sur chaque contrepartie ainsi

que l'évolution de cette exposition. La mesure du risque de crédit est donc une étape essentielle dans la mise en place d'une politique globale de gestion des risques et de rentabilité au sein de la banque. En effet, une mesure précise et quantitative de ce risque permet de :

- créer un langage commun du risque grâce à une échelle de notation unique afin de communiquer sur la base d'une mesure objective servant à comparer les risques de crédits et de contreparties ;
- aider à l'évaluation de la rentabilité et à la structuration des crédits ;
- renforcer le pilotage des risques et gérer activement le portefeuille d'engagements de la banque afin d'optimiser notamment l'allocation des fonds propres aux différentes activités en fonction de leurs performances. Ainsi, la mesure du risque et de la rentabilité effectuée sur le portefeuille de crédit, sous différents axes d'analyse, permet une gestion plus facile du portefeuille : cessions, couvertures...
- de diversifier et de réduire le risque en imposant, par exemple, des limites à l'exposition au risque de crédit par contrepartie, ligne de métiers, secteurs géographiques et industriels.
- de fournir des informations fiables sur son intégrité financière aux régulateurs, aux investisseurs et aux agences de notation ;

La mise en place des processus de gestion du risque correspond à l'évolution des accords réglementaires [1].

I.2 Les paramètres du risque de crédit

L'approche de modélisation de risque de crédit considérée dans cette étude se base sur les estimations internes de la banque sur les composantes du risque. Ces composantes comprennent le système de notation interne de la banque, la Perte en Cas de Défaut (Loss Given Default - LGD), l'Exposition en cas de Défaut (EAD) et le Coefficient de Conversion en Crédit hors-bilan (CCF).

I.2.1 L'exposition en cas de Défaut

L'exposition en cas de défaut (Exposure at Default, EAD) est définie comme le montant de la banque exposé au risque de défaut de la contrepartie. L'EAD comprend les expositions (prêts, créances, produits à recevoir, opérations de marché, etc.) inscrites au bilan ainsi qu'une proportion des expositions hors-bilan calculée à l'aide de facteurs de conversion (Credit Conversion Factor, CCF). Étant donné que le risque porté par un prêt à terme, un crédit renouvelable ou d'autres instruments de crédit n'est pas le même, le nominal de ces instruments est converti en EAD. Elle est calculée sans tenir compte des garanties et des collatéraux. Le CCF dépend de l'instrument de crédit concerné, du type de contrepartie et de la maturité.

I.2.2 La matrice de transition des probabilités

Un système de notation interne permet de couvrir la notation de l'ensemble des contreparties, de servir de base au calcul des indicateurs de performance ajustés au risque et aux méthodologies de gestion de portefeuille qui en découlent.

En supposant que le système de notation de la banque à n grades. Une matrice de transition d'un an est estimée à partir des transitions de notation observées dans le système de notation. La matrice résultante est notée $\mathbf{P}(1)$. Les entrées de la matrice sont notées par $p_{kl}, k, l = 1, \dots, n$ où p_{kl} est la probabilité qu'un emprunteur de notation k passe à la note l dans un délai d'un an. La matrice $\mathbf{P}(1)$ possède les propriétés suivantes :

1. Les entrées de $\mathbf{P}(1)$ sont non négligeables, c'est-à-dire $p_{kl} \geq 0, k, l = 1, \dots, n$.
2. Toutes les lignes de $\mathbf{P}(1)$ se somment à un, c'est-à-dire $\sum_{l=1}^n p_{kl} = 1, l = 1, \dots, n$.
3. La dernière colonne $p_{kn}, k = 1, \dots, n - 1$ contient les probabilités par défaut (PD) sur un an des notes de notation $1, \dots, n - 1$. La probabilité de défaut à un an est la probabilité qu'une contrepartie soit en défaut de paiement à un horizon de un an.
4. L'état par défaut est considéré comme étant absorbant. Cette hypothèse n'est pas vérifiée en pratique.

S'il est supposé que les transitions de notation sont markoviennes, (c'est-à-dire qu'elles dépendent uniquement de la note de notation actuelle d'un emprunteur) et que les probabilités de transition sont homogènes dans le temps (c'est-à-dire que la probabilité d'une transition de notation entre deux points de temps dépend uniquement de la durée de l'intervalle de temps), alors il est possible d'appliquer la théorie des chaînes de Markov pour construire des matrices de transition $\mathbf{P}(h)$ pour une année complète arbitraire h simplement en multipliant $\mathbf{P}(1)$ avec elle-même :

$$\mathbf{P}(h) = \underbrace{\mathbf{P}(1) \cdot \dots \cdot \mathbf{P}(1)}_{h \text{ times}}$$

Une fois $\mathbf{P}(h)$ calculé, la probabilité de défaut à l'horizon h $p_{kn}(h), k = 1, \dots, n - 1$ peut être lue à partir de la dernière colonne de la ligne k .

Les probabilités de défaut avec un horizon temporel d'un an sont généralement l'un des résultats du système de notation de la banque. Elle peut être évaluée selon différentes méthodes.

- Une approche de type scoring : Elle consiste à associer directement une probabilité de défaut à une contrepartie grâce à des modèles statistiques internes. De tels modèles donnent la probabilité directe de défaut associée à une contrepartie. De cette probabilité de défaut est déduite la notation de la contrepartie sur l'échelle de notation interne. Ces modèles étant basés sur l'historique des défauts observés sur le portefeuille de la banque, ce type d'approche ne peut être envisagé que sur les portefeuilles de clients pour lesquels la banque dispose d'un échantillon interne important, avec des informations historiques suffisamment approfondies et comprenant suffisamment de cas de défaut.

- Une approche de type cartographie hors clientèle de détail, le système de notation consiste à attribuer une note à chaque contrepartie selon une échelle interne dont chaque niveau correspond à une probabilité de défaut déterminée à partir d'un historique observé par Standard & Poor's sur plus de 20 ans. La note attribuée à une contrepartie est généralement proposée par un modèle, puis elle est ajustée et validée par les experts notateurs de la filière Risques à la suite de l'analyse individuelle de chaque contrepartie. Les modèles de notation de contreparties se structurent notamment en fonction de la nature de la contrepartie (entreprises, institutions financières, entités publiques...), du pays, de la zone géographique, de la taille de l'entreprise (mesurée le plus souvent par son chiffre d'affaires annuel). Les modèles de notation des entreprises s'appuient sur des modélisations statistiques (méthodes de régression) de défaut des clients. Ils combinent des éléments quantitatifs issus des données financières évaluant la pérennité et la solvabilité des contreparties et des éléments qualitatifs d'appréciation des dimensions économiques et stratégiques.¹

I.2.3 La perte en cas de Défaut

La perte en cas de défaut (LGD) est le pourcentage de perte (sur l'EAD) encourue suite au défaut d'une contrepartie. Elle constitue une perte économique qui se mesure en prenant en compte tous les éléments inhérents à la transaction, ainsi que les frais engagés pour le recouvrement de la créance en cas de défaut de la contrepartie. La perte en cas de Défaut reflète le degré de garantie et de collatéralisation d'un prêt.

I.3 Le risque de crédit sur un portefeuille de financement

Si les risques de défaut étaient indépendants, les paramètres présentés dans la précédente section suffiraient pour leur évaluation. En effet, par la loi des grands nombres, la variance des pertes du portefeuille de prêt deviendra arbitrairement faible si le nombre de prêts est suffisamment important et sans aucune concentration de volume. Il en résulterait des actifs et des passifs appariés de façon déterministe. Les risques de défaut ne sont toutefois pas indépendants puisque tous les emprunteurs sont touchés par le risque macroéconomique entraînant des défauts dépendants. Dans un mauvais environnement macroéconomique, les pertes de crédit sont plus élevées que prévue. Dans un scénario bénin, les pertes sont plus faibles. Pour éviter la faillite pendant les années de récession, les banques doivent détenir un coussin de fonds propres dans le but faire face à des situations pires que celles auxquelles elles pourraient s'attendre.

1. RAPPORT SUR LES RISQUES 2017 Version du 16/03/2017 SOCIÉTÉ GÉNÉRALE

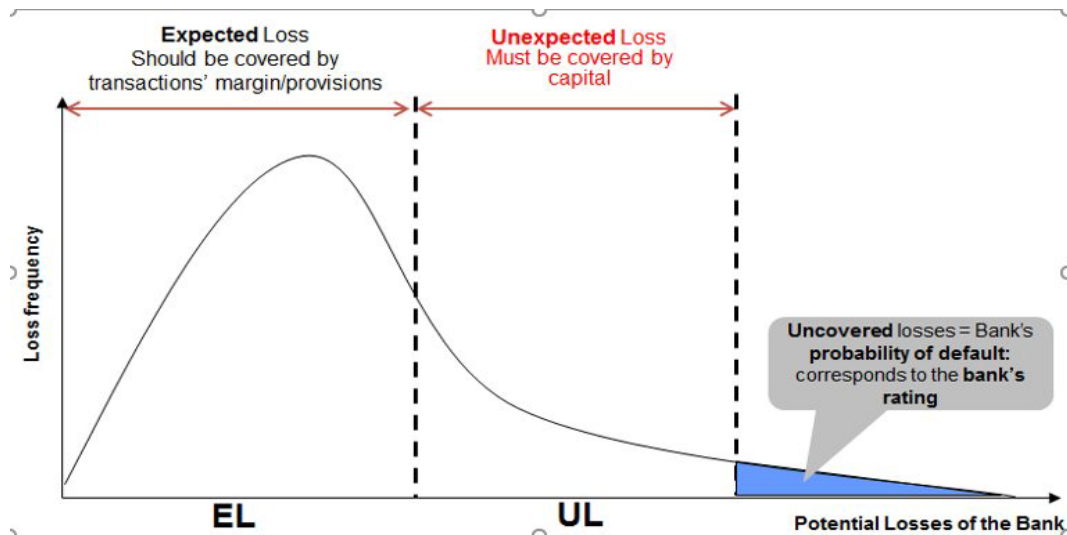


FIGURE I.1 – Distribution de perte sur un portefeuille de crédit

Le Comité de Bâle sur le contrôle bancaire propose des exigences minimales concernant la taille du coussin de fonds propres de la part des régulateurs [1]. Une approche basée sur les notations internes, permet aux banques de calculer des coussins de fonds propres minimaux à partir d'estimations internes des probabilités de défaut et des pertes en cas de défaut en utilisant la fonction dite « K ». Le modèle mono-factoriel expliquant cette fonction ainsi que l'expression de la fonction sont présentés ci-dessous.

I.3.1 Le modèle mono-factoriel

Les règles de capital fondées sur les notations, y compris celles de l'accord de Bâle attribuent une exigence de fonds propres à un instrument sur la seule base de ses caractéristiques propres. Le Modèle de base de Gordy-Vasicek permet d'établir que les règles de capital basées sur les notations peuvent être réconciliées avec la classe générale des modèles de VaR (Value at Risk) de crédit. Les deux hypothèses du modèle de Gordy (2002) et Vasicek (2003) sont :

- **hypothèse de facteur unique** : il n'y a qu'un seul facteur de risque systématique qui détermine les corrélations entre les débiteurs ;
- **hypothèse de granularité** : aucune exposition dans un portefeuille ne représente plus qu'une part arbitrairement petite de l'exposition totale.

Il s'agit d'un modèle mono-factoriel statique qui s'applique à un portefeuille de créances en principe bien diversifié, comprenant ainsi un nombre élevé de créances.

I.3.1.1 La probabilité de défaut conditionnelle

Selon Gordy (2003), une distinction importante est faite entre la probabilité de défaut conditionnelle et la probabilité de défaut inconditionnelle. La probabilité de défaut inconditionnelle d'une contrepartie, notée PD, correspond à sa probabilité de défaut à un certain horizon, compte

tenu de toute l'information observable au moment de son estimation. En revanche, la probabilité de défaut conditionnelle est la probabilité de défaut qu'on assignerait à cette contrepartie si on connaissait aussi quelle serait la valeur précise du/des facteur(s) de risque systématique à cet horizon. La probabilité de défaut inconditionnelle correspond à la valeur moyenne des probabilités de défaut conditionnelles obtenues en considérant toutes les réalisations possibles du/des facteur(s) de risque systématique.

La rentabilité standardisée des actifs de l'émetteur de la i -ème créance, Z^i , est exprimée en fonction d'un seul et même indice gaussien X représentant la conjoncture globale (facteur unique de risque systématique), et d'un facteur de risque spécifique diversifiable ε^i qui est également gaussien.

$$Z^i = \sqrt{1 - \rho^i} \varepsilon^i - \sqrt{\rho^i} X$$

Dans le modèle de Merton (1974), la probabilité de défaut de l'entreprise émettrice de la créance i correspond à la probabilité que la rentabilité standardisée de son actif total Z^i devienne inférieure ou égale à sa distance au défaut² DD^i , à l'horizon T (habituellement 1 an dans la pratique). L'expression de la probabilité de défaut conditionnelle à la valeur particulière de x de l'indice de la conjoncture macro-économique peut donc s'écrire :

$$PD(x)^i = \mathbb{P} \left(\sqrt{1 - \rho^i} \varepsilon^i - \sqrt{\rho^i} X \leq DD^i \mid X = x \right)$$

$$PD(x)^i = N \left(\frac{DD^i + \sqrt{\rho^i} x}{\sqrt{1 - \rho^i}} \right)$$

où N est la fonction de répartition d'une gaussienne centrée réduite. En rappelant l'expression de la distance au Défaut d'après le modèle de Merton, l'expression de la probabilité conditionnelle devient donc :

$$PD(x)^i = N \left(\frac{N^{-1}(p^i) + \sqrt{\rho^i} x}{\sqrt{1 - \rho^i}} \right)$$

I.3.1.2 La perte au niveau portefeuille

La variable aléatoire D , représentative du défaut peut s'écrire : $D^i = \mathbb{1}_{\sqrt{1 - \rho^i} \varepsilon^i - \sqrt{\rho^i} X \leq DD^i}$

La perte globale de crédit dans un portefeuille de N créances correspond à la somme de N variables $L^i * D^i$ conditionnellement à X , où $L^i = LGD^i * EAD^i$, correspond à la perte subie par la banque sur le titre de créance i en cas de défaut de son émetteur durant l'année. La perte globale de crédit du portefeuille P s'écrit alors :

$$L_P = \sum_{i=1}^N EAD^i * LGD^i * \mathbb{1}_{\sqrt{1 - \rho^i} \varepsilon^i - \sqrt{\rho^i} X \leq DD^i}$$

2. La « distance au défaut » DD représente la rentabilité standardisée de l'actif de l'entreprise au-dessous de laquelle l'entreprise fait défaut. Elle correspond au quantile d'ordre p^i de la loi normale centrée-réduite : $DD^i = N^{-1}(p^i)$ où p^i est la probabilité de défaut inconditionnelle

Le modèle de Gordy (2003) permet d'établir que pour un portefeuille bien diversifié (N grand avec aucun actif i dominant), la perte de crédit pour une unité d'exposition converge vers son espérance conditionnelle. En conséquence, la perte globale de crédit dans un portefeuille de N créances peut s'écrire comme suit :

$$L_P \approx E[L_P | (x)] = EL_P(x) = \sum_{i=1}^N EAD^i * LGD^i * PD(x)^i$$

$$L_P \approx EL_P(x) = \sum_{i=1}^N EAD^i * LGD^i * N \left(\frac{N^{-1}(p^i) + \sqrt{\rho^i x}}{\sqrt{1 - \rho^i}} \right)$$

Il s'en suit que la credit-VaR au niveau du portefeuille P peut s'écrire :

$$\text{Credit-VaR } (L^P)(q, 1an) \approx EL^P(x_q) = \sum_{i=1}^N EAD^i * LGD^i * N \left(\frac{N^{-1}(p^i) + \sqrt{\rho^i x_q}}{\sqrt{1 - \rho^i}} \right)$$

où $x_q = \text{Var}_q(X) = N^{-1}(q)$ correspond à une conjoncture macro-économique la plus défavorable au seuil de probabilité q (q fixé à 99.9% par le Comité de Bâle).

I.3.2 La fonction de calcul du capital réglementaire

Le régulateur (cadre de Bâle II et Bâle III) a proposé aux banques d'appliquer le modèle mono-factoriel asymptotiquement granulaire aux portefeuilles de crédit bien diversifiés, avec quelques modifications.

I.3.2.1 Pour le calcul de la perte attendue (EL) :

La perte attendue (EL) se calcule en prenant la probabilité de défaut inconditionnelle de l'émetteur de la créance i au lieu de sa probabilité de défaut conditionnelle $PD(x)$. Il s'agit de la perte espérée inconditionnelle.³

$$EL^P = \sum_{i=1}^N EAD^i * LGD^i * p^i$$

I.3.2.2 Pour le calcul de la perte inattendue :

Dans le pire des scénarios, la totalité d'un portefeuille de crédit (scénario peu probable mais pas impossible) pourrait être perdue en une année. Dans ce cas, les déposants ne seraient protégés que si la banque finançait 100% des crédits qu'elle octroie par des fonds propres. Cela

3. Selon Dhima & Julien (2020) [6] « la perte espérée inconditionnelle peut être obtenue approximativement à partir d'une valeur moyenne de la conjoncture qui correspond à une probabilité de la dépasser de 50% et qui donne un quantile de $\bar{X} = x_{0.5} = 0$ dans la loi normale centrée ». De même les valeur de ρ généralement en dessous de 25% avec la formule proposée par le régulateur permettent d'avoir le dénominateur proche de la valeur 1.

entraînerait un coût insupportable pour la banque et, de plus, cela serait incompatible avec son rôle de transformation (du passif de court-terme en actif de long-terme) dans l'économie. La fonction de calcul du capital réglementaire dans l'approche IRB (Internal Rating Based) développée depuis le Comité de Bâle II, appelée autrement fonction K, a été élaborée pour fixer le niveau optimal de fonds propre à trouver, en adoptant un intervalle de confiance de 99.9%. Plus précisément, les autorités de contrôle acceptent de voir un établissement subir des pertes dépassant le montant de ses fonds propres réglementaires une fois en mille ans, ce qui signifie qu'elles acceptent d'observer une faillite dans un environnement économique sur mille. L'intervalle de confiance de 99.9% correspond à la probabilité que les pertes effectivement réalisées soient inférieures ou égales aux pertes calculées à ce seuil par le modèle. Ce montant de pertes de crédit correspond à la Value at Risk de Crédit. Elle est estimée comme une espérance de perte de crédit conditionnelle à une conjoncture économique dégradée (cf.[8]). La partie de la Credit-VaR correspondant aux pertes attendues (EL) a été soustraite de la charge en capital, dans la mesure où celle-ci est déjà couverte par les dépréciations et les provisions normalement constituées par l'établissement. La charge en capital (K) est donc censée couvrir uniquement les pertes inattendues (UL). La formulation de la fonction K se base sur le modèle asymptotique mono-factoriel de Gordy-Vasicek.

La fonction K dans le cas général des expositions saines s'écrit :

$$K = \text{capital ratio} * RWA$$

$$RWA = 12.5 * \left[N \left(\frac{N^{-1}(PD) + \sqrt{R} * N^{-1}(0.999)}{\sqrt{1-R}} \right) - PD \right] * LGD * EAD * M_{Adj}$$

Quelques extensions liées à l'introduction d'une formule fixe pour calculer le coefficient ρ (fonction R), d'une maturité ajustée M_{Adj} et d'un facteur multiplicatif.

- Le calcul du coefficient R (fonction R)

Le coefficient R renseigne sur la corrélation de la rentabilité de l'actif de l'obligor au facteur de risque systématique. Selon la proposition de l'accord, la formule de calcul de ce coefficient dépend :

- du type de contrepartie (souverain, institution financière, grande entreprise, petite ou moyenne entreprise, clientèle de détail) : une entreprise est par exemple plus sensible à l'état global de l'économie qu'un particulier ;
- de la taille de l'entreprise mesurée par son chiffre d'affaires annuel S en million d'euros (M_e) : une grande entreprise ($S \geq 50M_e$) est plus sensible à l'état global de l'économie qu'une PME ($S < 50e$)
- du type de produit : on observe par exemple une corrélation historiquement beaucoup plus élevée entre les emprunteurs qui ont souscrit des prêts hypothécaires résidentiels qu'entre les porteurs de crédit renouvelable ;

- de la qualité de crédit de la contrepartie, mesurée par sa probabilité de défaut (PD) : les clients sains sont plus sensibles à l'état global de l'économie que les clients déjà au bord de la faillite,

Nous présentons ci-dessous différents cas de figure :

- pour les corporates ($S \geq 50Me$)

$$R = 0.12 * \frac{1 - \exp(-50PD)}{1 - \exp(-50)} + 0.24 \left[1 - \frac{1 - \exp(-50PD)}{1 - \exp(-50)} \right]$$

- pour les corporates ($S < 50Me$)

$$R = 0.12 * \frac{1 - \exp(-50PD)}{1 - \exp(-50)} + 0.24 \left[1 - \frac{1 - \exp(-50PD)}{1 - \exp(-50)} \right] - 0.04 * \left[1 - \frac{\min\{\max\{5, S\}, 50\} - 5}{45} \right]$$

- pour la clientèle retail

$$R = 0.03 * \frac{1 - \exp(-35PD)}{1 - \exp(-35)} + 0.16 \left[1 - \frac{1 - \exp(-35PD)}{1 - \exp(-35)} \right]$$

- La maturité ajustée M_{Adj}

Cet ajustement M_{Adj} vaut 1 pour la clientèle retail alors qu'il est fonction de la PD pour les corporates :

$$M_{Adj} = \frac{(1 + (M - 2.5) * b)}{(1 - 1.5 * b)}, \text{ avec } b = (0.11852 - 0.05478 * \ln(PD))^2$$

L'ajustement de maturité s'explique par le fait que les modèles de risque de crédit supposent que les crédits sont détenus jusqu'à échéance. Les pertes résultent alors exclusivement des défauts constatés à l'horizon du modèle. Or les modèles supposent un horizon d'un an alors que la maturité moyenne des crédits est généralement plus longue. Il faut alors prendre en compte le fait qu'un crédit de maturité supérieure à un an a une PD à moyen terme supérieure à sa PD à un an. Ainsi, par exemple, un crédit peu risqué peut être dégradé dans l'intervalle sans être en défaut. Cette dégradation se traduit par une augmentation effective de sa PD.

