



معهد التمويل والتنمية للمغرب العربي
INSTITUT DE FINANCEMENT DU DÉVELOPPEMENT DU MAGHREB ARABE



Institut De Financement Du Développement Du Maghreb Arabe

(36^{ème} Promotion)

مذكرة

Mémoire en vue de l'obtention du diplôme de l'IFID

*La prévention de la défaillance des PME par la
méthode scoring*

مذكرة

Elaboré par :

OURABIA Selma

Sous la direction de :

Monsieur Mohammed EL MONCER

Année universitaire 2017-2018

INTRODUCTION GENERALE

« Le marché du crédit est l'un des premiers marchés financiers mondiaux, bien plus important que le marché des actions. Il comprend l'ensemble des crédits directs consentis par les banques et les investisseurs, les marchés obligataires classiques et les expositions au risque de contrepartie générées par les transactions sur les produits dérivés.

Le risque de crédit est le risque de perte sur une créance ou celui d'un débiteur qui n'honore pas sa dette à échéance. Il dépend de trois paramètres : le montant de la créance, la probabilité de défaut et la part de non-recouvrement de la créance en cas de défaut.

Les réglementations prudentielles imposent aux acteurs de marché des contraintes strictes dans le pilotage de leurs risques et l'allocation des fonds propres. Ainsi, l'évaluation du risque de crédit est-elle une problématique centrale des institutions financières et des investisseurs sur le marché de la dette qui doivent analyser le risque individuel de chacun de leurs clients et le risque global de leur portefeuille de crédits »¹.

Le risque de crédit, pratiquement, ne peut être totalement éliminé mais peut être atténué s'il est bien géré et analysé par les institutions financières. L'objectif de la banque alors n'est pas d'éliminer ce risque mais de trouver la meilleure combinaison « risque – rendement ». Dans cette optique, les banques ont été appelées à développer leurs mesures de gestion du risque bancaire, particulièrement, le risque de crédit et renforcer leurs outils d'analyse permettant de détecter correctement les divers facteurs influençant et aggravant ce risque.

¹ Philippe THOMAS, « analyse du risque de crédit », 2eme édition, revue banque, 2016.

Outre la gestion classique mise en œuvre par les banques qui se base essentiellement sur le diagnostic financier et l'approche par les ratios et l'exigence des garanties bancaires, et vu les failles qu'a connu ce type de gestion, les banques tentent d'une manière continue à mettre en œuvre d'autres méthodes de gestion se basant sur des modèles mathématiques sophistiqués adaptés proprement à la stratégie de chaque banque.

Parmi ces outils, la méthode du crédit scoring présente l'une des plus connues méthodes d'appréciation du risque de contrepartie permettant ainsi à la banque de créer un système de notation propre aux exigences de la banque.

L'objectif de notre mémoire alors, est de présenter un outil moderne pour la banque capable de faire face au risque de crédit et permet de fournir un critère de décision facilitant la sélection des bons clients et en rejeter les mauvais : *la méthode logistique du crédit scoring*. Cela nous pousse à poser la problématique suivante :

« Comment la banque peut améliorer, à la fois, la prévention de la défaillance de ses crédits et sa qualité de service envers sa clientèle ? »

Afin de répondre à cette problématique, nous allons tout au long de notre travail répondre progressivement à certaines questions auxiliaires :

- Qu'est-ce qu'un modèle scoring ?
- Quels sont les types de modèles utilisés dans le cadre de crédit scoring ?
- Quelle est la méthodologie à suivre pour l'élaboration d'un modèle logistique ?
- Dans quelle mesure la méthode logistique diffère des autres méthodes de scoring, et selon quels critères peut-on juger la meilleure méthode ?

En vue d'avoir une réponse à nos questions et à notre problématique, nous avons divisé notre travail en trois chapitres à savoir deux chapitres théoriques et un dernier chapitre empirique :

- Le chapitre I qui nous permettra d'appréhender des notions de base sur les différents risques bancaires y compris le risque de crédit ainsi que la gestion externe et interne de ce risque
- Le chapitre II à travers lequel on va aborder dans un premier lieu, la technique de scoring, ses différents modèles, et dans un deuxième lieu s'intéresser d'une façon approfondie aux deux méthodes paramétriques : la méthode logistique principalement et la méthode de discrimination linéaire qui fera l'objet de comparaison en terme de performance avec la première méthode
- Le chapitre III à travers lequel on va mettre en œuvre les notions théoriques englobant les différentes démarches nécessaires à la construction d'un modèle scoring logistique tout en comparant sa performance avec celle de l'analyse linéaire.

CHAPITRE I

DEFINITION ET GESTION DU RISQUE DE CREDIT

INTRODUCTION :

La banque joue un rôle primordial dans le développement économique à travers son intermédiation financière entre les agents à besoin de capitaux et ceux à excédent de capitaux. Ce métier, expose la banque à plusieurs risques sur plusieurs niveaux notamment le risque de crédit qu'elle doit y faire face afin d'assurer la continuité de son métier.