- Un facteur multiplicatif :

Sur la base d'études quantitatives d'impact (Quantitative Impact Studies : QIS) réalisées auprès des banques du monde entier, le Comité de Bâle a estimé que la charge en capital telle que déterminée par la fonction K était insuffisante pour atteindre l'objectif de maintien du niveau global de fonds propres du secteur bancaire. Il a ainsi décidé d'introduire un facteur multiplicatif (scaling factor), fixé à 1.06.

Chapitre II

Présentation du portefeuille de financements structurés

L'objet de ce chapitre est d'introduire les types de financement structurés concernés par cette étude. Nous expliquerons les particularités de ces financements et présenterons également les caractéristiques de l'échantillon d'étude .

II.1 Définition

Les financements structurés regroupent les opérations pour lesquelles une structure est mise en place, afin de réduire les risques pour la banque ou de répondre à des besoins spécifiques du client. De la diversité des risques (juridiques, financiers, opérationnels) et des besoins clients naissent la variété des structures possibles. Les besoins clients nécessitant d'adapter le financement sont peuvent être :

- Le financement d'actif qui peut concerner le financement des exportations, le financement de l'aviation, le financement du transport maritime et de l'offshore, le financement de l'immobilier. Dans cette catégorie on peut également inclure les opérations de leasing sur les financements suscités. Le leasing ou crédit-bail permet à une entreprise de lever des fonds autrement que par un prêt classique. Il permet à l'entreprise (preneur) de louer un bien en payant un loyer au bailleur pour une durée déterminée. À une date ultérieure, le preneur peut acheter le bien.
- Les financements pour l'exploitation de ressources naturelles et la construction d'infrastructures qui s'appuient sur une expertise sectorielle spécifique en matière de financement des infrastructures, des projets énergétiques, des mines, des métaux et des industries et des matières premières, ainsi que sur une connaissance approfondie des marchés dans les pays développés et émergents ;
- Le financement d'investissement qui regroupe l'ensemble des financements consistant à apporter des ressources financières aux entreprises (le financement des opérations de fu-

sions et acquisition, equity capital markets, debt capital markets, financement d'acquisition stratégique, leverage buyout, ...)

II.2 L'échantillon de financements étudiés

L'échantillon d'étude est construit à partir des données obtenues sur 7915 deals enregistrés à l'étape d'origination. Ces deals ont tous une date de signature espérée entre 2016 et 2021. Au terme du processus d'approbation, un deal peut être signé, abandonné ou perdu du fait du retrait du client pour diverses raisons.

| Statut d'approbation | Fréquence |
|----------------------|-----------|
| Approved | 0.36 |
| Lost | 0.43 |
| Abandon | 0.21 |

TABLE II.1 – Distribution de la variable d'intérêt

Sur l'échantillon considéré, il est à noter que bon nombre de deals sont perdus au cours du processus de preclosing. Cette situation fait référence à la disponibilité d'options de financements alternatives selon l'expertise de l'emprunteur/des sponsors et les conditions du marché. Le manque d'informations sur les réelles conditions du marché (la situation de l'offre des autres institutions financière et de la demande) a constitué un frein à la compréhension de certaines observations réalisées sur les deals perdus.

Avec un taux d'abandon de 21% et un taux d'approbation de 36%, il convient de noter qu'une annonce de deal a, à priori, un peu plus d'une chance sur 3 de faire consensus à toutes les étapes du processus d'approbation.

II.3 Caractéristiques en terme de réduction des risques

Dans les opérations de financement structuré, plusieurs techniques de réduction ou de transfert de risques sont employées et participent à la structuration des opérations. Il peut être question de procéder par syndication du crédit pour associer d'autres banques, par titrisation pour associer des investisseurs professionnels, par l'utilisation d'une garantie ou par la mise à disposition de collatéral. Ces stratégies visent à créer des mécanismes spécifiques pour atténuer les pertes potentielles.

II.3.1 L'utilisation des garanties et/ou de collatéral

L'utilisation de garanties permet d'améliorer la notation de crédit des opérations de financement structuré. La mise à disposition de collatéral par les clients permet de limiter la perte

potentielle liée au défaut. Les garanties et/ou collatéraux peuvent être soit déterminées en interne dans la structure de la transaction, soit fournies en externe par un tiers. Dans le cadre de cette étude, 18.5% des deals sont couverts par une garantie.

II.3.2 L'utilisation d'un véhicule de financement

Les véhicules de financement appelés SPV (Special Purpose Vehicle) permettent aux grandes entreprises d'atteindre des objectifs spécifiques en obtenant des financements pour effectuer des activités spéciales d'investissement. Les investisseurs accordent donc un financement au véhicule et non aux parties (sponsors ou initiateurs) qui ont fondé cette société. Puisque l'actif à financer est pris en charge par une entité juridique créée dans ce but spécifique, toutes les conséquences économiques générées par l'investissement en question sont attribuées à cette entité. Compte tenu du fait que ce véhicule a sa propre valeur nette, les actifs qui servent à gérer le projet sont séparés des autres actifs des parties qui ont créé le véhicule. Ainsi, les actifs du SPV peuvent devenir une garantie pour les créanciers. Ces techniques impliquent que le remboursement du prêt est garanti principalement par la génération de liquidités par les actifs financés.

II.3.3 La syndication

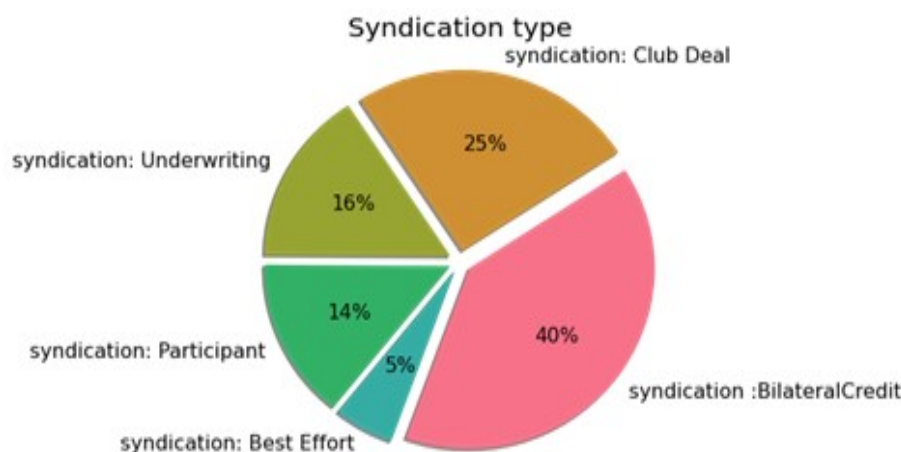


FIGURE II.1 – Répartition des deals étudiés par type de syndication

Le syndicat bancaire peut être défini comme la réunion de deux ou plusieurs banques, ayant pour objet la répartition de la charge, de la rémunération et du risque d'un crédit octroyé à un emprunteur. Les crédits syndiqués sont structurés, arrangés et administrés par une ou plusieurs banques - commerciales ou d'investissement - connues sous le nom d'arrangeurs. On distingue pour cette étude 4 formes d'implication de la banque dans une syndication :

- Underwritten deal (ou syndication en underwriting) : les arrangeurs s'engagent d'emblée sur la totalité du montant de l'opération, et procèdent ensuite à la syndication. S'ils ne

peuvent pas trouver d'autres prêteurs, ils sont contraints d'absorber la différence, qu'ils pourront par la suite essayer de vendre à des investisseurs. Ces deals représentent 17.3% de l'échantillon étudié.

- Syndication en Best Efforts : le montant sur lequel s'engage le groupe des arrangeurs est inférieur au montant total de l'opération. Le solde est laissé à l'appréciation du marché. Si le crédit n'est pas assez souscrit, le dossier peut ne pas aboutir, ou nécessiter une adaptation significative pour intéresser le marché. Les deals syndiqués en best effort représentent 6% de notre échantillon.
- Club deal : un club deal est un crédit d'un montant plus qui est pré-vendu à un groupe de prêteurs. L'arrangeur est généralement "primus inter pares" (premier parmi ses pairs), et chaque prêteur reçoit une part identique (ou quasi-identique) des commissions, en proportion de son engagement. 27.5% des deals considérés pour cette étude sont syndiqué de cette manière.
- Participant : Au cours de la syndication les prêteurs signataires du contrat de crédit peuvent faire appel à d'autres prêteurs non-signataires du contrat afin d'atteindre leur part finale dans le deal en cédant une partie de leurs engagements. Dans 16% des deals considérés pour cette étude, la banque est impliquée dans la syndication en tant que participant.

La syndication n'étant pas incontournable dans les financements structurés, notre échantillon d'étude est constitué de 33.7% de financements réalisés en bilatéral avec la contrepartie.

II.4 Typologie de financement

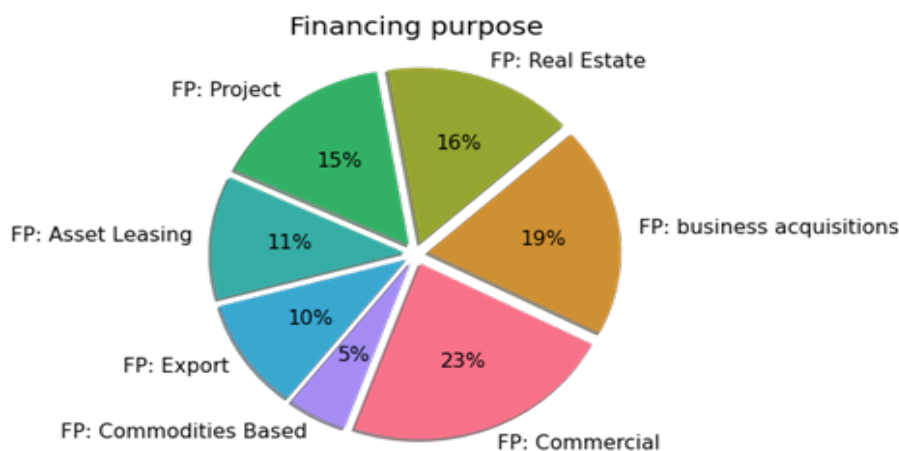


FIGURE II.2 – Répartition des deals selon l'objectif de financement

Dans cette étude nous considérons une typologie en lien avec les objectifs de financement et permettant de regrouper les différents financements accordés en 7 groupes :

- Les financements de projet :

Le financement de projet est un crédit sans recours accordé à une entité ad hoc, qui doit rembourser ses échéances grâce aux flux de trésorerie générés par le projet. Les activités concernées sont les suivantes : infrastructures, électricité, gaz-pétrole, énergies nouvelles. L'échantillon de deal concerné par cette étude est constitué à 15% de financements de ce type.
- Le financement d'actif immobilier :

Cet objet de financement couvre les expositions sur des contreparties propriétaires de biens immobiliers. Ces expositions sont généralement créées spécifiquement pour acheter un actif ou un groupe d'actifs. Elles sont sans recours pour les propriétaires/investisseurs de l'entité, de sorte que le remboursement provient uniquement des flux de trésorerie et/ou du produit de la vente des actifs. L'échantillon de deal concerné par cette étude est constitué à 16% de financements de ce type.
- Le financement d'autres actifs :

Cet objet de financement couvre les expositions spécifiquement structurées pour financer l'acquisition ou de leasing d'avions, de bateaux ou d'autres actifs. Dans la plupart des cas, l'actif appartient à une entité ad hoc et est loué soit directement par une compagnie aérienne, soit par une société de leasing d'avions, qui le sous-loue à une compagnie aérienne. La transaction peut porter sur le financement d'un seul avion exploité par une seule compagnie aérienne ou sur un portefeuille entier d'avions exploités par plusieurs compagnies aériennes. Les transactions sont basées sur des droits de propriété/sûreté solides, ou des droits économiquement équivalents, et sont remboursées par les revenus de la location ou potentiellement par la vente ou la relocation de l'avion. 11% des deals considérés pour cette étude sont de ce type.
- Les financements d'exportation :

Le financement des exportations concerne le financement des exportations de biens industriels et de services. Le financement des pré-exportations est conçu pour les transactions liées au commerce (à court ou à long terme) conclues avec des producteurs de matières premières (ou des négociants, sous certaines conditions). Il est principalement structuré par un préfinancement direct ou des prépaiements. 10% des deals considérés pour cette étude sont financés dans cet objectif.
- Les financements d'exploitation de ressources naturelles :

Le financement de matières premières est utilisé pour l'acquisition de matériaux d'origine naturelle, soumis à une transformation et à une utilisation économique. Cet objectif de financement regroupe les expositions (engagements financés ou hors bilan tels que les lettres de crédit) accordées aux entreprises de négoce ou de transformation de matières premières. Chaque exposition est garantie par une quantité spécifique de produits de base appartenant à la société de négoce, telle qu'une cargaison de pétrole brut en transit entre

les pays exportateurs et importateurs. 5% de l'échantillon de deals considérés sont financés pour cet objectif.

— Les financements d'investissement :

Cet objectif de financement couvre les acquisitions de business via des opérations de LBO (Leveraged Buy Out)¹ et via des opérations de non-LBO (crédit remboursé par la trésorerie du groupe acquéreur ou par un appel au marché des capitaux). Les financements d'investissement sont présents à 19% dans notre échantillon d'étude.

— Les financements pour des objectifs commerciaux :

Ces financements concernent essentiellement une partie des opérations de financement simplement accordées pour des raisons commerciales, et considérées comme un " investissement " pour la banque, lui permettant de développer sa relation commerciale avec ses clients. Ces financements, sont présents dans notre échantillon d'étude à hauteur de 23%.

1. rachat avec effet de levier

Chapitre III

Présentation de la métrique de rentabilité

Dans ce chapitre, il sera question de présenter et d'illustrer les caractéristiques de la métrique RARORC, les indicateurs intermédiaires ainsi que les étapes de calcul. Les deux dernières sections du chapitre seront consacrées à l'évaluation de cette métrique dans le cadre de l'octroi des financements structurés.

III.1 Expression des indicateurs RARO(R)C

Les banques ont investi dans des outils fournissant des mesures de performances dans une logique de gestion optimale du risque de crédit. C'est le rôle des métriques de type RAPM (Risk-Adjusted Performance Metrics). Durant les années 1990, une métrique de rentabilité ajustée au risque appelée RAROC, pouvant être calculée à l'échelle des transactions et des Business line a été proposée et largement adoptée par l'industrie bancaire. En substance, le RAROC (Risk-Adjusted Return on Capital) est défini comme le rapport de la somme des revenus nets ajustés des pertes attendues au capital nécessaires pour absorber les pertes inattendues.

En pratique, le schéma de calcul du RAROC dépend de la définition particulière du capital adopté. Il est donc observé plusieurs variantes de cet indicateur basées sur le capital économique ou le capital réglementaire. Deux définitions du capital s'appliquent alors au calcul du RARORC. Le capital réglementaire qui est défini par les règles de surveillance du Comité de Bâle et le capital économique qui, mesuré par un modèle de portefeuille de crédit, est supposé refléter la réalité économique plus fidèlement que les règles réglementaires, par exemple, en quantifiant les risques de concentration. Généralement, une mesure du risque est dérivée de la distribution simulée des pertes d'un portefeuille de crédit. La VAR¹ ou l'expected Shortfall² est utilisée

1. Pour rappel la VaR_α est le quantile d'ordre $\alpha \in]0, 1[$ de la distribution des pertes

2. L'Expected Shortfall ES_α est l'espérance de la distribution de pertes au-dessus du quantile d'ordre $\alpha \in]0, 1[$

$$ES_\alpha(L) = E[L | L \geq VaR_\alpha(L)] = \frac{1}{1-\alpha} \int_{VaR_\alpha(L)}^{\infty} l dP(l)$$

comme mesure de risque et la valeur obtenue est ensuite répartie entre les différents crédits dans un processus appelé allocation du capital [11]. Le capital économique correspond ainsi au montant des fonds propres économiques que l'établissement estime nécessaire pour couvrir ses risques.

Dans la mesure de la performance des investissements, le capital réglementaire est préféré puisque les règles réglementaires sont devenues bien plus conservatrices.

Dans cette étude, le RARORC (Risk-Adjusted Return On Regulatory Capital) est adoptée par la banque.

$$\text{RARORC} = \frac{\text{Cumulative Expected NBI} - \text{Cumulative expected loss}}{\text{capital ratio} * \text{Cumulative RWA}}$$

Cette approche rend le calcul de la métrique plus facile à mettre en oeuvre tout en conservant certaines propriétés intéressantes. Calculé au niveau de chaque facilité, le rarorc à un niveau deal s'obtiendra simplement par une moyenne pondérée par le capital réglementaire.

III.2 Les indicateurs intermédiaires

Dans cette sous section, DF_t est le taux d'actualisation appliqué à la fin de la période t . Les éléments de modélisation du risque de crédit que sont la probabilité de défaut, la matrice de transition, la perte en cas de défaut, l'exposition en cas de défaut ainsi que le RWA (Risk weighted Asset) sont présentés dans le premier chapitre (I.3.2.2) de ce document.

III.2.1 Le NBI cumulé

Le NBI est composé des revenus nets de la transaction ajustés des coûts de liquidité. Il est calculé à chaque période et probabilisé (Expected NBI).

$$\text{Cumulative Expected NBI} = \sum_{t=0}^{T-1} v_t * NBI_t * DF_t$$

où v_t représente la probabilité de survie de la contrepartie (ou du sponsor) entre l'instant initial et l'instant t d'encaissement de ces flux et

$$\text{NBI} = \text{fees} + \text{gross margin} - \text{funding costs}$$

Le coût de liquidité (ou coût de funding, spread de liquidité, spread de funding, funding cost) désigne le coût d'emprunt de la liquidité de la banque sur les marchés. Il s'agit des coûts internes de la banque facturés aux lignes d'activité par les services de trésorerie afin d'obtenir des fonds sur le marché au meilleur prix, conformément aux caractéristiques des transactions (devise, maturité initiale, du profil d'amortissement, ...)

III.2.2 L'Expected Loss (EL) cumulée

La perte espérée est calculée pour chaque période entre la date de début et la date d'échéance d'une transaction. L'EL réglementaire serait soumise à des transitions de notation. L'Expected Loss cumulée pour la période est obtenue par actualisation des pertes espérées incrémentielles cumulées :

$$\text{Cumulative expected loss} = \sum_{t=0}^{T-1} Pd_{t,t+1} * EAD_t * LGD_t * DF_t$$

$Pd_{t,t+1}$, est la probabilité de défaut du client sur un an ou du sponsor et EAD_t l'exposition de la banque en cas de défaut.

III.2.3 Le RWA cumulé

Tout au long de la vie d'une transaction et sur une base annuelle, il est calculé une consommation de capital pour chaque période. Les valeurs sont ensuite actualisées et additionnées. Le RWA cumulé de la transaction peut être exprimé comme suit :

$$\text{CumulativeRWA} = \sum_{t=0}^{T-1} \text{SpotRWA}_t * DF_t$$

La notion de SpotRWA_t (le RWA est présenté plus en détail dans le chapitre précédent) fait référence au fait que la probabilité de défaut d'un an initial est gardée constante dans le temps. Le calcul du RWA cumulé, en gardant la probabilité de défaut constante sur la vie du deal (pour des raisons de simplicité), pourrait poser la problématique de la non prise en compte du risque de migration de rating.

III.3 Distribution du RARORC par type de financement

Le graphique III.1 présente les valeurs obtenues pour le RARORC par type de métier pour les deals approuvés et les deals non approuvés (axe de gauche) ainsi que le taux d'approbation des deals selon l'objectif de financement. De ce graphique il ressort que :

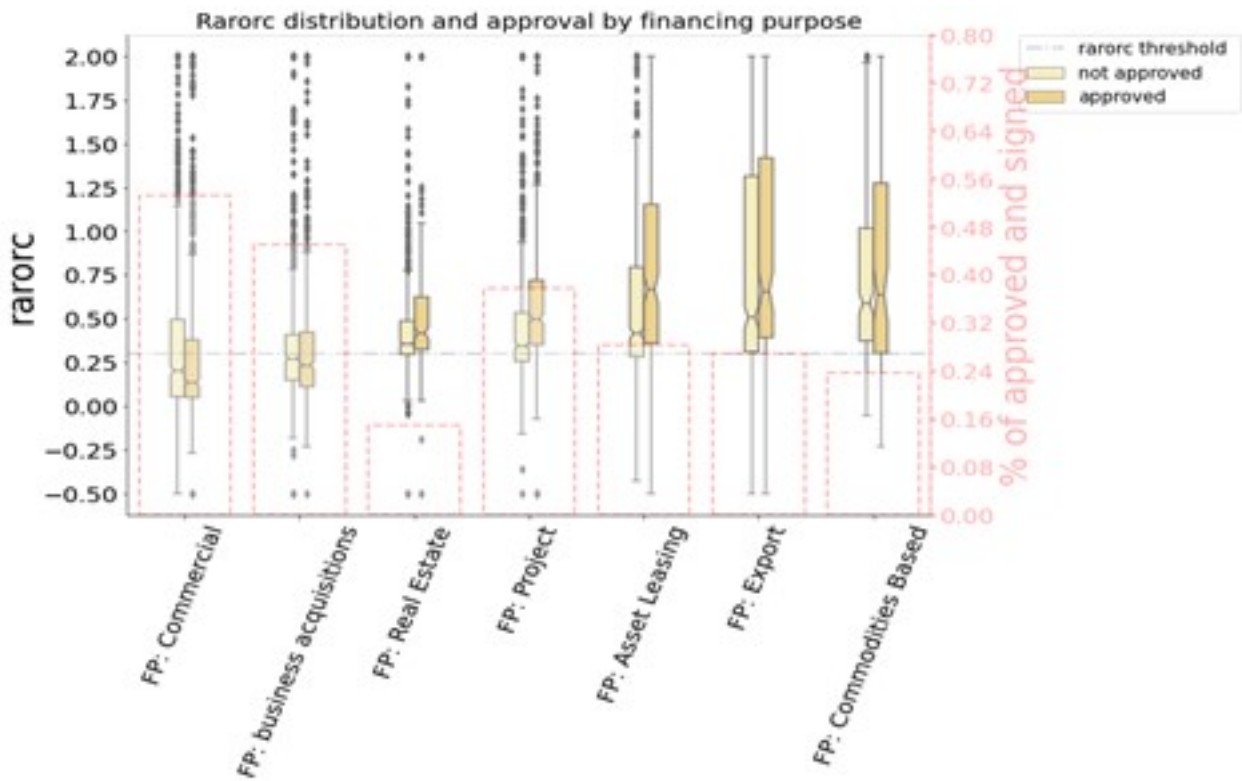


FIGURE III.1 – RARORC et approbation des deals par objectifs de financement

Aide à l'interprétation : Le graphique présente avec l'axe de gauche la distribution de rarorc obtenus sur les deals pour chaque objectif de financement. Avec l'axe de droite les barplot permettent d'observer le taux d'approbation de ces deals pour chaque objectif de financement.

- La distribution du RARORC diffère d'un type de financement à un autre. On observe une distribution plus dispersée pour les financements d'actifs, d'exportations et surtout pour des financements basés sur des transformations de matières premières. Cela peut se justifier par l'utilisation des techniques de réduction de risque (telles que les divers types de garanties) qui permettent sur ces financements de réduire considérablement la LGD et donc le capital réglementaire. En conséquence, le RARORC atteint assez souvent sur ces financements des valeurs très fortes.
- Le niveau médian des deals proposés diffère aussi selon le type de financement. Il faudra rappeler que le NBI proposé par la banque pour un financement donné tiendra compte premièrement de la rémunération proposée sur le marché. Lorsque la concurrence est importante sur un type de financement, on peut s'attendre à marge d'intérêt réduit. De même, la banque peut être amenée à proposer des offres moins rentables mais dans un but purement commercial (fidélisation d'un client, gestion d'une position stratégique sur un secteur). Cette réalité explique le faible niveau de rentabilité observé pour les financements dans un objectif commercial.
- En comparant les deals approuvés à ceux non approuvés, il en ressort que sur la plupart des

types de financement, les deals signés ont une meilleure médiane de RARORC. Cependant la distribution des deals signés n'est pas si différente de ceux non signés. Cette situation pourrait s'expliquer par la perte d'un certain nombre de deal (presque autant de deals signés que perdu sur l'ensemble du périmètre). En effet le lot de deal perdu peut être constitué de deals aussi rentables que ceux signés.

- L'analyse des taux d'approbation de deals par type de financement permet de noter que ce taux n'est pas constant d'un type de financement à un autre et les sous périmètres présentant les meilleures rentabilités n'ont pas forcément les meilleurs taux d'approbation. Plusieurs facteurs peuvent justifier ce constat. En premier lieu, la pression commerciale est différente d'un type de financement à un autre et peut conduire à un certain nombre de dérogation. L'exemple le plus évident est celui des financements réalisés dans un objectif dit commercial : la médiane de rarorc est autour de 10\% et le taux d'approbation est le plus important. En deuxième lieu, il faut noter que la gestion d'autres sources de risques (et non uniquement la rentabilité ajustée du risque de crédit) peut être déterminant sur certains types de financement. Le risque opérationnel, le risque de concentration, le risque pays, l'impact environnemental, la limite de capital alloué à la ligne de métier, la préférence pour certaines formes de garantie sont autant d'éventuels facteurs potentiellement explicatifs que le rarorc n'intègre pas. En troisième lieu, il convient de noter que le rarorc est calculé sur la maturité effective du deal. Cette hypothèse devient moins vraisemblable lorsque certains types de financement peuvent passer du marché primaire vers le marché secondaire. Sur des transactions d'investissement au capital, on pourrait analyser le rarorc en tenant compte d'une possible surestimation de la maturité effective du financement dans le portefeuille. Des niveaux plus bas de rarorc auront donc du sens sur les sous périmètres concernés.
- Le graphique III.1 ci-dessus permet aussi de comparer les distributions de rarorc et le seuil de rentabilité minimum pour l'approbation des deals. En ce qui concerne le pilotage de la rentabilité au niveau des lignes de métiers, on note que le seuil minimum fixé est en phase avec la rentabilité médiane (et moyenne) sur l'ensemble des types de financement à l'exception des financements pour un objectif commercial ou pour un objectif d'investissement au capital. Par contre au niveau deal, on peut noter qu'un pourcentage important de deal est approuvé avec un rarorc inférieur au seuil quelque soit l'objectif de financement considéré. Cette remarque conduit à analyser le pouvoir discriminant de la métrique de rentabilité dans le processus d'octroi des financements.

III.4 Distribution du rarorc par type de syndication

Le graphique III.9 ci-dessous présente la distribution du rarorc et le taux d'approbation selon le type de syndication. Les deals sur lesquels l'engagement de l'institution financière nécessite un rôle actif dans la syndication et les deals bilatéraux sont les plus rentables. Un faible taux

d'approbation est aussi noté sur ces groupes de transaction.

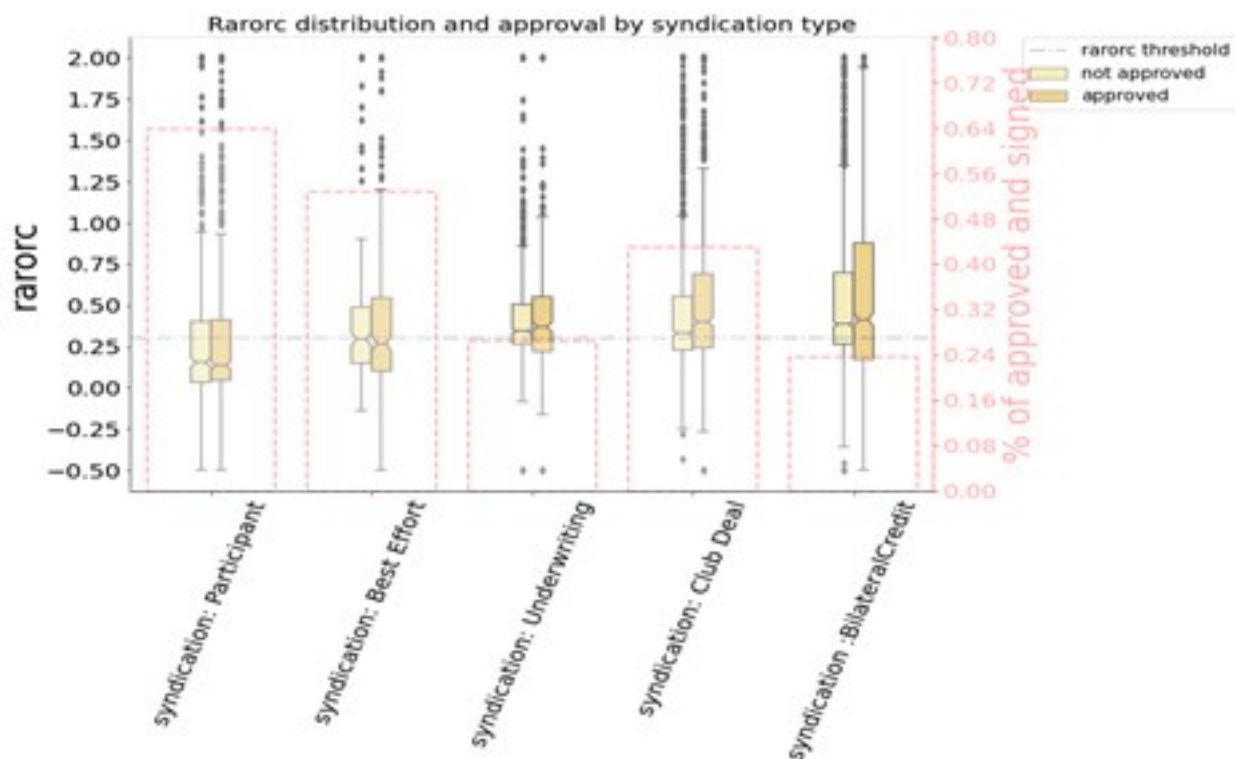


FIGURE III.2 – Rarorc et approbation des deals par type de syndication

Aide à l'interprétation : Le graphique présente avec l'axe de gauche la distribution de rarorc obtenus sur les deals selon le type de syndication. Avec l'axe de droite les barplot permettent d'observer le taux d'approbation de ces deals selon le type de syndication.

La syndication offre aux prêteurs syndiqués la possibilité de bénéficier de la connaissance de l'emprunteur et de l'expertise des banques du pool en matière de prêt sur un marché donné. Par exemple les prêts syndiqués permettent aux banques étrangères de s'associer à des banques nationales et peuvent donc offrir de meilleures opportunités à ces banques étrangères. Aussi lorsqu'une première banque propose la syndication, elle a probablement analysé les risques et approuvé le deal de son côté. Pour toutes ces raisons, il est naturel d'obtenir un meilleur taux d'approbation sur deal pour lesquels le rôle de la banque est moins active. Elle partage la communication financière du pool sur le deal mais aussi la pré-sélection de l'arrangeur.

En termes de tarification, une autre distinction importante peut être faite entre les prêts bilatéraux et les autres sources d'endettement. Les revenus sont plus importants car l'institution financière peut bénéficier des marges pour éventuellement couvrir un risque d'anti-sélection ou d'économie d'information de la part de du client.

III.5 Pouvoir discriminant du RARORC sur le stock de financements étudiés

Pour évaluer le pouvoir discriminant du RARORC sur l'octroi, une première approche consiste à utiliser les outils statistiques associés à l'analyse des courbes ROC ou de la courbe CAP. Ces outils nous permettent d'analyser si le signal envoyé par la rentabilité ajustée au risque de crédit constitue un élément discriminant les financements approuvés de ceux qui n'ont pas passés cette étape d'approbation. Nous présenterons dans un premier temps les outils statistiques suscités et dans un deuxième temps nous analysons les résultats.

III.5.1 Les courbes ROC et de Lorentz

III.5.1.1 La courbe ROC

Nous partons de l'hypothèse qu'on dispose de deux groupes de financements (ceux qui sont approuvés après la phase d'étude et ceux qui ne le sont pas) sur lesquelles nous cherchons à prendre une décision dès la connaissance de la rentabilité ajustée au risque. Quatre évènements sont ainsi possibles une fois la décision d'approbation est effectivement prise (l'approbation ou non selon que le RARORC est supérieur ou non à un certain seuil). En raisonnant en termes de probabilités conditionnelles on peut noter $Y \in \{0; 1\}$ l'approbation ou non et R le Rarorc sur un financement. Le seuil idéal est celui qui permet d'approuver un deal dès lors que le rarorc est au dessus de ce seuil et de ne point l'approuver dans le cas contraire. Autrement dit, ce seuil permettra de réaliser $\mathbb{P}(R \geq s|Y = 1) = 1$ et $\mathbb{P}(R < s|Y = 0) = 0$. En notant ces deux probabilités respectivement Se et Sp et $\Phi_{Y=i}$ la distribution conditionnelle de R , il peut être remarqué que : $Se = 1 - \Phi_{Y=1}(s)$ donc $Se = 1 - \Phi_{Y=1}(\Phi_{Y=0}^{-1}(Sp))$. De ce constat, lorsque $Sp = 1$ on obtient $Se = 0$ et pour $Sp = 0$ on a $Se = 1$. Il paraît donc impossible de maximiser l'un avec un seuil s sans réduire l'autre.

Ce contexte correspond en fait à une problématique abordée dans le domaine de la théorie de la détection du signal. Celle-ci est posée comme un problème de décision : lors d'une expérience, deux évènements sont possibles mais un seul est effectivement réalisé à chaque fois. Dans tous les cas, les évènements sont inaccessibles à l'opérateur. Il souhaite cependant prendre une décision en faveur de la réalisation de l'un ou l'autre de ces évènements. On suppose qu'il a tout de même la possibilité de relever indirectement des informations par l'intermédiaire d'une variable R , à valeurs dans un espace continu et unidimensionnel, dans l'une des conditions suivantes :

- soit l'opérateur a un accès total à R et les données qu'il a recueillies contiennent ainsi toute l'information sur R .
- soit l'opérateur n'a qu'un accès partiel à R . Dans ce cas, les données qu'il a relevées ne contiennent pas toute l'information sur R .

Quel que soit le cas de figure, l'objectif pour l'opérateur est de parvenir à évaluer les capacités intrinsèques de la variable R à dissocier les deux évènements. Les courbes ROC constituent un premier outil intéressant pour visualiser les forces et faiblesses de la variable R utilisée à cette fin.

L'idée sur laquelle repose la courbe ROC est de représenter dans un même repère $1 - Sp$ et Se en modifiant la valeur du seuil de décision. Construire la courbe ROC revient donc à représenter dans un même graphique les points $\{(1 - Sp, Se)\}_{s \in \mathcal{S}}$ où \mathcal{S} représente l'ensemble des seuils. La Figure III.4 présente un exemple de courbe ROC.

Muni de ces définitions, nous pouvons préciser certaines propriétés des courbes ROC :

- Propriété 1 : Les points (0,0) et (1,1) appartiennent toujours à la courbe ROC.
- Propriété 2 : Le point $(1 - Sp, Se) = (0,1)$ correspond à la situation idéale.
- Propriété 3 : La courbe ROC est inaltérée si R subit une transformation monotone.
- Propriété 4 : La pente de la courbe ROC correspond au rapport de vraisemblance évalué au seuil correspondant.
- Propriété 5 : L'aire sous la courbe ROC (AUC) est égale à $\mathbb{P}(R_{|Y=1} > R_{|Y=0})$

III.5.1.2 La courbe CAP

Un autre outil d'évaluation du pouvoir discriminant d'un indicateur ou score est la courbe CAP (Cumulative accuracy profile) aussi connu comme une courbe de Lorentz. Cette courbe permet de représenter dans un même repère les probabilités $p_x = \mathbb{P}(R \geq s)$ et $Se = \mathbb{P}(R \geq s | Y = 1)$ en modifiant la valeur du seuil de décision. Construire cette courbe revient donc à représenter dans un même graphique les points $\{(p_x, Se)\}_{s \in \mathcal{S}}$ où \mathcal{S} représente l'ensemble des seuils. Dans le cas idéal la courbe CAP augmentera linéairement et reste ensuite à un puisque tous les deals approuvés auront un excellent rarorc. Dans le cas où le rarorc n'a aucun pouvoir discriminant la fraction p_x de tous les deals ayant un bon rarorc contiendra la même fraction p_x de deal approuvé. Ces deux cas constituent les deux cas extrêmes entre lesquelles se trouve la courbe CAP en pratique.

Le pouvoir discriminant de l'indicateur peut être résumé par la statistique de l'Accuracy Ratio (AR). Il est défini comme le rapport de l'aire de la zone entre la courbe CAP et la CAP du modèle aléatoire, à l'aire de la zone entre le CAP du modèle aléatoire et du modèle idéal. L'accuracy Ratio est relié à l'AUC [7] par la relation $AR = 2 * AUC - 1$.

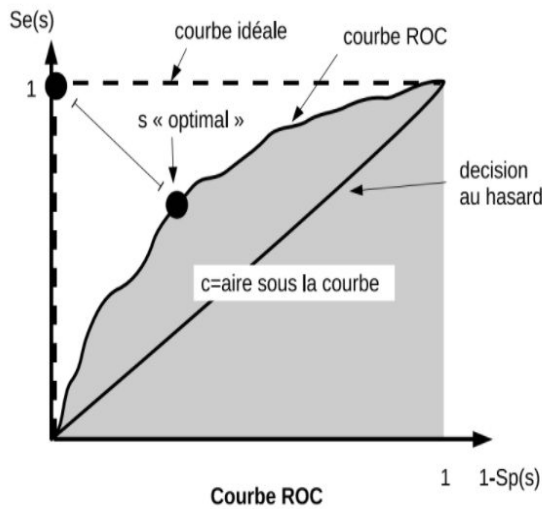


FIGURE III.3 – Exemple de courbe ROC

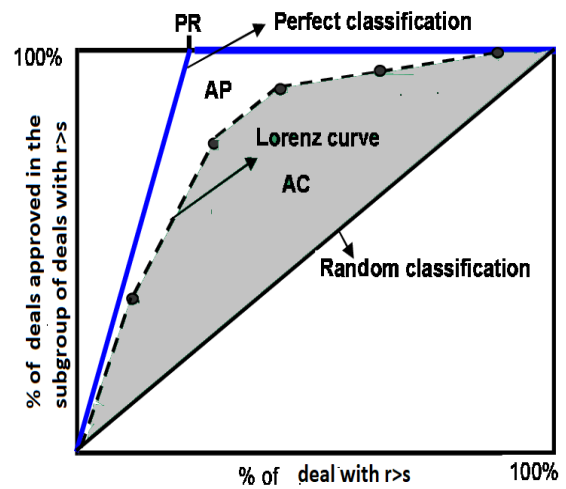


FIGURE III.4 – Exemple de courbe CAP

III.5.1.3 Estimation des paramètres de ces courbes

Dans le cadre de cette étude l'estimation de cette courbe se fera par une approche non paramétrique sur l'ensemble des données et par objectifs de financement. Ce choix est en lien avec les constats de la section précédente sur le fait qu'il est plus pertinent d'avoir différents seuils de RARORC pour différents objectifs de financement. L'estimation non paramétrique de ces courbes ROC repose sur les fonction de répartition empirique.

Le tableau III.1 permet de résumer les informations nécessaires pour chaque seuil s lors de la construction de ces courbes.

| Approbation | Rarorc | | Sensibilité (Se)- Spécificité (Sp) |
|--------------------|--------------------------|---------------------|---------------------------------------|
| | $r \geq s$ | $r < s$ | |
| $Y = 1$ | True Positive (TP) | False Negative (FN) | $Se = \frac{TP}{TP+FN}$ |
| $Y = 0$ | False Positive (FP) | True Negative (TN) | $Sp = \frac{TN}{TN+FP}$ |
| precision(p_x) | $p_x = \frac{TP}{TP+FP}$ | | |

TABLE III.1 – Matrice de confusion

III.5.2 Analyse des résultats

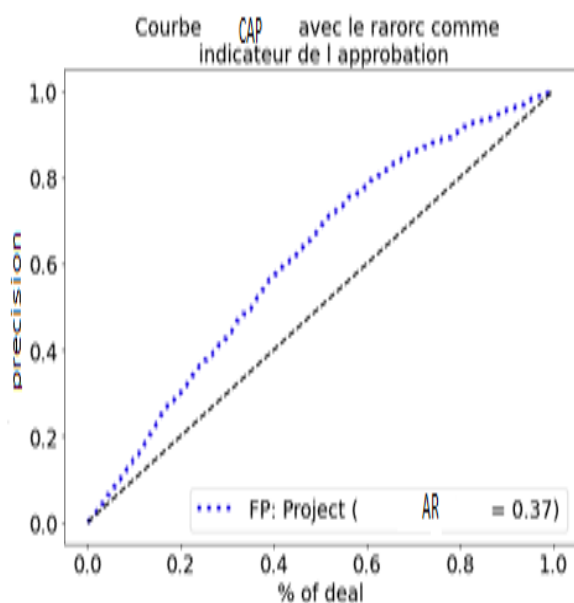


FIGURE III.5 – Courbe de lorentz sur les financements de de projets

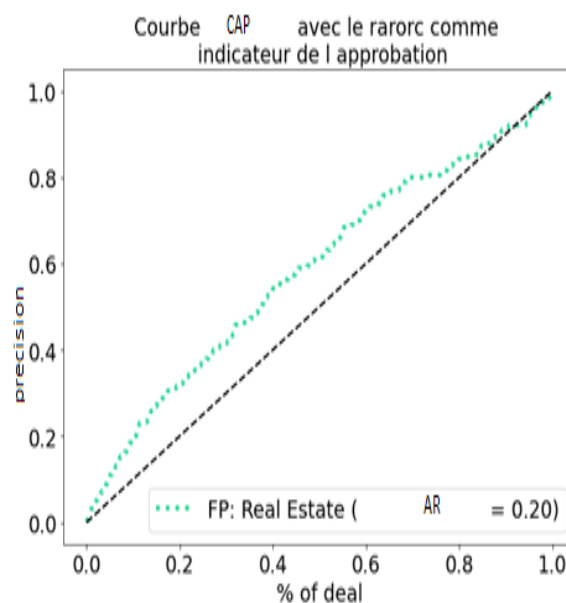


FIGURE III.6 – Courbe sur les financements en lien avec l'immobilier

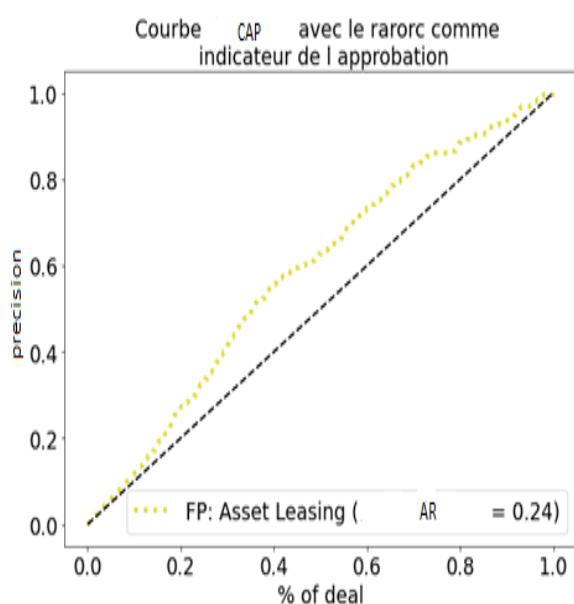


FIGURE III.7 – Courbe sur les financements d'actifs

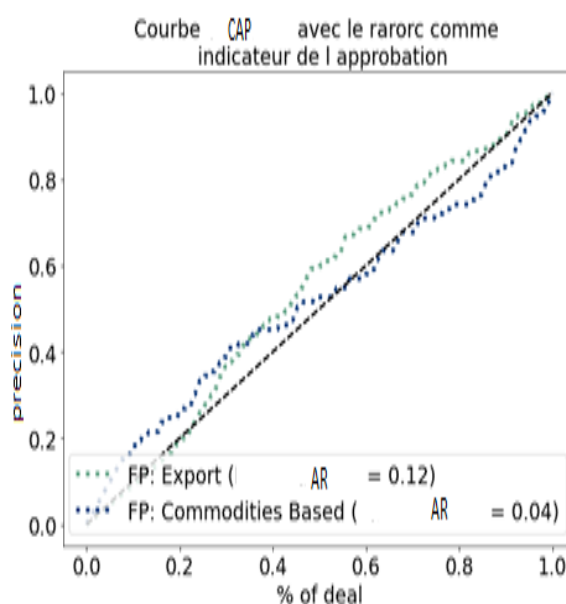
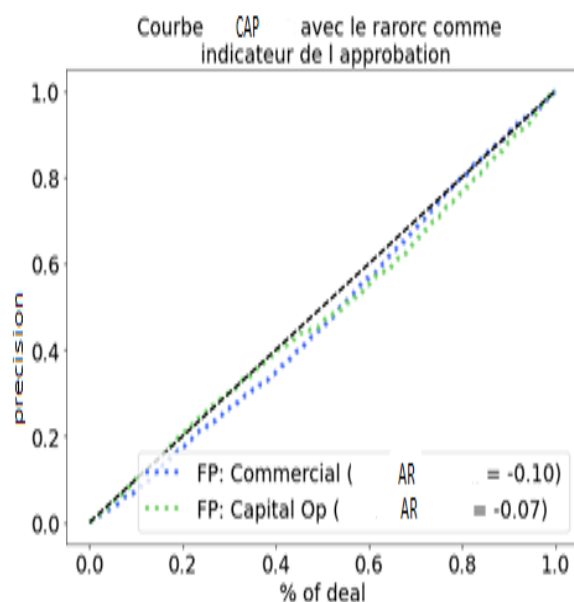