Dans ce premier chapitre, nous allons dans un premier temps, présenter et définir le risque de crédit, ses types et expliciter en deuxième lieu les mesures mises en œuvre à l'échelle externe ainsi qu'à l'échelle interne pour la gestion du risque de crédit.

SECTION 1 : DEFINITION ET TYPOLOGIE DU RISQUE DE CREDIT

1. Définitions

1.1. Définition de la notion du risque :

Le petit Robert définit le risque comme un « *Danger éventuel prévisible* », « *Eventualité d'un événement ne dépendant pas exclusivement de la volonté des parties et pouvant causer la perte d'un objet ou tout autre dommage* ».

Le risque peut se définir comme un danger éventuel plus ou moins prévisible. La caractéristique propre du risque est donc l'incertitude temporelle d'un événement ayant une certaine probabilité de survenir et de mettre en difficulté la banque. Le risque inhérent au secteur bancaire se distingue par sa multiplicité et par son caractère multidimensionnel ne pouvant être mesuré par un seul indicateur (Chiappori et Yanelle, 1996).

De nombreuses recherches et discussions, ont donné la description suivante du risque : « *Le risque se rapporte à l'incertitude qui entoure des événements et des résultats futurs. Il est l'expérience de la probabilité et de l'incidence d'un événement susceptible d'influencer l'atteinte des objectifs de l'organisation* »².

En effet, il existe plusieurs types de risque aux quels la banque est confrontée dont le principal étant le risque de crédit.

1.2. Définition du risque de crédit

Le risque de crédit pèse lourd au sein des risques financiers que doit supporter une institution de dépôt. Les provisions pour pertes sur prêts, qui découlent de ces risques, grignotent bien souvent une bonne partie de ses bénéfices³. C'est la forme la plus ancienne du risque sur le marché des capitaux. Il se concrétise par la perte résultant de l'impossibilité pour une contrepartie d'honorer ces engagements.⁴

² Pascal HODONOU DANNON, « *Mécanisme interne de gouvernance bancaire et risques financiers dans la zone UEMOA : Une analyse économétrique par les données de panel* », CERME -RII-ULCO, Université du Bénin, 2009, P4

³ Raymond THEORET, « *Traité de gestion bancaire* ». 2ème édition, Presse de l'Université de Québec, 1999, P233

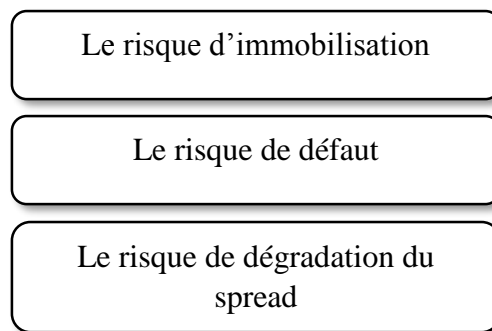
⁴ Mondher BELLALAH, « *Gestion des risques de taux d'intérêt et de change : Théories et exercices corrigés* ». 1ère édition. Edition De Boeck, Bruxelles, 2005, P75

Jorion 2001 définit le risque de crédit comme suit : « *Credit risk is the risk of an economic loss from the failure of counterparty to fulfill its contractual obligations* »⁵« *Credit risk is the risk arising from the uncertainty of an obligor's ability to perform its contractual obligations* »⁶.

De plus, Le risque de crédit est« *la perte potentielle consécutive à l'incapacité par un débiteur d'honorer ses engagements* »⁷. Il s'agit du principal risque pour une banque qui prend aussi diverses autres formes ou appellations : risque de contrepartie (dans les transactions sur les marchés financiers ou interbancaires), risque de faillite ou risque de crédit au sens propre (dans les transactions sur les marchés de crédits).

2. Typologie du risque de crédit

On distingue de façon générale trois formes de risque de crédit :



• **Le risque d'immobilisation** : appelé aussi le risque de liquidité, il est propre à la politique et la stratégie de la banque puisqu'il est lié à la qualité de gestion de la trésorerie de ressources et emplois de la banque. Le banquier doit toujours analyser et apprécier les risques de crédit efficacement avant toute décision d'octroi de crédit.

Par ailleurs cette immobilisation des capitaux se traduit par l'incapacité de la banque à transformer son portefeuille de crédit en liquidité, et ceci pour pouvoir satisfaire la demande des fonds par ses déposants et de poursuivre le financement de sa clientèle. La couverture de ce type de risque se fait par une gestion efficace des ressources de la banque et sa trésorerie dans la cadre de l'ALM.