FIGURE III.8 – Courbe sur les financements d'exploitation de matières premières et d'exportation



| Financing purpose | AUC |
|------------------------|------|
| FP : Capital Op | 0.0 |
| FP : Commercial | 0.0 |
| FP : Commodities Based | 0.52 |
| FP : Export | 0.56 |
| FP : Real Estate | 0.60 |
| FP : Asset Leasing | 0.62 |
| FP : Project | 0.70 |

TABLE III.2 – AR obtenu avec le rarorc

FIGURE III.9 – Courbe sur les financements d'opérations d'investissement en capital

| | seuil unique | accuracy | sensibilité | specificité |
|------------------------|--------------|----------|-------------|-------------|
| FP : Commodities Based | 0.3 | 0.29 | 0.75 | 0.14 |
| FP : Real Estate | 0.3 | 0.36 | 0.80 | 0.28 |
| FP : Export | 0.3 | 0.38 | 0.84 | 0.21 |
| FP : Commercial | 0.3 | 0.44 | 0.30 | 0.60 |
| FP : Asset Leasing | 0.3 | 0.46 | 0.86 | 0.31 |
| FP : Capital Op | 0.3 | 0.50 | 0.41 | 0.57 |
| FP : Project | 0.3 | 0.58 | 0.86 | 0.41 |

TABLE III.3 – Score de classification avec le seuil unique fixé sur l'ensemble du périmètre

| | Score différencié | accuracy | sensitivité | specificité |
|------------------------|-------------------|----------|-------------|-------------|
| FP : Commercial | 0.121 | 0.468 | 0.527 | 0.400 |
| FP : Capital Op | 0.306 | 0.504 | 0.405 | 0.584 |
| FP : Project | 0.378 | 0.640 | 0.718 | 0.592 |
| FP : Real Estate | 0.398 | 0.606 | 0.553 | 0.615 |
| FP : Asset Leasing | 0.544 | 0.620 | 0.583 | 0.635 |
| FP : Export | 0.581 | 0.568 | 0.593 | 0.559 |
| FP : Commodities Based | 0.806 | 0.615 | 0.452 | 0.666 |

TABLE III.4 – Score de classification avec le meilleur seuil

De cette analyse, il ressort le rarorc n'a pas la même importance quel que soit l'objet de financement considéré. Sur les financements pour un objectif commercial ou pour un investissement en capital, cette métrique paraît sans importance. En revanche pour les financements d'actifs (immobilier ou non) ou de projet, le RARORC est un peu plus discriminant. Les tableaux III.3 III.4 permettent d'analyser les scores d'une étude d'approbation des deals exclusivement basée sur le RARORC. Les seuils différenciés selon les objectifs de financement permettent d'obtenir une meilleure classification. On notera par ailleurs que ces seuils sont en cohérence avec les niveaux de distributions de rarorc obtenus à la figure III.1. Cette remarque corrobore l'ensemble des commentaires de la section III.3. Le pilotage de la rentabilité au niveau deal à partir d'un seuil unique de rarorc pour l'ensemble des deals ne serait pas très pertinent. Ce résultat empirique confirme aussi les conclusions de plusieurs publications sur le sujet. Crouhy et al. [5] démontrent comment le RAROC varie avec la volatilité et le risque systématique des actifs d'une institution financière. Sur la base d'hypothèses paramétriques plausibles, ils montrent que les conséquences économiques pourraient être importantes si les variations de la volatilité et/ou du risque systématique sont ignorées. Cela pourrait conduire à accepter (rejeter) à tort des projets à haut risque (faible risque), et donc à prendre des décisions non optimales. Cependant, Crouhy et al. n'examinent pas comment prendre en compte de manière appropriée les risques systématiques dans la mesure de la performance ajustée au risque de chaque ligne financement. En considérant une économie avec un facteur unique, Milne et Onorato [10] démontrent théoriquement que l'utilisation d'un seul seuil raroc n'est appropriée que sous des hypothèses de distribution très strictes (par exemple, les rendements des actifs suivant une distribution normale multivariée) qui pourraient être irréalistes. Ils montrent que, s'il existe des différences dans la forme des distributions des rendements des différents actifs, le rendement par rapport au risque global du portefeuille ne peut pas être optimisé en utilisant un seul seuil raroc pour l'ensemble du portefeuille.

Cependant cette étude n'a pas pour objet la réflexion sur le choix des seuils différenciés car, comme décrit précédemment, la métrique RARORC elle-même n'est pas très pertinente sur certains segments du portefeuille.

III.6 Conclusion partielle

L'information sur la rentabilité ajustée au risque de crédit ne peut conditionner de façon exclusive l'approbation d'un deal sur un portefeuille de financements structurés. Les premiers éléments d'imperfection de cet indicateur sont liés à sa formulation. Avec l'utilisation d'instrument de couverture du défaut de l'emprunteur, il est fréquent sur certaines lignes de métier de se retrouver avec un capital réglementaire faible et un rarorc très élevé. Cette situation rend difficile le pilotage de la rentabilité à partir du rarorc sur ces lignes de métier. Sur d'autres lignes de métiers, les accords de transactions peuvent finir rapidement sur le marché secondaire. De

ce fait le rarorc calculé en tenant compte de la maturité initiale peut être une surestimation (ou une sous-estimation si l'opération est associée à des frais de commissions importantes) de la rentabilité effective. Un troisième point négatif relatif au rarorc concerne la prise en compte partielle d'autres sources de risque comme le risque de migration de rating associé aux transactions. Pour toutes ces raisons, il est impossible d'avoir un seuil unique de rarorc pour l'acceptation des deals. Toutefois l'objectif de respecter un seuil sur une ligne de métier spécifique est réaliste tant que cette dernière ne réalise pas de façon prépondérante des financements pour des raisons commerciales.

Chapitre IV

Modélisation de la décision d’approbation

La rentabilité, même ajustée au risque porté par le deal, ne constitue pas toujours le point susceptible de conduire à l’approbation d’un deal. Des considérations commerciales, opérationnelles, environnementales ou sociales peuvent également peser dans la décision d’approbation. La logique voudrait tout de même que l’objectif de rentabiliser les fonds propres soit des plus prioritaires. Ce chapitre nous permet de proposer une modélisation de cette décision à partir des modèles linéaires généralisés et d’un algorithme de tree boosting (Extreme gradient boosting). Cette modélisation nous permettra d’analyser l’importance des différents facteurs en jeu dans le processus d’approbation et d’étudier les interactions entre ces facteurs. Les facteurs considérés (désignés ci-après comme X_j) sont l’ensemble des variables entrant dans le cadre décisionnel en matière de crédit.¹

IV.1 Théorie du modèle linéaire généralisé.

Le modèle linéaire généralisé constitue une extension souple des moindres carrés introduite par Nedler & Wedderburn (1972)[12]. Dans cette Section, nous présentons la théorie sur laquelle repose ce type de modèle afin de mieux comprendre les indicateurs et les analyses utiles pour notre étude.

IV.1.1 Étendre le modèle linéaire

Le modèle linéaire cherche à expliquer une variable Y par une fonction linéaire appliquée à un ensemble de variables explicatives X_1, \dots, X_p . Nous considérons dans tout ce chapitre n observations et p variables explicatives. Plus précisément, on dit que l’espérance conditionnelle de Y sachant X est une transformation affine en les paramètres et on peut représenter le modèle à l’aide de l’écriture matricielle suivante :

1. Selon le point 67 la guideline EBA (European Banking Authority) EBA/GL/2020/06 « Le cadre décisionnel en matière de crédit doit tenir compte de la dimension du risque dans la prise de décision. Il doit également tenir compte des particularités des produits de crédit et des emprunteurs, notamment le type de produit, la taille de la facilité ou de la limite de crédit, et le profil de risque de l’emprunteur. »

$$\mathbb{E}[Y | X] = X\beta + \varepsilon$$

où Y représente le vecteur réponse de taille $n \times 1$, X la matrice $n \times (p + 1)$ des variables explicatives, β le vecteur $(p + 1) \times 1$ des coefficients et ε le vecteur $n \times 1$ des erreurs ou résidus. Nous écrivons la dimension de certains éléments du modèle avec $(p + 1)$ afin d'indiquer l'ajout d'une variable explicative constante qui constitue l'ordonnée à l'origine.

Le modèle linéaire va estimer β tel que la prédiction \hat{Y} obtenue soit la meilleure possible au sens des moindres carrés. Ce choix de méthode d'estimation repose sur les hypothèses de Gauss-Markov et sur les hypothèses de normalité suivantes :

1. Non colinéarité des variables explicatives ;
2. Indépendance des erreurs : $\forall i \neq j, \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$;
3. Variables explicatives exogènes (non corrélées aux erreurs) : $\mathbb{E}[\varepsilon | X] = 0$;
4. Homoscédasticité : les erreurs sont de variance constante ;
5. Normalité des erreurs : $\varepsilon | X \sim \mathcal{N}(0, \sigma^2 I_n)$.

$X\beta$ étant déterministe, l'hypothèse 2 implique l'indépendance des Y et l'hypothèse 5 leur normalité avec $Y | X \sim \mathcal{N}(X\beta, \sigma^2 I_n)$. Cette dernière hypothèse ne se vérifie pas dans le cadre où Y prends des valeurs binaires, des valeurs positives, ou est une variable de comptage. Le modèle linéaire généralisé permet de s'affranchir de cette hypothèse de normalité.

Pour ce faire, le GLM relie le modèle linéaire a la variable de réponse à partir d'une fonction de lien g , un prédicteur linéaire η et en laissant la variance de chaque prédiction dépendre de sa valeur espérée. Par conséquent, pour une variable Y appartenant à une certaine famille que nous décrivons dans la sous-section suivante, l'équation du GLM devient :

$$g(\mathbb{E}[Y | X]) = \eta = X\beta$$

Nous allons voir à présent les différentes composantes de ce modèle et les conditions nécessaires à son estimation.

IV.1.2 La famille exponentielle

Les modèles linéaires généralisés étendent les distributions acceptées à l'ensemble de la famille exponentielle. Cette dernière regroupe un grand nombre de lois courantes : normale, exponentielle, gamma, Poisson, Bernoulli et d'autres. Pour Y une variable aléatoire et θ un paramètre, Y appartient à la famille exponentielle si sa densité par rapport à une mesure dominante peut s'écrire de la forme suivante :

$$f_Y(y, \theta) = a(y)b(\theta) \exp[\Lambda(\theta) - T(y)]$$

où $a(\cdot), b(\cdot), \Lambda(\cdot)$ et $T(\cdot)$ sont des fonctions mesurables. Dans cette définition, la mesure dominante est celle de Lebesgue pour les lois continues et une combinaison de masses de Dirac pour les lois discrètes. Les densités de la famille exponentielle peuvent également s'écrire de la manière suivante :

$$f_{e,\phi}(y) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)$$

où :

- $a(\cdot)$ et $c(\cdot)$ sont des fonctions dérivables avec $c(\cdot)$ définie sur \mathbb{R}^2 ;
- $b(\cdot)$ est de classe C^3 et de dérivée première inversible ;
- θ est le paramètre d'intérêt ;
- ϕ est le paramètre de dispersion.

L'écriture de la densité fait apparaître une propriété fondamentale de la famille exponentielle :

$$\begin{cases} \mathbb{E}[Y] = b'(\theta) \\ \mathbb{V}[Y] = b''(\theta)a(\phi) \end{cases}$$

Dans notre étude, nous nous plaçons tout d'abord dans un contexte de classification binaire, c'est-à-dire que nous cherchons tout d'abord à discriminer les propositions de contrats ayant abouti de tous les autres. Pour modéliser un tel phénomène la loi de bernoulli constitue une approche pertinente. La densité de cette loi étant donnée par :

$$f_p(y) = p^y(1-p)^{(1-y)}$$

où $x \in 0; 1, p = \mathbb{P}(Y = 1)$. Nous pouvons réécrire cette densité de la manière suivante :

$$f_\lambda(y) = \exp\left(\frac{y \ln\left(\frac{p}{1-p}\right) + \ln(1-p)}{1}\right)$$

où nous retrouvons l'écriture 4.3 avec :

$$\begin{aligned} - \theta &= \ln\left(\frac{p}{1-p}\right) \Leftrightarrow p = \frac{\exp(\theta)}{1 + \exp(\theta)}; \\ - a(\phi) &= 1; \\ - b(\theta) &= -\ln(1-p) \Leftrightarrow b(\theta) = \ln(1 + \exp(\theta)); \\ - c(y, \phi) &= 0. \end{aligned}$$

Nous pouvons de plus vérifier que nous retrouvons bien :

$$\begin{cases} \mathbb{E}[Y] = b'(\theta) = p \\ \mathbb{V}[Y] = b''(\theta)a(\phi) = p(1-p) \end{cases}$$

Enfin, la fonction g telle que $g(\mathbb{E}[Y]) = \theta$, appelée fonction de lien canonique, est la fonction

logit. Nous allons détailler à présent le rôle de la fonction de lien dans les modèles linéaires généralisés.

IV.1.3 La fonction de lien

La fonction de lien établit la relation entre le prédicteur linéaire $\eta = X\beta$ et la moyenne de la distribution de $Y | X$. Il existe pour toute fonction de la famille des lois exponentielles une fonction de lien canonique qui lie cette moyenne au paramètre d'intérêt θ .

Voici quelques exemples de fonctions de lien canoniques usuelles :

| Modèle | Distribution | Support | Nom du lien | Fonction de lien |
|--------------|--------------|--------------|-------------|--|
| linéaire | Normale | \mathbb{R} | Identité | $g(y) = y$ |
| log-linéaire | Poisson | \mathbb{N} | Logarithme | $g(y) = \ln(y)$ |
| logistique | Bernoulli | $\{0, 1\}$ | Logit | $g(y) = \ln\left(\frac{y}{1-y}\right)$ |

TABLE IV.1 – Liens canoniques de lois usuelles

Nous retrouvons, dans le cas de la distribution normale, le modèle linéaire qui se construit avec une fonction de lien identité ce qui revient à exprimer directement la moyenne de la variable réponse en fonction du prédicteur linéaire et correspond à un modèle additif. Bien qu'il serait théoriquement plus rigoureux de choisir la fonction de lien canonique à chaque construction de modèle, toute fonction réelle, monotone, différentiable et définie sur le bon ensemble peut être utilisée. Dans le cas où la variable à expliquer Y est binaire, outre la transformation logit, On trouve notamment dans la littérature les transformations suivantes :

- probit, qui n'est autre que l'inverse de la fonction de répartition de la loi normale centrée réduite :

$$\forall p \in [0, 1], \text{probit}(p) = \varepsilon \quad \text{avec} \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\varepsilon} \exp\left(-\frac{1}{2}u^2\right) du = p$$

- log-log définie par :

$$\forall p \in [0, 1], \text{loglog}(p) = \log(-\log(1 - p)).$$

La transformation log-log est bien appropriée aux cas où l'on souhaite modéliser les probabilités de succès de manière asymétrique. Les transformations logit et probit possèdent des propriétés identiques. La transformation logistique est préférée puisqu'elle est plus simple à manipuler d'un point de vue numérique (résolution des équations d'estimation des paramètres) et permet de fournir une interprétation claire des coefficients en terme de odds-ratio.

IV.1.4 Estimation des paramètres

Nous avons détaillé les trois éléments essentiels qui composent un GLM :

1. Une distribution de probabilité issue de la famille exponentielle pour Y ;
2. Un prédicteur linéaire $\eta = X\beta$;
3. Une fonction de lien telle que $\mathbb{E}[Y | X] = g^{-1}(\eta)$. Il s'agit à présent d'estimer les paramètres β_0, \dots, β_p du modèle. Nous utilisons la méthode du maximum de vraisemblance qui est utilisée par défaut dans la plupart des logiciels, bien que d'autres méthodes existent.

IV.1.4.1 Équation du maximum de vraisemblance

Sous l'hypothèse d'indépendance des Y_1, \dots, Y_m , la log-vraisemblance du modèle s'écrit :

$$L(\beta) = \sum_{i=1}^m \ln(f_{\theta, \phi}(y_i)) = \sum_{i=1}^m \left[\frac{y_i \theta - b(\theta)}{a(\phi)} + c(y_i, \phi) \right] := \sum_{i=1}^m l_i$$

Les paramètres du modèle sont inclus de manière implicite dans l'équation ci-dessus via la relation suivante :

$$\mathbb{E}[Y | X] = \mu = b'(\theta) = g^{-1}(X\beta)$$

Nous devons annuler la dérivée pour satisfaire à notre problème d'optimisation. En dérivant terme à terme la log-vraisemblance de chaque observation nous obtenons :

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta} \frac{\partial \theta}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$$

Calculons chacune des dérivées partielles :

$$\left\{ \begin{array}{l} \frac{\partial l_i}{\partial \theta} = \frac{y_i - b'(\theta)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)} \\ \frac{\partial \theta}{\partial \mu_i} = \frac{\partial^2 b(\theta)}{\partial \theta^2} = \frac{\mathbb{V}[Y_i | X]}{a(\phi)} \\ \frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} = (g^{-1})'(\eta_i) \\ \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial X_i \beta}{\partial \beta_j} = x_{ij} \end{array} \right.$$

Nous obtenons ainsi, pour tout $j = 1, \dots, p$, les équations de vraisemblance suivantes :

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^m \frac{y_i - \mu_i}{\mathbb{V}[Y_i | X]} \cdot x_{ij} \cdot (g^{-1})'(\eta_i) = 0$$

IV.1.4.2 Résolution

La résolution de ces équations nécessite l'utilisation de méthodes itératives car il n'existe généralement pas de formules fermées pour trouver l'estimateur du maximum de vraisemblance $\hat{\beta}$. L'algorithme IRLS (Iteratively Re-weighted Least Squares) s'applique avec les méthodes de type Newton-Raphson ou du score de Fisher. Leur principe consiste à estimer successivement des jeux de paramètres avec certaines grandeurs mises à jour à chaque étape jusqu'à obtenir une convergence au sens des moindres carrés. Le lecteur pouvant se référer Breheny & al. (2013)[2] pour plus de détails. Avec la méthode de Newton-Raphson par exemple la mise à jour des coefficients se fait de manière itérative d'après la formule suivante :

$$\hat{\beta}_{k+1} = \hat{\beta}_k - H^{-1} \nabla$$

où les éléments de H , la matrice Hessienne des dérivés secondes de l , sont définis comme $h_{i,j} = \frac{\partial^2 l}{\partial \beta_i \partial \beta_j}$.

IV.2 Approche par les méthodes d'apprentissage automatique

Dans cette section nous présentons le cadre de modélisation par des approches de machine learning. Ce cadre nous permet de construire un socle commun pour l'ensemble des modèles utilisés dans cette étude.

IV.2.1 Principe général

Le point de départ en apprentissage automatique est un échantillon Z_1, \dots, Z_n . Il existe deux grands types de machine learning caractérisés par la nature de l'échantillon : l'apprentissage supervisé et l'apprentissage non supervisé.

Dans le cadre de l'apprentissage non supervisé, les observations Z_i ne sont pas étiquetées. Parmi les différents objectifs de cet apprentissage, on trouve la réduction de dimension via des méthodes de partitionnements ou clustering qui peut servir à mieux comprendre la structure de nos données (segmentation), créer des groupes d'individus qui sont proches au sens d'une certaine métrique (classification), et à préparer les données pour une prochaine étude. Cela permet également de faire du feature engineering, soit préparer nos données à l'application de méthodes d'apprentissage automatique et créer des nouvelles variables construites à partir de celles existantes afin de capturer de nouvelles parts d'information et améliorer la performance d'un modèle.

En apprentissage supervisé, les observations $Z_i = (X_i, Y_i)$, suivant théoriquement une loi \mathbb{P} , sont composées d'une variable d'entrée $X_i \in \mathbb{R}^d$ (appelée prédicteur ou feature) et d'une variable de sortie Y_i appelée étiquette ou label appartenant à un ensemble fini ou à \mathbb{R} . On observe donc $\mathcal{D}_n = (Y_i, X_i)_{i=1 \dots n}$. Pour une nouvelle entrée $X_{n+1} \in \mathbb{R}^d$, l'objectif est de prévoir Y_{n+1} , la

valeur d'étiquette correspondante en supposant que la paire (X, Y) est une nouvelle réalisation indépendante issue du même phénomène que les réalisations précédemment observées et utilisées pour la construction du modèle. Pour ce faire, les différentes méthodes disponibles reprennent certains éléments que nous allons détailler.

IV.2.2 Fonction de prédiction

On appelle \mathcal{X} l'espace des variables d'entrée X_i et \mathcal{Y} l'espace des variables de sortie Y_i . Une fonction de prédiction est une fonction mesurable qui va de \mathcal{X} dans \mathcal{Y} et qui fait partie de l'ensemble des fonctions de prédiction que l'on note $\mathcal{F}(\mathcal{X}, \mathcal{Y})$.

Un algorithme d'apprentissage est une méthode qui, à partir d'un échantillon d'apprentissage Z_1, \dots, Z_n , renvoie un estimateur de la « meilleure » fonction de prédiction f qui appartient à $\mathcal{F}(\mathcal{X}, \mathcal{Y})$. Nous allons voir comment définir cette meilleure fonction.

IV.2.3 Fonction de perte

Soit $L(Y, Y')$ la perte encourue lorsque notre modèle prédit Y' et que la sortie réelle observée est Y . La fonction $L : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ est alors appelée fonction de perte.

- Exemple 1. $\ell : (y, y') \mapsto 1(y \neq y')$ (binary classification)
- Exemple 2. $\ell : (y, y') \mapsto |y - y'|^2$ (squared error)
- Exemple 3. $\ell : (y, y') \mapsto \max(0, 1 - yy')$ (hinge loss)

La fonction de perte doit être conçue en fonction de l'application concernée.

IV.2.4 Risque théorique et risque empirique

On souhaite prédire au mieux Y_{n+1} à partir de X_{n+1} et \mathcal{D}_n , i.e. résoudre :

$$f^* = \arg \min_{f \in \mathcal{F}(\mathcal{X}, \mathcal{Y})} E_{\mathbb{P}} [L(Y_{n+1}, f(X_{n+1}, \mathcal{D}_n))]$$

La qualité d'une fonction de prédiction f est mesurée par son risque défini par :

$$\mathcal{R}(f) = E_{\mathbb{P}} [L(Y_{n+1}, f(X_{n+1}, \mathcal{D}_n))]$$

si bien que $\mathcal{R}(f)$ représente de la perte esperée encourue sur la donnée (X_{n+1}, Y_{n+1}) par la fonction f . Le risque conditionnel est défini par

$$\mathcal{R}(f, x) = E_{\mathbb{P}} [L(Y_{n+1}, f(X_{n+1}, \mathcal{D}_n)) \mid X_{n+1} = x]$$

et représente la perte espérée conditionnellement à l'observation de la valeur x pour X_{n+1} . En pratique, on se focalise sur la fonction de risque empirique définie par :

$$\hat{\mathcal{R}}_n(f) := \frac{1}{n} \sum_{i=1}^n L(Y_i, f(X_i, \mathcal{D}_n))$$

IV.2.5 Meilleure fonction de prédiction

La meilleure fonction de prédiction est une fonction f^* de $\mathcal{F}(\mathcal{X}, \mathcal{Y})$ qui minimise le risque $\mathcal{R}(f)$. Une telle fonction appelée fonction oracle ou prédicteur de Bayes est inconnue puisqu'elle dépend de la loi de probabilité \mathbb{P} sous Z .

— Exemple 1. La fonction oracle pour la perte quadratique :

$$f_{\mathbb{P}}^*(x) := \mathbb{E}_{\mathbb{P}}[Y | X = x]$$

L'espérance conditionnelle est la "meilleure approximation" de Y avec l'information induite par

— Exemple 2. La fonction oracle pour la perte binaire (0 et 1) :

$$f_{\mathbb{P}}^*(x) := 1(\mathbb{E}_{\mathbb{P}}[Y | X = x] \geq 1/2)$$

Dans ce cas de la perte binaire où $Y \in \{-1, 1\}$ qui n'est pas un espace vectoriel, donc $\{f : \mathcal{X} \rightarrow \{-1, 1\}\}$ non plus, on observe donc une difficulté pour résoudre un problème de minimisation. Il faudra donc chercher $f : \mathcal{X} \rightarrow \mathbb{R}$ et considérer alors la fonction : $x \rightarrow \text{sign}(f(x))$ comme prédicteur. Le risque devient alors $\mathcal{R}(f) = \mathbb{P}(\text{sign}(f(X)) \neq Y) = E\Phi_{0-1}(Yf(X))$ où : $\Phi_{0-1} = \mathbf{1}_{Yf(X) \leq 0}$ est la perte 0-1. Le risque empirique s'écrit donc :

$$\hat{\mathcal{R}}(f) = \frac{1}{n} \sum_{i=1}^n \Phi_{0-1}(Y_i f(X_i))$$

On cherche à minimiser le risque empirique, mais on ne peut pas le faire directement car Φ_{0-1} est ni continue, ni convexe. Pour cette raison on peut introduire d'autres fonctions de pertes qui permettront d'obtenir les mêmes prédicteurs optimaux que Φ_{0-1} . Il s'agira par exemple de la perte logistique définie par $\Phi(u) = \log(1 + e^{-u})$ donc $\Phi(Yf(X)) = \log(1 + e^{-Yf(X)})$. Le risque empirique associé s'écrit donc :

$$\hat{\mathcal{R}}(f) = \frac{1}{n} \sum_{i=1}^n -\log\left(\frac{1}{1 + e^{-Y_i f(X_i)}}\right)$$

Lorsqu'on considère le modèle défini par : $\mathbf{P}(Y = 1 | X = x) = \frac{1}{1 + e^{-f(x)}}$ et $f(x)$ linéaire en x . Alors le risque est égal à l'opposé de la vraisemblance conditionnelle telle que définie

dans le cadre du modèle GLM ci-dessus ($-L$).

— Exemple 3. La fonction oracle quand $Y \in \{1, \dots, p\}$ avec une perte binaire :

$$f_{\mathbb{P}}^*(x) := \arg \max_k \mathbb{P}(Y = k | X = x)$$

Notre but est donc de trouver une fonction de prédiction qui s'approchera au maximum de l'erreur obtenue avec la fonction oracle. L'ensemble de toutes les fonctions possibles est trop grand en général. On considère pour \mathcal{F} un sous-ensemble de fonctions équivalent à se placer dans le cadre d'un modèle. De plus, nous réduisons généralement le sous ensemble $\mathcal{F}(\mathcal{X}, \mathcal{Y})$ afin de lutter contre le sur-apprentissage, tout en le gardant assez grand pour pouvoir approcher toute fonction. En effet, l'excès de risque qui découle de la fonction de prédiction obtenue par rapport à la fonction oracle se décompose en erreur stochastique (ou erreur d'estimation) et erreur systématique (ou erreur d'approximation). Plus l'espace de nos fonctions candidates retenu est grand, plus l'erreur d'approximation est faible mais celle d'estimation est grande (généralement).

Il y a donc un compromis à trouver : nous retrouvons un dilemme généralement appelé « le dilemme biais-variance. ». De nombreuses techniques permettent de gérer ce compromis. La plupart de ces techniques peuvent être présentées comme une certaine pénalisation. Elles consistent à lutter contre le sur-apprentissage en introduisant une mesure de pénalisation lorsque la fonction obtenue est trop irrégulière. Elles dépendront de la classe d'algorithme à laquelle on restreint $\mathcal{F}(\mathcal{X}, \mathcal{Y})$:

| Function class \mathcal{F} (and its parameterization) | Regularizer |
|---|--|
| Global/parametric predictors | |
| Linear $\beta'x$ (and generalizations) | Subset selection $\ \beta\ _0 = \sum_{j=1}^k \mathbf{1}_{\beta_j \neq 0}$ LASSO $\ \beta\ _1 = \sum_{j=1}^k \beta_j $ Ridge $\ \beta\ _2^2 = \sum_{j=1}^k \beta_j^2$ Elastic net $\alpha\ \beta\ _1 + (1 - \alpha)\ \beta\ _2^2$ |
| Local/non-parametric predictors | |
| Decision/regression trees | Depth, number of nodes/leaves, minimal leaf size, information gain at splits |
| Random forest (linear combination of trees) | Number of trees, number of variables used in each tree, size of bootstrap sample, complexity of trees, ... |

| | |
|---|---|
| Function class \mathcal{F} (and its parameterization) | Regularizer |
| Combined predictors | |
| Bagging : unweighted average of predictors from bootstrap draws | Number of draws, size of bootstrap samples (and individual regularization parameters) |
| Boosting : linear combination of predictions of residual | Learning rate, number of iterations, and individual predictor parameters regularization |
| Ensemble : weighted combination of different predictors | Ensemble weights, and individual predictor parameters regularization |

TABLE IV.2 – Exemple de pénalisation par famille d'algorithme

IV.3 Présentation de l'algorithme du Gradient Boosting

Dans le cadre de cette étude il a été décidé de mesurer les interactions entre les différents facteurs explicatifs à partir d'une modélisation non paramétrique : le gradient boosting. Les algorithmes de gradient boosting d'arbre offrent une certaine continuité avec le GLM. L'idée sous-jacente du gradient boosting découle d'observations faites par Leo BREIMAN [3] qui, se basant sur certains travaux antérieurs, montre que le boosting peut s'interpréter comme un programme d'optimisation sur une certaine fonction de perte. Les meilleurs résultats expérimentaux de Leo Breiman [3] ont été obtenus en utilisant des arbres de décision de taille fixe comme prédicteurs faibles. Ainsi, à chaque étape du processus de boosting, un arbre de décision CART est entraîné sur le gradient de la fonction de perte issu de l'agrégation des arbres créés lors des étapes précédentes.

IV.3.1 Les arbres de décision CART

Les arbres CART (Classification And Regression Tree) sont des arbres de décision permettant de résoudre selon une approche d'apprentissage statistique non paramétrique des problèmes de classification et de régression.

La construction de ces arbres se fait globalement en deux étapes :

- La première est récursive : à chaque étape, l'espace de dimension p formé par les n observations est scindé de façon binaire selon une des p variables explicatives, de façon à réduire l'erreur de prédiction. Un nœud est donc formé à chaque étape en effectuant une division binaire des observations qui le composent de manière à diminuer un critère d'impureté au sein des deux nœuds qui le suivent et qui sont appelés « fils gauche » et « fils

droit ». L'algorithme se poursuit ainsi jusqu'à l'étape r , étape où le critère d'arrêt fixé est vérifié. L'espace de dimension p est alors découpé en r régions R_1, \dots, R_r distinctes.

En supposant qu'il y ait $p = 2$ variables explicatives quantitatives, une représentation d'un arbre CART serait la suivante :

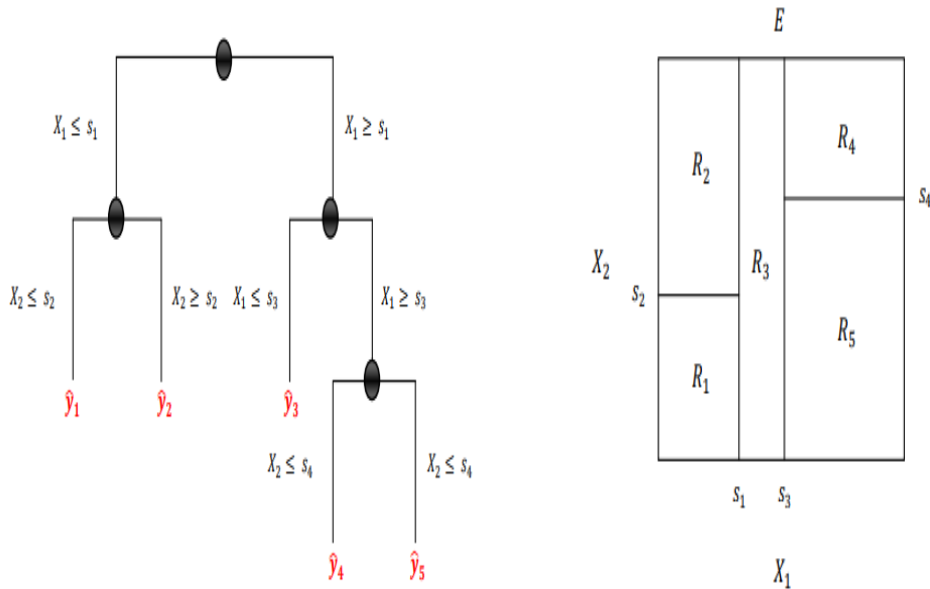


FIGURE IV.1 – Exemple d'arbre CART

FIGURE IV.2 – Pavage didactique de l'espace correspondant à un arbre CART

L'espace est formé par les variables explicatives et est ainsi partitionné en rectangles. La problématique réside autour du choix de cette partition : quel est le meilleur fractionnement de l'ensemble formé par les variables explicatives ?

- La seconde permet d'étiqueter chacune des classes obtenues. Dans le cas d'une régression, la valeur prédite pour une région R_i sera simplement la moyenne de toutes les observations qui s'y trouvent. Dans le cas d'une classification, la règle de prédiction est celle du vote : la classe majoritaire de la région R_i sera attribuée à chaque observation de cette région.

Nous nous focaliserons ici sur la construction d'arbres de classification dont la définition générale est la suivante :

Soient $C = 1, \dots, c$ les classes présentes dans notre échantillon de taille N .

On note : $\pi_j = N_j/N$ la probabilité à priori d'appartenir à la classe j . Avec :

- N_j la taille de la classe j .
- $N(t)$: le nombre d'observations dans le nœud t de l'arbre, et
- $N_j(t)$: le nombre d'observations de la classe j dans le nœud t .

On estime ensuite :

- $P(j, t)$: la probabilité d'appartenance au groupe j dans le nœud t pour une observation.

$$P(j, t) = \pi_j \frac{N_j(t)}{N_j}$$

— $P(t)$. la probabilité qu'une observation soit dans le noeud t , par $P(t) = \sum_j P(j, t)$

— $P(j | t)$, la probabilité à posteriori de la classe j dans t , par $P(j | t) = \frac{P(j, t)}{P(t)}$

On définit ensuite h une fonction de $\{(p_1, \dots, p_C) \mid p_j \geq 0, \sum_j p_j = 1\}$ dans \mathbb{R} .

On dit que h est une fonction d'hétérogénéité si :

— h est symétrique en (p_1, \dots, p_j) ;

— h est maximale en $(1/C, \dots, 1/C)$;

— h est minimale en $(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 1)$.

On définit l'hétérogénéité d'un noeud t par : $i(t) = h(P(1 | t), \dots, P(j | t))$ avec h une fonction d'hétérogénéité. Dans le cas des arbres CART on utilise souvent la fonction d'hétérogénéité de Gini : $1 - \sum_{i \neq j} p_i p_j$ ou la fonction de Shannon $-\sum_j p_j \log(p_j)$. Les deux fonctions donnent une hétérogénéité de 0 à un noeud ne possédant que des observations d'une seule classe et une hétérogénéité maximale à un noeud ayant une répartition égale entre les deux classes. L'hétérogénéité de Gini est en fait la probabilité d'obtenir deux observations de classes différentes.

Pour la construction de l'arbre, on commence par définir à chaque étape un noeud t . On note t_d son descendant droit et t_g son descendant gauche, engendré par une partition δ . On a les proportions d'observations dans les partitions données par $p_g = (P(t_g)) / (P(t))$ et $p_d = (P(t_d)) / (P(t))$, respectivement. On définit enfin la variation d'hétérogénéité engendrée par δ par :

$$\Delta i(\delta, t) = i(t) - P_g i(t_g) - P_d i(t_d)$$

Ainsi on obtient une division optimale par :

$$\delta^*(t) := \delta^* = \operatorname{argmax}_{\delta} \Delta i(\delta, t)$$

En effet, nous cherchons à obtenir des descendants dont l'homogénéité est la plus importante possible. On retient donc la division qui maximise la variation d'hétérogénéité parmi toutes celles qui sont possibles. La construction de l'arbre ne fonctionne que dans le cas où la fonction d'hétérogénéité est concave, car cela permet de garantir que l'optimisation de l'arbre est globale. La construction de l'arbre se fait par récursivité de la méthode précédente, en partant du noeud initial contenant l'ensemble de la population.

Ensuite, on définit une règle d'arrêt, ce qui correspond à définir à partir de quel moment on ne divise plus la population présente dans un noeud. La règle pour l'algorithme CART est de s'arrêter quand toute la population d'un noeud appartient à la même classe réponse. Les feuilles qui ne sont pas scindées sont alors appelées feuilles terminales.

Après ces étapes, nous obtenons un arbre maximal qui ne possède que des feuilles terminales avec une seule classe. On s'aperçoit cependant que l'on obtient selon les cas des arbres avec énormément de feuilles et qui sont très difficiles à interpréter et à utiliser car très complexes. Il y a donc encore une étape à faire pour rendre ces arbres plus simples : il s'agit de l'élagage. En

effet, il serait également possible de modifier la règle d'arrêt pour obtenir un arbre plus simple. Cependant, il faudrait alors choisir une règle et justifier ce choix. Il faudrait de plus que cette règle puisse s'adapter aux données que l'on utilise.

L'élagage

L'élagage intervient donc pour pallier la très grande complexité de l'arbre maximal. En effet celui-ci est très précis mais possède une grande variance, à l'inverse d'un arbre minimal qui est constitué uniquement du noeud initial. Il faut alors trouver un compromis entre les deux.

L'élagage en lui-même consiste à réduire le nombre de feuilles totales de l'arbre. Pour cela on observe de nombreux arbres avec différents degrés d'élagage et pour chacun on utilise une fonction de pénalité en fonction du nombre de feuilles de l'arbre. On calcule aussi une erreur d'ajustement sur chaque arbre obtenu. Cette erreur représente la probabilité qu'une observation soit mal placée dans une feuille. On cherche alors une sorte d'équilibre entre le nombre de feuilles et la précision obtenue par optimisation et on retient l'arbre qui répond le mieux à ce critère pénalisé.

Les avantages de la méthode CART

La méthode CART présente de nombreux avantages. Premièrement, il s'agit d'une méthode non paramétrique (et donc aucune hypothèse n'est faite sur la distribution sous-jacente des données). De plus, et cela peut être particulièrement intéressant dans notre cas, la méthode CART prend en charge toutes sortes de données, qualitatives ou quantitatives dans l'ensemble des variables explicatives. Cela permet d'éviter d'avoir à retraiter les données avant l'utilisation de la méthode. Il existe des algorithmes qui permettent de traiter le problème des données manquantes et grâce auxquels la méthode CART peut effectuer des prédictions même dans le cas où certaines observations sont manquantes pour certaines variables. De par sa structure et son algorithme, on peut noter également que la méthode CART permet une sélection au fur et à mesure des variables significatives en éliminant celles qui sont inutiles. Cela pourra nous servir par la suite, Enfin, la facilité d'interprétation reste l'un des atouts majeurs de ce modèle.

Les inconvénients de la méthode CART

La méthode CART peut être fortement perturbée par des valeurs aberrantes et le bruit en général. De plus la stratégie d'élagage doit être soigneusement calibrée. En outre, on peut reprocher aux arbres CART le fait de n'utiliser qu'une seule variable à chaque fois pour effectuer les divisions. En effet, on pourrait imaginer des effets croisés de certaines variables qui seraient potentiellement expliqués par une combinaison de celles-ci dans la condition de scission, ce qui n'est pas rendu possible avec CART. Enfin, ces arbres sont peu robustes : leur construction dépend beaucoup des données de la base d'apprentissage. Afin de lutter contre cette limite, une méthode souvent utilisée consiste à agréger de nombreux arbres simples.

IV.3.2 Principe de l'eXtrême Gradient Boosting

Le XGBoost fait partir des méthodes de boosting qui construisent de manière itérative un ensemble de prédicteurs faibles puis les agrègent en un unique prédicteur fort. Un prédicteur est considéré « faible » lorsque sa performance est à peine supérieure à un tirage aléatoire ; cette désignation implique généralement une faible complexité et un temps de calcul faibles.

IV.3.2.1 Le boosting

Le boosting est une méthode ensembliste se basant sur l'idée d'utiliser une série de classifieurs dits « faibles » dans laquelle chaque nouveau classifieur servira de complément aux précédents pour corriger l'erreur commise. L'intérêt de cette approche est de se concentrer sur la réduction du biais tout en n'augmentant pas la variance alors que des algorithmes comme Random Forest se concentrent sur la réduction de la variance, potentiellement au détriment du biais. L'approche est ici récursive, à savoir que chaque nouvel arbre est constitué en fonction du précédent car le poids des observations dépendent de la performance de l'arbre précédent. Un accroissement du poids de l'échantillon mal classé et une réduction du poids pour l'échantillon correctement classé est proposé à chaque étape de l'algorithme pour construire un nouveau estimateur. A la fin de l'algorithme un vote pondéré est effectué sur l'ensemble des estimateurs afin d'obtenir le prédicteur final.

IV.3.2.2 Le Xgboost

De nombreux algorithmes ont été développés autour de cette idée. Le Xgboost (pour Extrême Gradient Boosting) est un cas particulier de l'algorithme de gradient boosting qui a été proposé en 2016 par Chen et Guestrin [4]. Cette version vise à corriger les problèmes de l'algorithme initial en introduisant notamment des paramètres permettant de contrôler le sur-apprentissage. L'algorithme de base est celui du boosting auquel on ajoute la technique de descente de gradient sur une fonction de perte. En notant $\mathcal{F} = \{f(\mathbf{x}) = w_{q(\mathbf{x})}\} (q : \mathbb{R}^p \rightarrow T, w \in \mathbb{R}^T)$ l'espace des arbres CART, avec T représentant le nombre de feuilles, q la structure de l'arbre, w le vecteur de poids associé à ces feuilles, l'estimateur \hat{y}_i et l'objectif d'optimisation peuvent s'écrire :

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F},$$

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

$$\text{où } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

avec l la fonction de perte différentiable et convexe (la logloss dans notre cas de classification binaire) et Ω une fonction de pénalisation de la complexité du modèle. Dans cette fonction de pénalisation, γ représente la réduction minimale de la perte requise pour effectuer une nouvelle

partition sur un nœud feuille de l'arbre. λ représente un autre terme de régularisation qui affecte directement les poids à l'image de la pénalisation ridge. Plus γ ou λ est grand, plus l'algorithme sera conservateur. La fonction objective à l'étape t peut se réécrire :

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t)$$

L'algorithme utilise des statistiques de gradient de premier et second ordre de la fonction de perte pour minimiser l'objectif régularisé. Le lecteur peut se référer à Chen et Guestrin (2016)[4].

IV.4 Mesure d'influence relative des variables

L'objet de cette section est de présenter le cadre théorique qui permet de mesurer l'influence d'une variable explicative dans un modèle prédictif. Dans un premier temps nous expliquons comment quantifier l'importance d'une variable dans un GLM et dans un second temps nous exposerons la mesure d'influence utilisée dans le cadre des modèles complexes.

IV.4.1 Mesure d'influence dans le modèle linéaire généralisé

IV.4.1.1 Tests de significativité des variables

Afin de tester la significativité des variables, il est possible de recourir à un test se basant sur un écart de vraisemblance. Ce test consistera à calculer, pour chacune des variables explicatives choisies, l'écart entre la vraisemblance du modèle choisi et la vraisemblance du modèle choisi privé de la variable i (de toutes ses modalités). Rappelons que le modèle étudié s'écrit :

$$g(\mathbb{E}[Y | X]) = \beta_0 + \beta_1 * X_1 + \dots + \beta_p * X_p$$

Sous l'hypothèse que $\beta_i = 0$, en notant $\hat{\beta}_{-i}$ l'estimateur du maximum de vraisemblance de β dans le modèle privé de la variable t , la statistique de test S définie comme suit :

$$S = \frac{2 * (l(\hat{\beta}) - l(\hat{\beta}_{-i}))}{\Phi}$$

suit une loi du Khi ² à (nombres de modalités de la variable i) - 1 degrés de liberté. Ainsi, en définissant le test suivant, on peut tester la significativité de chaque variable :

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

En effet, en définissant la p - value de ce test comme :

$$p - \text{value} = \mathbb{P}(K > S) \quad \text{où } K \sim \chi_{p-1}^2$$

alors si $p - \text{value} > 5\%$, la variable i n'est pas significative. Il est à noter que plus la valeur de S est grande, plus la variable i contribue à expliquer la variable d'intérêt Y .

IV.4.1.2 Les odds ratio dans le modèle logistique

Le modèle linéaire généralisé utilisé avec le lien logit possède l'avantage non négligeable de permettre la mesure de l'influence de chaque facteur qualitatif sur le risque à travers les « Odds Ratios », lesquels expriment le degré de dépendance entre des variables explicatives. L'odds-ratio d'une variable explicative calcule le rapport des chances d'apparition de l'événement $Y = 1$ contre $Y = 0$ lorsque X_j passe de x à $x + 1$. Autrement dit, ils se calculent de la manière suivante :

$$OR = \frac{\pi(x+1)/1 - \pi(x+1)}{\pi(x)/1 - \pi(x)} = \exp \beta_j$$

Ce qui s'écrit également (dans le cas binaire) :

$$OR = \frac{P(Y = 1 | X_j = 1) / P(Y = 0 | X_j = 1)}{P(Y = 1 | X_j = 0) / P(Y = 0 | X_j = 0)} = \exp \beta_j$$

En règle générale, un odds-ratio strictement inférieur à 1 (i.e. un coefficient $\beta < 0$) signifie une influence négative de la variable explicative sur la variable à prédire et inversement s'il est strictement supérieur à 1. Afin que ce dernier soit significatif il ne faut pas que l'intervalle de confiance à 95% de celui-ci contienne la valeur 1.

IV.4.2 Mesure d'influence dans des modèles complexes

L'interprétabilité des modèles d'apprentissage automatique permet de rationaliser leurs décisions. Ainsi, comprendre les raisons pour lesquelles une prédiction est faite par un modèle complexe réduit ou élimine son caractère de boîte noire. À cette fin, la méthode SHAP (SHapley Additive exPlanations) a été récemment introduite. L'approche SHAP nous a permis d'identifier et d'hierarchiser les caractéristiques qui déterminent la finalité (approuvée, non approuvée ou perdue) d'une demande de financement structuré dans un modèle complexe. Pour déterminer la juste contribution d'une variable dans la prédiction d'un résultat, le calcul de la valeur SHAP (Lundberg et Lee 2017) [9] se base sur une mesure de shapley introduite par la théorie des jeux et des modèles d'approximation locale.

IV.4.2.1 Valeur de Shapley classique

Un jeu de coalition avec P joueurs, pouvant former une coalition S est considéré. Chaque coalition a un payoff v ($v : 2^P \rightarrow \mathbb{R}$). La valeur de Shapley est la seule qui assigne de manière « juste » un score à chaque joueur suivant leur contribution dans toutes les coalitions possibles. Ainsi, un opérateur ϕ est cherché de sorte qu'il assigne au jeu $J = (\{1, \dots, p\}, v)$, un vecteur $\phi =$

(ϕ_1, \dots, ϕ_p) de gains. Dans le contexte de la prédiction d'une variable à partir d'une observation x , les joueurs sont les variables explicatives, le jeu est la prédiction et la fonction de payoff est la dépendance partielle, qui peut être décalée par rapport à la prédiction moyenne c'est-à-dire :

$$v(x_S) = \mathbb{E}_{X_{P \setminus S}} [\hat{f}(x_S, X_{P \setminus S})] - \mathbb{E}_X [\hat{f}(X)] = f_S(x_S) - f_\emptyset(x_\emptyset)$$

Ainsi, pour chaque coalition S est définie la contribution marginale de la variable $j \in \{1, \dots, p\}$ par :

$$\Delta_j(x_S) = v(x_{S \cup \{j\}}) - v(x_S) = f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)$$

Pour une observation $x = (x_1, \dots, x_p)$, la valeur prédite $f(x)$ se décompose de la façon suivante :

$$f(x) = \sum_{j=1}^n \underbrace{\Phi_j(\Delta^x)}_{\substack{\text{contribution de la} \\ \text{variable explicative } X_j \\ \text{à la prédiction } f(x)}} + \underbrace{\mathbb{E}(f(X))}_{\substack{\text{espérance des prédictions} \\ \text{sur la base de apprentissage}}}$$

$\Phi_j(\Delta^x)$ est la valeur de shapley pour la variable X_j du jeu coopératif $(P = \{X_1, \dots, X_p\}, v)$:

$$\Phi_j(\Delta^x) = \sum_{S \in \mathcal{P}(P \setminus \{X_j\})} \frac{|S|!(p - |S| - 1)!}{p!} * \Delta_j(x_S)$$

La contribution finale ou valeur de Shapley de la variable X_j est donc déterminé comme la moyenne de ses contributions dans toutes les coalitions possibles sur l'ensemble des variables. Pour construire un indicateur d'importance globale, noté I_j pour la variables X_j en agrégeant les valeurs shapley sur la base d'apprentissage :

$$I_j = \frac{1}{k} \sum_{i=1}^N |\Phi_j^{(k)}|$$

où N est la taille de la base d'apprentissage et $\Phi_j^{(k)}$ est la valeur de shapley de la variable j associée à l'individu k .

Les contributions sont généralement réparties entre les variables d'entrées, mais des informations supplémentaires peuvent être fournies en séparant les effets d'interactions des effets principaux. Ainsi, les effets d'interactions d'ordre deux seront donnés par l'impact de tous les couples de variables sur une prédiction du modèle. L'extension naturelle aux effets d'interaction, à partir des valeurs shapley, est obtenue par la matrice symétrique des indices d'interaction de Shapley :

$$\Phi_{i,j} = \sum_{S \subseteq P \setminus \{i,j\}} \frac{|S|!(n - |S| - 2)!}{2(n - 1)!} \nabla_{i,j}(S)$$

avec $i \neq j$ et

$$\begin{aligned} \nabla_{i,j}(S) &= f_x(S \cup \{i, j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S) \\ &= f_x(S \cup \{i, j\}) - f_x(S \cup \{i\}) - [f_x(S \cup \{j\}) - f_x(S)] \end{aligned}$$

Le calcul de la valeur de Shapley classique présente de nombreux inconvénients. Un calcul exact de cette dernière est coûteux en temps et calcul. Par exemple pour une seule variable le calcul de Φ_j nécessite la construction de 2^{p-1} modèles prédictifs. Le nombre de variables utilisées accroît les difficultés de calcul et l'interprétabilité des prédictions. De plus, le moindre changement dans les données utilisées entrainera de nouveaux calculs.

IV.4.2.2 L'approximation locale

Pour l'explication de prédictions individuelles, une compréhension globale du modèle d'apprentissage automatique n'est pas indispensable. Au contraire, des approximations ou des explications locales sont suffisantes pour rationaliser les décisions du modèle. Les explications des décisions individuelles ont été proposées par Ribeiro et al. [14] et désignées sous le nom de LIME (Local Interpretable Model-agnostic Explanations). L'approche LIME vise à trouver un modèle simple qui se rapproche localement du modèle initial complexe au voisinage chaque observation. De tels modèles explicatifs locaux peuvent être définis comme une fonction linéaire de variables binaires selon l'équation suivante :

$$g(x') = \Phi_0 + \sum_{j=1}^{|p|} \Phi_j x'_j$$

où $x' \in \{0, 1\}^p$ correspond à l'absence présence des p variables explicatives, $\Phi_j \in \mathbb{R}$.

Ainsi, un modèle explicatif local approprié est obtenu en minimisant une fonction de perte et en pénalisant la complexité du modèle par un terme de régularisation selon l'équation :

$$\xi(x') = \underset{g}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_{x'}) + \Omega(g)$$

Ici, f est le modèle complexe original, $\pi_{x'}$ est une fonction kernel, et Ω le terme de régularisation qui permettra de sélectionner les variables pertinentes. La fonction kernel définit la similarité, par rapport à x entre l'observation et la valeur prédite à expliquer et détermine donc la localité du modèle.

L'objectif est donc de trouver le meilleur modèle interprétable g traduisant le comportement de f pour la prédiction d'une instance donnée x . Par meilleur, nous entendons celui décrivant au mieux f localement à l'aide des variables les plus pertinentes. Pour ce faire, il est primordial de définir une fonction de fidélité L , fonction qui mesure le niveau d'infidélité de g dans l'approximation de f dans la localité définie par π_x .

Le fait que LIME sélectionne les variables pertinentes lors de la recherche du meilleur modèle

explicatif local possible permet d'obtenir une prédiction qui ne dépend pas toujours de toutes les variables prises en compte. L'approche LIME présente aussi quelques inconvénients. LIME se base sur une interprétation locale, donc si localement, le comportement du modèle d'origine reste complexe, alors LIME ne pourra fournir une bonne approximation de ce dernier. De plus il n'est pas évident de définir le bon voisinage à partir du choix d'une fonction kernel.

IV.4.2.3 Le kernel SHAP

L'approche du kernel SHAP est basée sur une combinaison de l'approche LIME et des l'approche classique des shapley value. Plus précisément, les paramètres de (c'est-à-dire la perte, le noyau et la complexité) de l'approche LIME sont définis selon le formalisme de la valeur de Shapley. Ainsi, le kernel SHAP approxime les contributions des variables comme des valeurs de Shapley, tandis que l'approche originale LIME définit la localité d'une instance à expliquer de manière heuristique.

La méthode SHAP à noyau nécessite un ensemble de données de base pour l'apprentissage. L'absence d'une variable est simulée en remplaçant les valeurs de cette variable par des valeurs moyennes. Ensuite, un modèle de régression linéaire pondéré est estimé en tant que modèle d'explication g sur la base d'échantillons artificiels générés en activant ou désactivant la présence des variables, ce qui correspond à la prise en compte de différents ensembles de variables. Les coefficients du modèle g sont les valeurs SHAP déterminant l'importance des caractéristiques.

IV.5 Application pratique

IV.5.1 Préparation des données et sélection des variables

Une fois les données regroupées, il est impérative de passer à leurs nettoyage avant la phase de modélisation. Cette étape permet de minimiser l'erreur de prédiction causée par les valeurs manquantes ou aberrantes.

IV.5.1.1 Préparation des données

La vérification et l'imputation des valeurs manquantes des variables explicatives font parties du processus de nettoyage des données pour créer un ensemble de données valides et fiables. Chaque variable contenant des valeurs manquantes a été imputée de manière à conserver la distribution de la variable sans supprimer aucun enregistrement ou information. L'imputation des valeurs manquantes a permis de conserver la taille complète de l'échantillon, ce qui peut être avantageux pour gagner en précision au vu de la volumétrie très limitée de données dont nous disposons. Notons que cette imputation n'est indispensable que dans le cadre des modèles linéaires généralisés. La stratégie d'imputation adoptées est celle de l'algorithme d'imputation

itérative MissForest[15] en utilisant la valeur médiane des variables comme point d'initiation et les arbres de décision comme estimateurs des valeurs manquantes.

Une valeur aberrante est un objet qui présente des caractéristiques différentes de celles d'un objet commun ou un objet qui se trouve à une distance anormale des autres objets d'une caractéristique. Dans le cadre du nettoyage des données, les valeurs aberrantes peuvent être détectées et corrigées par une méthode statistique. En se basant sur les caractéristiques de la distribution de chaque variable, une valeur peut être déclarée comme aberrante si sa valeur est supérieure à une limite supérieure ou inférieure à une limite inférieure (définies en fonction d'un seuil quantile s_{sup} et s_{inf}). La façon la plus simple de traiter une valeur aberrante est de remplacer sa valeur par le seuil. Ce traitement a mené sur la plupart des variable en lien avec l'exposition et la rentabilité. Dans le cas particulier du rarorc, les valeurs aberrantes trouvaient du sens puisque une valeur du rarorc trop élevé traduit un une consommation de capital réglementaire trop faible voir nulle sur le financement. Nous avons jugé cohérent, au vu de la distribution de cette variable, de retenir comme seuil $s_{sup} = 200\%$ et $s_{inf} = -50\%$.

Compte tenu de la diversité des financements structurés et des caractéristiques de ses financements, nous avons eu besoin de procéder à des regroupement sur plusieurs variables nominales. Ces regroupements ont lieu selon des avis des experts métiers ou selon l'objectif de construire des classes équilibrées.

IV.5.1.2 Processus de sélection des variables explicatives

Une fois la base d'analyse constituée, il convient d'effectuer une analyse exploratoire (analyse univariée et bivariée) afin d'avoir une vue d'ensemble des variables contenues dans cette base et d'identifier les valeurs atypiques. Cette analyse permet également de constituer une liste de variables basée sur un avis métier (pré-sélection experte) et sur des indicateurs statistiques (pré-sélection statistique). L'objectif est d'avoir des variables les plus discriminantes.

Au début du processus de sélection des variables explicatives, l'avis expert-métier est indispensable. Il permet d'exclure certaines variables susceptibles d'être utilisées. L'idée étant de se référer aux expériences et compétences des experts-métiers afin d'éliminer les variables qui semblent moins pertinentes dans le cadre de l'étude. Après avoir effectué cette première pré-sélection des variables, la phase suivante est d'étudier la liaison entre les variables afin d'éviter certains problèmes tels que la multicolinéarité pour les modèles GLM. Lorsque que nous prenons par exemple deux variables qui sont très corrélées, outre les problèmes de convergence algorithmique, la contribution en termes d'information de l'une des deux variables est presque la même que celle de la seconde. Les estimations des paramètres du modèles peuvent donc être perturbées et l'interprétation des résultats devient difficile. L'analyse de la corrélation entre les variables se fait au moyen du V de cramer pour chaque combinaison de variables qualitatives et au moyen tau de kendall ou du rho de spearman pour les variables continues. De plus, pour évaluer la multicolinéarité entre les variables, l'indicateur utilisé est le VIF. Un

algorithme permettant d'éliminer de façon récursive les multicollinéarités à partir d'un contrôle sur la VIF est utilisé. Cet algorithme élimine les variables en vérifiant leurs facteurs d'inflation de la variance (VIF), qui évaluent dans quelle mesure la variance d'un coefficient de régression estimé augmente lorsque les variables sont corrélées. Les VIF sont calculés sur la base de l'équation :

$$VIF = \frac{1}{1 - R^2}$$

où R^2 peut être obtenu en effectuant une régression linéaire (log linéaire pour les variables faisant référence à des montants) de la variable sur toutes les autres variables. Le seuil VIF peut être calculé approximativement à l'aide de l'équation en définissant le niveau maximum de R^2 entre les variables. Nous avons défini un seuil VIF correspondant à 80% de R^2 . Ce seuil assez élevé permet d'éliminer les cas de forte multicollinéarité et se justifie par l'application de pénalisation aux différents modèles (régression logit pénalisée, xgboost avec un beta non nul).

IV.5.1.3 Les variables retenues

| Variables | modalités |
|-------------------|---|
| syndication type | syndication : Participant, syndication : Club Deal, syndication : Best Effort, syndication : BilateralCredit, syndication : Underwriting |
| garantie type | garantie : credit agency, garantie : garantor, garantie : no garantie, garantie : multilat agency, garantie : private insurer |
| financing purpose | FP : Commercial, FP : Export, FP : Project, FP : Real Estate, FP : Capital Op, FP : Commodities Based, FP : Asset Leasing, FP : Structured grid |
| country | country : France, country : in Others EU, country : UK, country : in North America, country : USA, country : in South America, country : in Oceania, country : in Asia, country : in Africa |
| rating | rating missing, rating sup 6, rating 5, rating 4, rating inf 3 |

TABLE IV.3 – variables sélectionnées

Au cours de la sélection des variables, les caractéristiques du deal relatives aux indicateurs intermédiaires du RARORC retenues sont la lgd synthétique, et le NBI de la première année en

pourcentage.

IV.5.2 Stratégie de validation

Séparation de l'échantillon en train set et validation set

La division de l'ensemble de données en deux parties (ensemble de données d'apprentissage et ensemble de données de validation) est une étape préparatoire à la création d'un modèle de classification valide avec une précision fiable dans la phase de modélisation. L'ensemble de données d'apprentissage a été utilisé pour estimer les paramètres des modèles, tandis que l'ensemble de données de validation a été conçu pour mesurer la capacité des modèles de classification à extrapoler une nouvelle étiquette pour une nouvelle observation qui n'est pas incluse dans l'ensemble de données d'apprentissage; on peut ainsi savoir si l'algorithme est fiable ou non pour produire une prédiction précise. La proportion de l'ensemble de données d'apprentissage est plus grande que celle de l'ensemble de données de test en raison de la construction d'un modèle de classification à faible variance. Dans cette étude, l'ensemble de données a été divisé en deux parties : 70 % pour l'ensemble de données d'apprentissage et 30 % pour l'ensemble de données de validation.

Optimisation des hyperparamètres

On appelle hyper-paramètres les paramètres de l'algorithme fixés avant la phase d'apprentissage. Par opposition, les valeurs des autres paramètres sont déterminées selon les résultats de l'apprentissage. Une fois les hyper-paramètres déterminés, l'algorithme fixe les autres paramètres optimaux à partir des données d'apprentissage.

La technique traditionnelle pour optimiser les hyperparamètres est la méthode de Grid Search, qui consiste à tester toutes les combinaisons possibles sur des ensembles hyper-paramètres fournis.

L'algorithme de Grid Search doit être guidé par l'évolution d'une métrique d'erreur, mesurée généralement par validation croisée sur l'algorithme d'apprentissage. C'est typiquement ce que nous avons réalisé dans le cadre de cette étude en utilisant la métrique du score de Fisher.

Afin d'optimiser les hyperparamètres des modèles (paramètre de pénalisation pour le GLM et le xgboost et d'autres hyperparamètres évoqués dans le tableau IV.2 pour les arbres CART), nous avons procédé par cross validation [13] en nous focalisant sur une optimisation du score de Fisher connu comme une moyenne harmonique de la précision et du rappel. La méthode de validation croisée utilisée est la k -fold cross validation. On rappelle qu'elle consiste à diviser l'ensemble des données d'apprentissage en k partitions distinctes de tailles égales pour utiliser $k - 1$ partitions pour l'apprentissage et la dernière pour évaluer la performance du modèle, en répétant ce processus k fois.

IV.5.3 Évaluation des modèles

Les modèles de classification ont été évalués en calculant la précision, la sensibilité, la spécificité, l'AUC et l'Accuracy Ratio (AR).

| Classe réelle | Classe prédite | |
|---------------|---------------------|---------------------|
| | Positive | Negative |
| Positive | True Positive (TP) | False Negative (FN) |
| Négatif | False Positive (FP) | True Negative (TN) |

TABLE IV.4 – matrice de confusion.

Les résultats de la matrice de confusion sont utilisés pour calculer la précision, la sensibilité (taux de vrais positifs) et la précisions de la prédiction résultant du modèle de classification avec les formulations indiquées dans l'équation ci-dessous. (taux de vrais positifs) et la spécificité (taux de vrais négatifs) de la prédiction résultant du modèle de classification avec les formulations indiquées dans l'équation ci-dessous.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

$$\text{Sensibilité (ou rappel)} = \frac{TP}{TP + FN}$$

$$\text{Spécificité} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

Le score du modèle est mesuré par la valeur de l'aire sous la courbe (AUC) de la fonction ROC. La qualité de la classification peut être déduite avec les formules et tableau de catégorisation indiqués ci-dessous [Logistique2017].

| AUC | AR | Description |
|-------------|-----------|-------------------------------|
| 0.90-1.00 | 0.80-1.00 | Discrimination exceptionnelle |
| 0,80-0,90 | 0.60-0.80 | Discrimination excellente |
| 0,70-0,80 | 0.40-0.60 | Discrimination acceptable |
| 0.60-0.70 | 0.20-0.40 | Mauvaise discrimination |
| 0,50 - 0,60 | 0-0.20 | Échec |

TABLE IV.5 – Catégorisation des valeurs de l'aire sous la courbe

L'échantillon de test de (30%) tiré de façon stratifiée par rapport à la variable d'intérêt permet d'obtenir les scores finaux afin de comparer les modèles.

| | accuracy | precision | recall | roc | AR |
|---------|----------|-----------|--------|-------|-------|
| logit | 0.656 | 0.652 | 0.665 | 0.723 | 0,446 |
| CART | 0.633 | 0.638 | 0.648 | 0.700 | 0,4 |
| xgboost | 0.685 | 0.666 | 0.674 | 0.747 | 0,494 |

TABLE IV.6 – Scores obtenus par modèle

Avant l'analyse des scores du modèle présenté dans le tableau IV.6, il convient de rappeler qu'il est en effet difficile de systématiser un processus de décision complexe qui recoupe un si grand nombre d'expertise. Il peut être noté que les informations prises en compte dans cette étude n'incluent aucune donnée pour comprendre le comportement des clients. De même des informations en lien avec le risque juridique, le risque opérationnel ou encore le risque environnemental pourraient être considérées comme non négligeables mais n'ont pas été observées dans cette étude. Toutefois le modèle logit est globalement significatif et les scores présentés indiquent que nos modèles ont une performance tout au moins moyenne.

| | rarorc | logit | CART | xgboost |
|----------------------------|--------|-------|-------|---------|
| FP : Asset Leasing | 0.238 | 0.346 | 0.398 | 0.454 |
| FP : Commercial | -0.098 | 0.105 | 0.124 | 0.210 |
| FP : Commodities Based | 0.042 | 0.563 | 0.522 | 0.769 |
| FP : Export | 0.119 | 0.379 | 0.349 | 0.440 |
| FP : Project | 0.315 | 0.424 | 0.381 | 0.533 |
| FP : Real Estate | 0.191 | 0.674 | 0.657 | 0.723 |
| FP : business acquisitions | -0.075 | 0.290 | 0.166 | 0.333 |

TABLE IV.7 – Scores (Accuray Ratio) obtenus sur différents objectifs de financement

Le tableau A.1 présente les scores (AR) pour chaque objectif de financement. Pour tous les types de financement, la prise en compte de l'ensemble des informations disponibles sur les financements permet d'améliorer les scores de classification sur l'approbation. Cette amélioration est particulièrement remarquable sur les financements d'actifs (immobiliers et autres), les financements de projet et dans une moindre mesure les financements d'exportations. Par contre les informations disponibles dans cette étude ne permettent pas une si bonne classification sur les autres périmètres.

IV.5.4 Résultats du modèle logistique

Le test de rapport de vraisemblance permet d'évaluer l'hypothèse de significativité jointe de tous les paramètres associés aux modalités d'un facteur dans le modèle logit.

| | chi 2 score | pvalue |
|---------------------------|-------------|-------------|
| Financing purpose | 23.21 | 3.70 e – 60 |
| syndication | 18.06 | 5.44 e – 36 |
| country | 12.56 | 8.05 e – 38 |
| rating | 6.34 | 2.74 e – 12 |
| rarorc | 4.04 | 9.57 e – 08 |
| synthetic lgb | 3.78 | 1.38 e – 04 |
| cost of liquidity | 0.65 | 0.11 |
| 1 ^{rst} yearNBI% | 0.25 | 0.33 |
| Transaction Type | 0.21 | 0.37 |

TABLE IV.8 – Test de significativité sur les variables

Il ressort des résultats de ce test au tableau IV.8 que l'objectif du financement, le type de syndication, le pays de localisation des actifs financés, le rating de l'obligor, le rarorc, le type de garantie, et la lgd sont des facteurs qui expliquent de façon significative la validation d'une proposition de deal.

Le graphique IV.3 des odds ratio permet d'analyser l'influence des différentes variables dans la décision d'approbation des deals. Seules les modalités présentant un impact significativement non nul sont considérées. Toutes les autres modalités ainsi que celles mises en référence sont considérées comme associées à un coefficient nul.

Il ressort du modèle logit que :

- Pour les mêmes caractéristiques (rating, type de syndication, pays, ...) un deal avec le niveau de rentabilité de plus de 0.5% aura une vraisemblance d'approbation plus importante que les deals de niveau de rentabilité inférieur à 0.5%. On notera par ailleurs que le niveau de rentabilité ne différencie pas les deals perdus de ceux abandonnées (voir graphique ?? en annexe B.3).
- Le risque de voir une proposition de deal non approuvée croit avec le niveau d'engagement de la banque (par rapport au type de syndication) sur deal. Un deal syndiqué en best effort ou en club aura une probabilité d'approbation plus forte qu'un deal réalisé en bilatéral.
- Ce modèle permet de noter une plus faible probabilité d'approbation lorsque le deal est soumis pour un financement dans le secteur real estate. Le graphique B.3 en annexe permet de remarquer que les lignes de métier relatives aux flux sur les matières premières sont particulièrement concernées par la perte.
- Présenté dans certains pays (particulièrement les USA et dans une moindre mesure les pays de l'UE et d'ASIE), un deal aura une probabilité d'approbation plus forte (contrairement aux deals présentés en UK Ou en Afrique).



FIGURE IV.3 – Coefficients de la régression logistique approved vs abandon/lost
 Aide à l'interprétation : Les valeurs des odds ratio (e^{β^j} où β^j représente la variable j) significatifs associés aux différentes variables de l'étude sont affichés. L'intervalle de confiance à 5% sur ces valeurs est ajoutée. Les valeurs de référence ou les odds ratio non significatifs sont également représentés à avec leurs valeur à 1.

- Par rapport au rating des obligor, un financement pour un obligor de bon rating inférieur à 3 (ie AAA, AA, A) ou de rating 4 (BBB) a une vraisemblance pour l'approbation plus importante. Le graphique B.3 en annexe permet de noter que lorsque ces financements ne sont pas approuvés ils sont plutôt perdus qu'abandonnés.
- Notons par ailleurs qu'avec la figure B.3 en annexe, les informations obtenues dans le cadre de cette étude ne permettent véritablement de mettre en exergue les facteurs explicatifs de la perte des deals compte tenu de la non significativité de la plupart des facteurs).

IV.5.5 Résultats de l'algorithme xgboost

IV.5.5.1 Importance des facteurs explicatifs

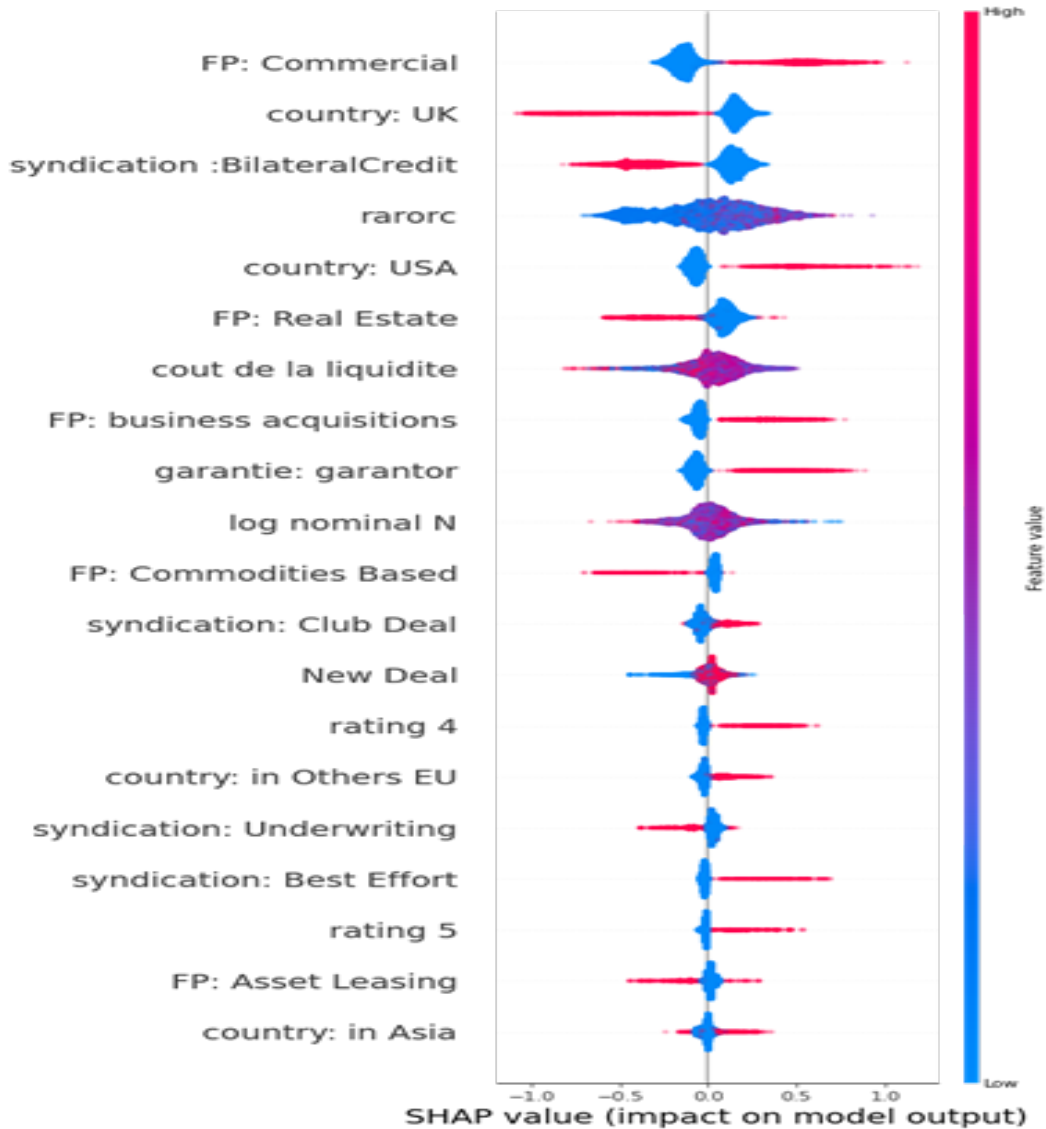


FIGURE IV.4 – Importance par variable

Aide à l'interprétation : Ce graphique présente le nuage de point des valeurs SHAP obtenues sur chaque variable. En rouge les valeurs élevées de la variable et en bleu les valeurs faibles de la variable. Sur l'axe des abscisse on retrouve les valeurs SHAP obtenues en chaque point. Le shap positif signifie que la variable contribue à l'approbation. Une variable dont les grandes valeurs (en rouge) contribue à l'approbation auront donc sur ce graphe des shaps positif. Quand la variable contribue à l'abandon/perte ces shaps sont négatifs

Les figure IV.4 et IV.5 permet d'analyser l'influence de chaque variable sur la prédiction du modèle xgboost. Il ressort des résultats de ce modèle que les financements pour un objectif commercial se distingue particulièrement par une approbation plus importante, alors que les financement en bilateral credit ou en UK vont se caractériser par une non approbation plus importante.

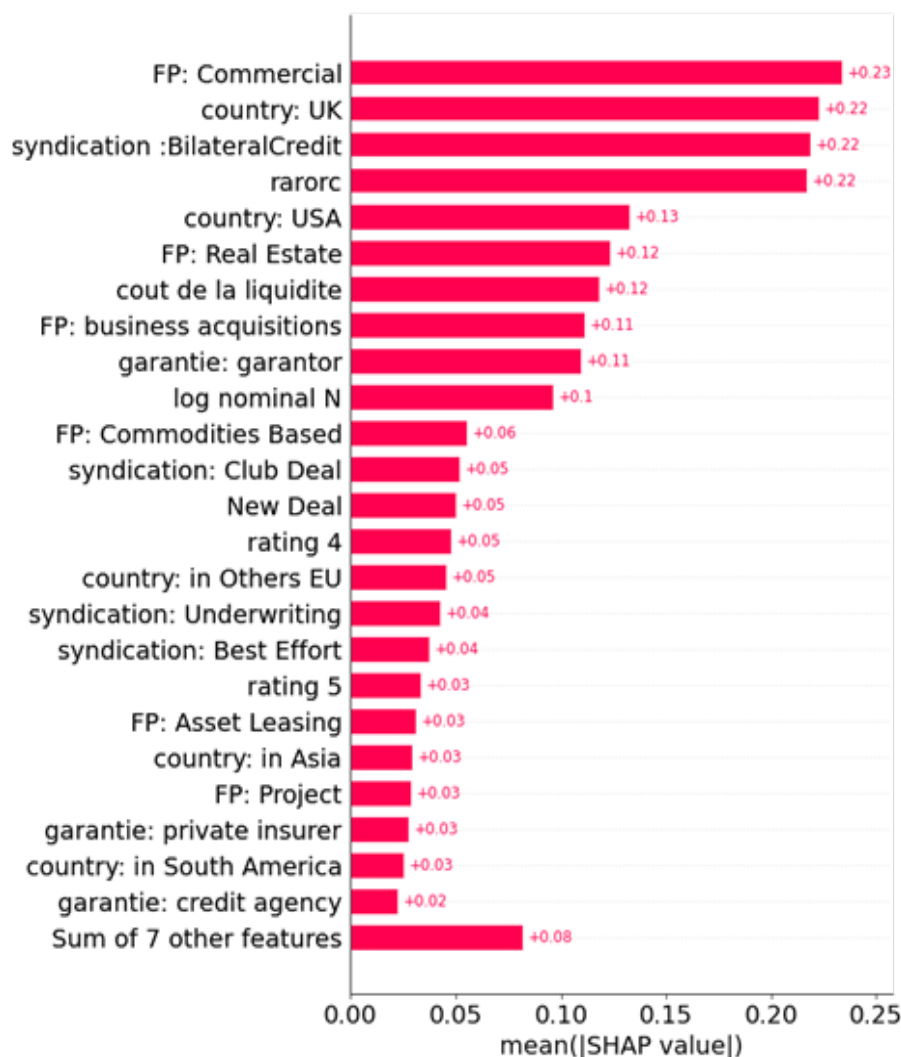


FIGURE IV.5 – Nuage des valeur SHAP par variable

Ce graphique représente l'influence totale (moyenne des valeurs absolues des shap sur la base d'apprentissage) pour chaque variable.

Par rapport au rarorc on note, en interprétant la figure IV.5, que le fait d'avoir un bon rarorc améliore la probabilité d'approbation du deal, cependant les deals approuvés n'ont pas tous des rarorc élevé. Le graphique sur les interactions IV.6 révèle que l'impact du rarorc est différencié :

- selon que le deal est financé pour un objectif commercial, rarorc faible mais shap positif (voir graphique B.2). Cette interaction renseigne que sur ces financements le rarorc bien que faible ne constitue pas un frein à l'approbation.
- dans une moindre mesure selon que le deal est syndiqué en club ou réalisé en bilatéral.

Concernant les interactions, on peut noter la particularité des deals réalisés pour des financements immobiliers en UK. Cette interaction signalée par le modèle et qui correspond à une faible approbation des deals sur ce segment pourrait traduire des tensions commerciales. Des interactions sont également signalées entre le coût de la liquidité et les deals réalisés en bilatéral. Sur les deals n'impliquant aucune syndication, le coût de la liquidité apparaît plus important

dans le processus d’approbation.

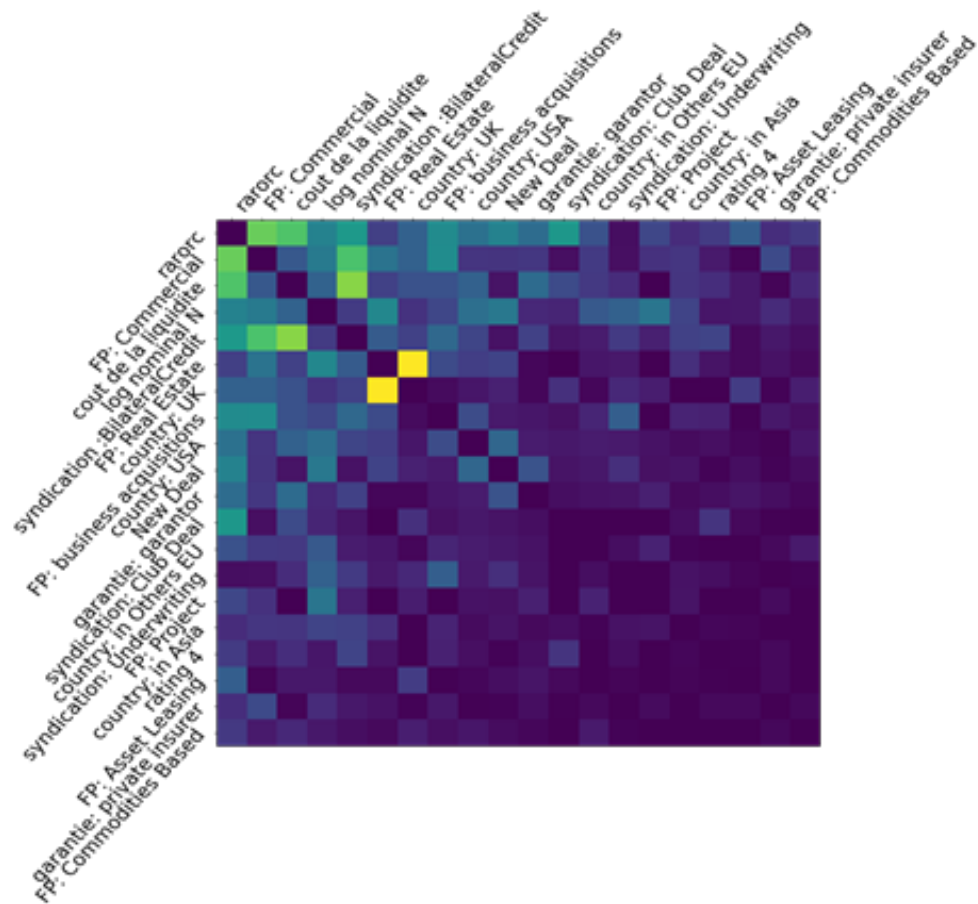


FIGURE IV.6 – Importance des interactions par variables

Aide à l’interprétation : Il s’agit d’une moyenne de la valeur absolue du shap d’interaction pour différents couples de variables et sur l’ensemble des points de notre base d’apprentissage. Les couleurs les plus vives traduisent des interactions plus importantes

IV.6 Conclusion partielle

Ce chapitre a permis d’établir que la rentabilité ajustée du risque de crédit est certes, un indicateur important dans le processus d’approbation des deals, mais loin d’être le seul. La structure du marché sur certains types de financement peut pousser à être moins exigeant sur le niveau de rentabilité. La taille des marchés primaire et secondaire et la force commerciale de l’institution financière peuvent constituer des éléments d’explication pour la bonne performance (probabilité d’approbation importante, probabilité de perte faible) sur certains types de financement (ou pays) et une mauvaise performance sur d’autres types de financement (ou pays).

Conclusion

L'approche RARORC d'évaluation de la performance ajustée du risque présente de nombreux avantages. L'indicateur rarorc est facile à calculer, tient compte des cash-flow internes et externes sur toute la durée de vie des financements, est compatible à la diversité des financements réalisés. Cet indicateur permet de s'assurer que les fonds propres sont correctement rentabilisés à la maille de chaque transaction mais aussi par rapport à une ligne de métier.

Sur un périmètre de financement structuré, la règle de l'activité de financement consiste à pouvoir réduire ou transférer le risque de crédit. Suivant cette règle, la LGD et par suite le capital réglementaire consommé peut être faible sur certaines transactions. Le rarorc calculé comme rentabilité sur ces transactions prendra donc une valeur anormalement élevée. On aboutit à des rentabilités différentes en niveaux et en volatilité sur différents centres d'intérêt. Il est de ce fait inapproprié de définir un seuil unique pour toutes les transactions et de piloter la rentabilité sur toutes les ligne de métier à partir de cet indicateur. Cependant, ce seuil unique est plus adapté lorsqu'il est fixé par rapport à la ligne de métier.

De plus, l'approbation des deals ne peut être de façon exclusive associée à la bonne rentabilité des transactions. Il peut être noté une importance significative de la prise en compte de l'environnement concurrentiel et du besoin de formuler assez souvent des offres commerciales. Le facteur commercial est particulièrement important sur certains types de financement et sur certains pays. Les niveaux d'approbation différenciés selon ces pays peuvent aussi refléter des pratiques opérationnelles différentes. La forme de syndication est aussi déterminante dans la décision d'approbation. Les formes de syndications impliquant un engagement limité (best effort et club deals) de la banque sont les plus susceptibles de conduire à une approbation de la demande de financement.

Cette étude permet d'exposer les limites de l'approche d'évaluation de la rentabilité par le rarorc sur un portefeuille de financement structuré et de présenter les facteurs déterminants dans le processus d'approbations. Cependant elle présente certaines limites. L'évaluation de l'impact des migrations de rating sur le portefeuille ou de l'impact des clôtures anticipées des deals sont des exemples de problème à traiter en complément aux réflexions menées dans cette étude.

Table des figures

| | | |
|-------|---|----|
| I.1 | Distribution de perte sur un portefeuille de crédit | 7 |
| II.1 | Répartition des deals étudiés par type de syndication | 14 |
| II.2 | Répartition des deals selon l'objectif de financement | 15 |
| III.1 | RARORC et approbation des deals par objectifs de financement | 21 |
| III.2 | Rarorc et approbation des deals par type de syndication | 23 |
| III.3 | Exemple de courbe ROC | 26 |
| III.4 | Exemple de courbe CAP | 26 |
| III.5 | Courbe de lorentz sur les financements de de projets | 27 |
| III.6 | Courbe sur les financements en lien avec l'immobilier | 27 |
| III.7 | Courbe sur les financements d'actifs | 27 |
| III.8 | Courbe sur les financements d'exploitation de matières premières et d'exportation | 27 |
| III.9 | Courbe sur les financements d'opérations d'investissement en capital | 28 |
| IV.1 | Exemple d'arbre CART | 41 |
| IV.2 | Pavage didactique de l'espace correspondant à un arbre CART | 41 |
| IV.3 | Coefficients de la régression logistique approved vs abandon/lost | 56 |
| IV.4 | Importance par variable | 57 |
| IV.5 | Nuage des valeur SHAP par variable | 58 |
| IV.6 | Importance des interactions par variables | 59 |
| A.1 | Approbation des deals par objectif de financement | b |
| A.2 | Approbation des deals par type de syndication | c |
| A.3 | Répartition des deals par type pays de localisation des actifs | c |
| A.4 | Approbation des deals par type pays de localisation des actifs | d |
| B.1 | Deux premiers niveaux du meilleur arbre de classification | f |
| B.2 | Valeur SHAP pour le rarorc sur la base d'apprentissage | g |
| B.3 | Interprétation de la régression logistique perte vs abandon | h |

Liste des tableaux

| | | |
|-------|---|----|
| II.1 | Distribution de la variable d'intérêt | 13 |
| III.1 | Matrice de confusion | 26 |
| III.2 | AR obtenu avec le rarorc | 28 |
| III.3 | Score de classification avec le seuil unique fixé sur l'ensemble du périmètre | 28 |
| III.4 | Score de classification avec le meilleur seuil | 28 |
| IV.1 | Liens canoniques de lois usuelles | 34 |
| IV.2 | Exemple de pénalisation par famille d'algorithme | 40 |
| IV.3 | variables sélectionnées | 51 |
| IV.4 | matrice de confusion. | 53 |
| IV.5 | Catégorisation des valeurs de l'aire sous la courbe | 53 |
| IV.6 | Scores obtenus par modèle | 54 |
| IV.7 | Scores (Accuray Ratio) obtenus sur différents objectifs de financement | 54 |
| IV.8 | Test de significativité sur les variables | 55 |
| A.1 | Corrélation des variables avec le rarorc | d |
| A.2 | Association | e |

Bibliographie

- [1] Arne BENZIN, Stefan TRÜCK et Svetlozar T. RACHEV. « Approaches to Credit Risk in the New Basel Capital Accord ». In : (2003). Sous la dir. de Georg BOL et al., p. 1-33.
- [2] Patrick BREHENY et Jian HUANG. « Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors ». In : *Statistics and Computing* 25.2 (nov. 2013). ISSN : 1573-1375.
- [3] Leo BREIMAN. « Bagging Predictors ». In : (1994).
- [4] Tianqi CHEN et Carlos GUESTRIN. « XGBoost : A scalable tree boosting system ». In : 13-17-August-2016 (2016).
- [5] Michel CROUHY, Stuart TURNBULL et Lee WAKEMAN. « Measuring risk-adjusted performance ». In : 2 (1999).
- [6] Julien DHIMA. « Evolution des méthodes de gestion des risques dans les banques sous la réglementation de Bale III : une étude sur les stress tests macro-prudentiels en Europe Julien Dhima To cite this version : HAL Id : tel-02440557 ». In : (2020).
- [7] Bernd ENGELMANN, Evelyn HAYDEN et Dirk TASCHE. « Measuring the Discriminative Power of Rating Systems Bernd Engelmann Dirk Tasche ». In : (2003).
- [8] Michael B. GORDY. « A risk-factor model foundation for ratings-based bank capital rules ». In : (2003).
- [9] Scott M. LUNDBERG et Su In LEE. « A unified approach to interpreting model predictions ». In : *Advances in Neural Information Processing Systems* 2017-Decem (2017). ISSN : 10495258.
- [10] Alistair MILNE et Mario ONORATO. « Risk-Adjusted Measures of Value Creation in Financial Institutions ». In : 18 (2012).
- [11] Peter MIU et al. « Determining Hurdle Rate and Capital Allocation in Credit Portfolio Management ». In : *Journal of Financial Services Research* 50 (mai 2015).
- [12] J. A. NELDER et R. W. M. WEDDERBURN. « Generalized Linear Models ». In : *Journal of the Royal Statistical Society. Series A (General)* 135.3 (1972), p. 370-384. ISSN : 00359238.

- [13] Sebastian RASCHKA. « Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning ». In : (2018).
- [14] Marco Túlio RIBEIRO, Sameer SINGH et Carlos GUESTRIN. « "Why Should I Trust You ?" : Explaining the Predictions of Any Classifier ». In : (2016).
- [15] Daniel J. STEKHOVEN et Peter BÜHLMANN. « Missforest-Non-parametric missing value imputation for mixed-type data ». In : 28 (2012).

Annexe A

Statistiques pour la présélection des variables

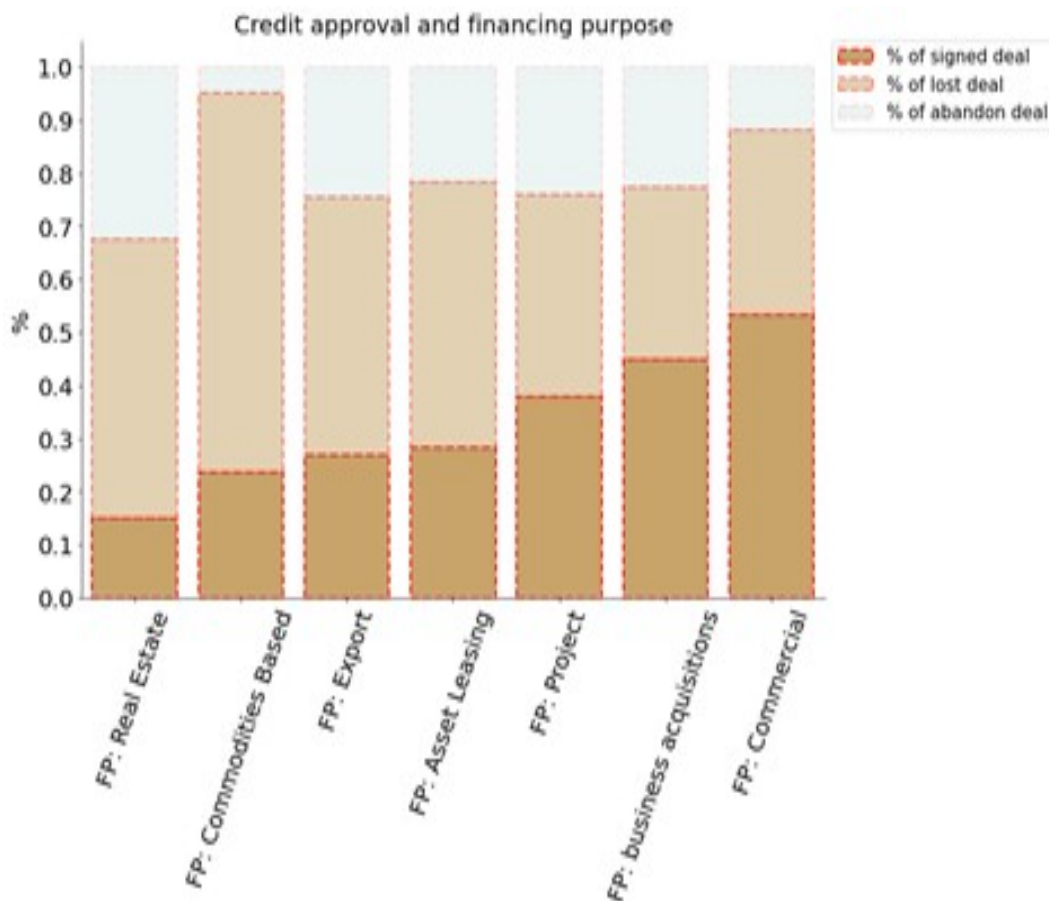


FIGURE A.1 – Approbation des deals par objectif de financement

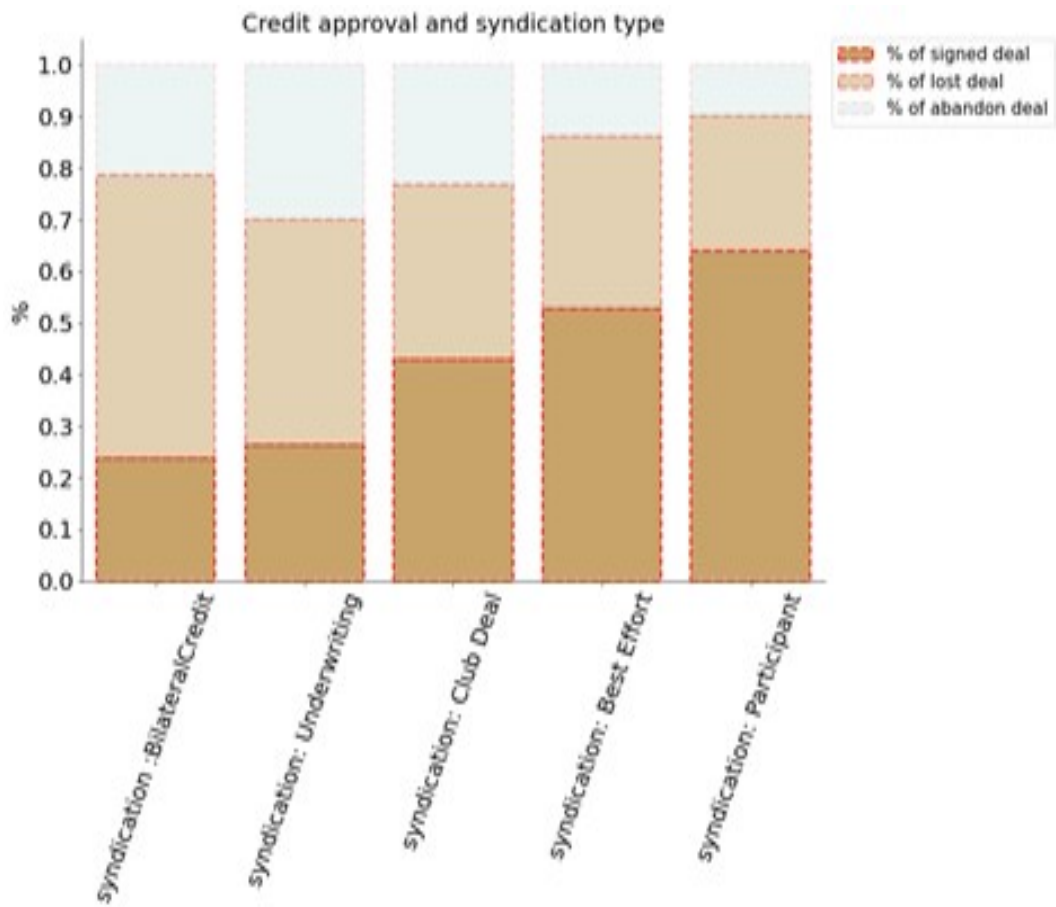


FIGURE A.2 – Approbation des deals par type de syndication

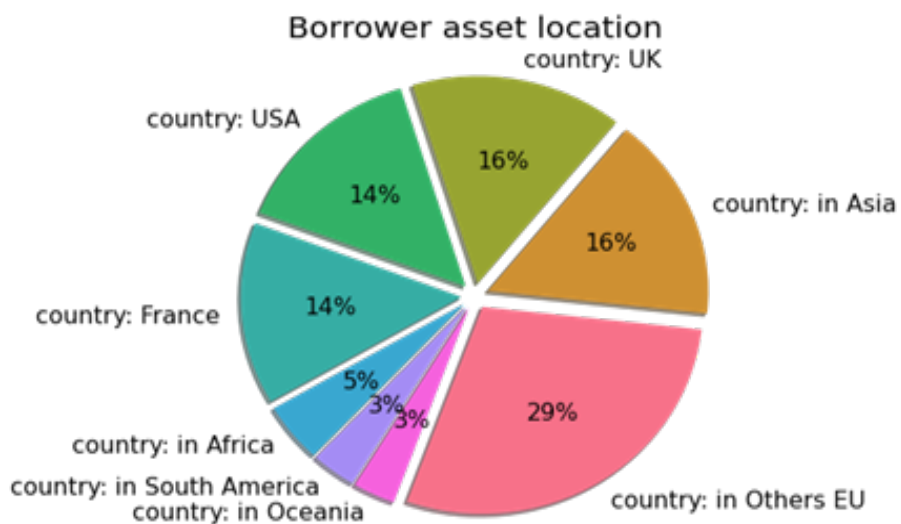


FIGURE A.3 – Répartition des deals par type pays de localisation des actifs

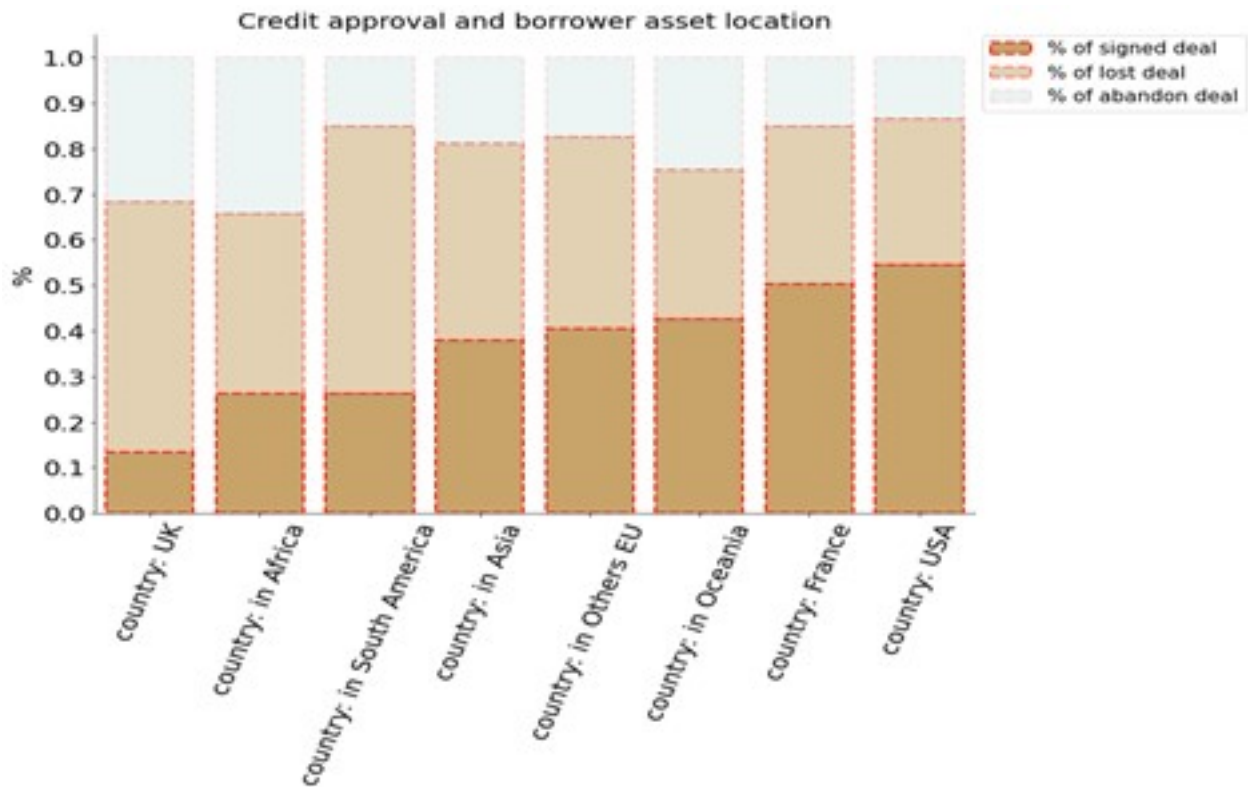


FIGURE A.4 – Approbation des deals par type pays de localisation des actifs

| label | r pearson | thau kendall | rho spearman | p value spearman |
|-------------------------|-----------|--------------|--------------|------------------|
| synthetic lgd | -0.359 | -0.286 | -0.379 | 0 |
| cost of liquidity | -0.124 | -0.223 | -0.312 | 0 |
| share of initial amount | -0.014 | -0.087 | -0.129 | 0 |
| global initial amount | -0.154 | -0.043 | -0.064 | 0 |
| expected nbi | -0.001 | 0.008 | 0.006 | 0.607 |
| 1st year nbi (%) | 0.008 | 0.041 | 0.06 | 0 |
| duration | 0.143 | 0.107 | 0.145 | 0 |
| 1st year RWA | 0.061 | 0.146 | 0.209 | 0 |
| rarorc | 1 | 1 | 1 | 0 |

TABLE A.1 – Corrélation des variables avec le rarorc

| | p value | v cramer |
|-----------------------|---------|----------|
| business line | 0 | 0.277 |
| produ Revolving | 0 | 0.243 |
| syndication type | 0 | 0.227 |
| financing purpose | 0 | 0.224 |
| transaction type | 0 | 0.224 |
| country recoded | 0 | 0.2 |
| rating fc | 0 | 0.116 |
| total amount currency | 0 | 0.088 |
| guarant | 0 | 0.08 |
| expected signing year | 0 | 0.059 |
| expected nbi currency | 0.078 | 0.034 |

TABLE A.2 – Association

Annexe B

Résultats complémentaires sur l'influence des variables

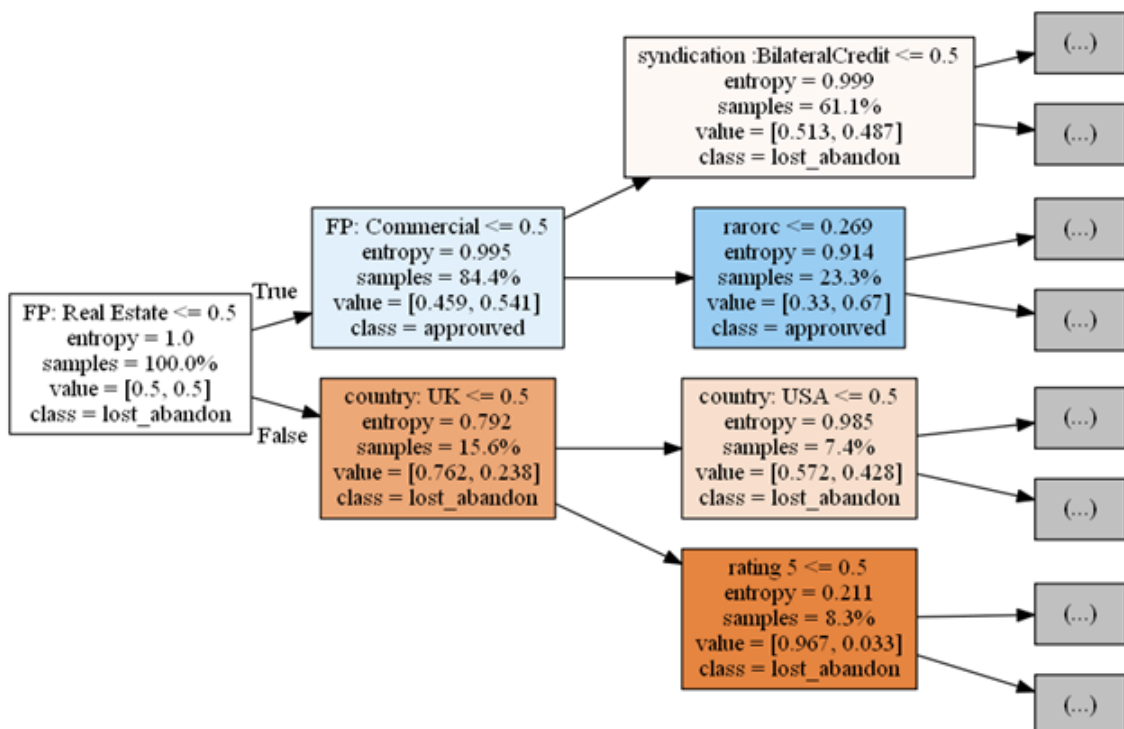


FIGURE B.1 – Deux premiers niveaux du meilleur arbre de classification

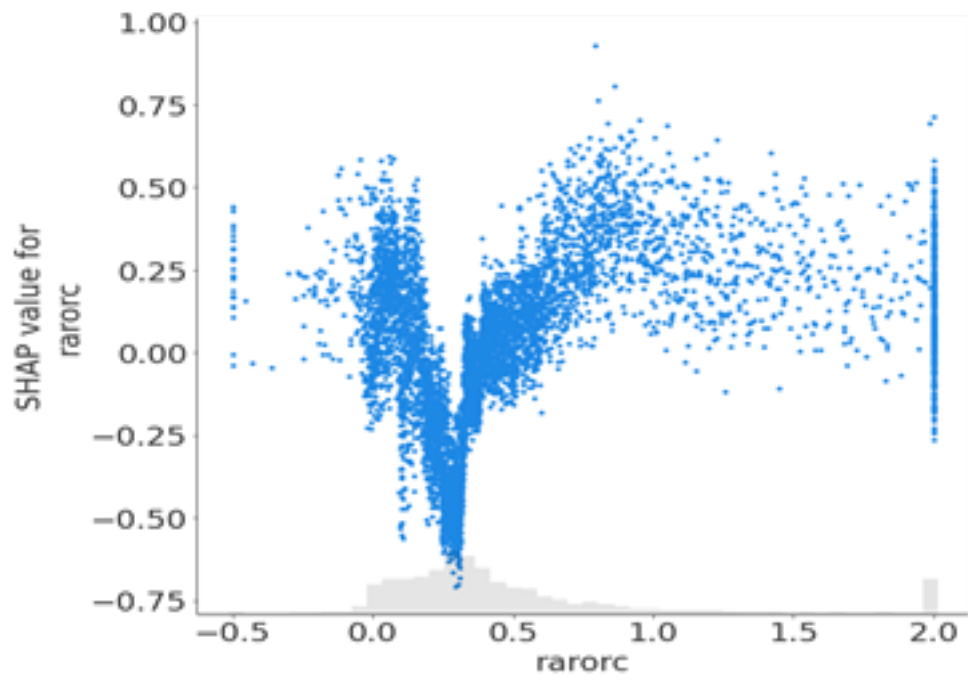


FIGURE B.2 – Valeur SHAP pour le rarorc sur la base d'apprentissage

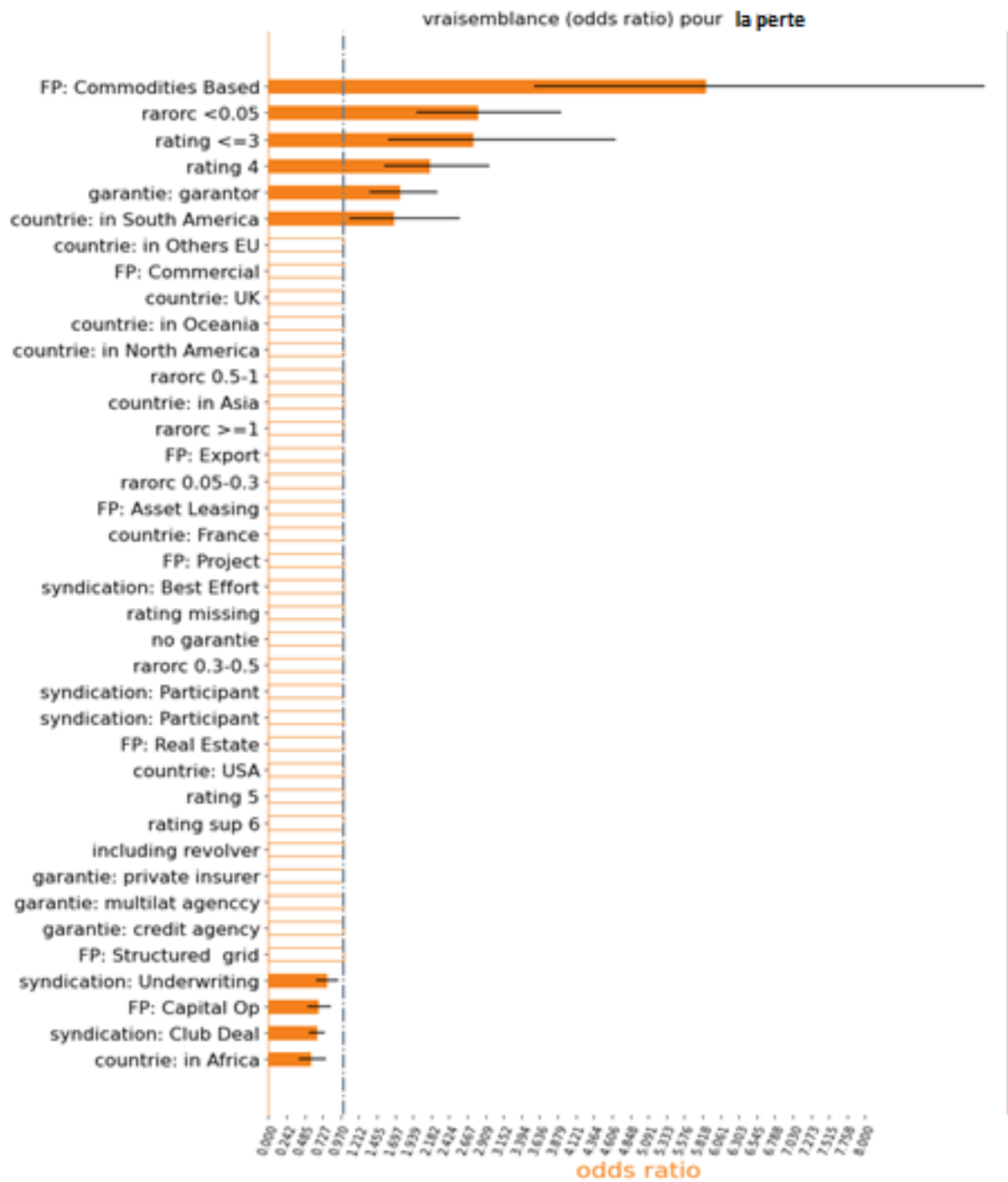


FIGURE B.3 – Interprétation de la régression logistique perte vs abandon