⁵ Philippe JORION, « *Financial Risk Manager Handbook 2001-2002* », New York. NK: Wile Finance, Risk management Library- Global Association of risk professionals, 2001, P433

⁶ Monetary Authority of Singapore (MAS), « *Credit risk* », February, 2006, P3

⁷ H.JACOB & A.SARDI, « *Management des risques bancaires* ». Ed AFGES. Paris, 2001, P19

• **Le risque de défaut** : C'est le risque majeur, appelé également « risque d'insolvabilité de l'emprunteur », qui est selon Moody's « *tout manquement ou tout retard sur le paiement du principal et/ou des intérêts* ». C'est le risque le plus inquiétant pour la banque. Il est lié à l'état de solvabilité de l'entreprise utilisatrice du crédit. Une entreprise ou tout autre bénéficiaire est considéré solvable s'il est capable d'honorer la totalité de ses engagements y compris les intérêts aux échéances prévues et conformément aux clauses du contrat de crédit.

• **Le risque de dégradation du Spread** : Le Spread de crédit représente la prime de risque qui lui est associée. Sa valeur dépend du niveau de risque encouru (plus le risque est élevé, plus le Spread l'est). Le risque de dégradation du Spread quant à lui est le risque de dégradation de la qualité et la note de la contrepartie et donc l'accroissement de sa probabilité de défaut. Cela induit une hausse de sa prime de risque et donc une baisse de la marge sur intérêts.

SECTION 2 : LA GESTION EXTERNE DU RISQUE DE CREDIT

Pour des raisons de concurrence, la banque est incitée parfois à prendre des risques importants, ce qui pourrait la mettre en péril et même menacer la stabilité de tout le système bancaire à cause des effets de contagion appelé aussi risque systémique. Afin de limiter les conséquences néfastes de la prise de risque et d'assurer la stabilité et la sécurité du système financier, les autorités réglementaires ont mis en disposition des banques des normes prudentielles à respecter en toute circonstance.

1. La réglementation prudentielle bancaire au sens de Bâle I

L'approche du Comité a fixé une globale approximation du risque crédit en fixant un minimum de fonds propres à adosser à un pourcent du portefeuille de crédit de la banque. Ce minimum a été fixé en mettant en place un ratio minimal de 8% de fonds propres par rapport à l'ensemble des crédits accordés par les banques. Les banques sont donc tenues de respecter des normes de gestion destinées à garantir leur liquidité et leur solvabilité à l'égard des déposants, et plus généralement des tiers, ainsi que de préserver l'équilibre de leur structure financière.

1.1.Le ratio de Cooke (ratio de solvabilité) :

Le ratio de Cooke tient ce nom de Peter Cooke, qui avait été un des premiers à proposer la création du Comité de Bâle et fut son premier président. Par ailleurs, le ratio Cooke ou ratio de solvabilité bancaire fut l'une des premières recommandations du comité de Bâle, il fixe une limite à l'encours pondéré des prêts accordés par un établissement financier en fonction de ses capitaux propres. Il représente le rapport entre les fonds propres (capital pur) et quasi fonds

propres (réserves + certaines provisions + titres subordonnés) et l'ensemble des engagements qui sont classés et pondérés selon la catégorie de risque à laquelle appartient le bénéficiaire ou l'actif concerné. Il est donné sous la formule suivante :

$$\frac{\text{fonds propres}(\text{capital} + \text{quasi fonds propres})}{\text{engagements pondérés}} \geq 8\%$$

Tel que les pondérations des engagements sont comme suit :

- 0% : Les créances liquides y compris celles des banques centrales
- 20% : Les créances sur les banques et institutions financières
- 50% : Les crédits hypothécaires pour les logements.
- 100% : Les autres créances

Ce ratio fut aménagé en 1996 afin d'y intégrer la gestion des risques de marché. Les risques de pertes des positions du bilan et du hors bilan à la suite des variations des prix du marché furent retenus pour recouvrir :

- Les risques liés aux taux d'intérêt des instruments et titres de propriété du portefeuille de négociation.
- Le risque de change et le risque sur les produits de base encourus pour l'ensemble de la banque.

Le nouveau ratio devint donc :

$$\frac{\text{fonds propres réglementaires}}{(\text{risque de crédit} + \text{risque de marché})} \geq 8\%$$

2. La réglementation prudentielle au sens de Bâle II

En réponse aux insuffisances de Bale I, le Comité a élaboré et approuvé formellement une nouvelle réglementation relative aux exigences en fonds propres des banques introduisant ainsi de nouvelles méthodologies pour calculer le capital réglementaire nécessaire pour couvrir le risque. En effet, une nouvelle mesure plus pertinente du risque de crédit a été définie prenant en compte la qualité de l'emprunteur à travers le système de notation interne propre à chaque établissement (approche IRB) ainsi que la prise en compte du risque du marché et du risque opérationnel.

L'architecture du dispositif repose sur trois piliers complémentaires :

- L'exigence de fonds propres (ratio de solvabilité McDonough) ;
- La procédure de surveillance prudentielle ;
- La discipline de marché.

❖ **Pilier I : l'exigence de fonds propres**

Ce ratio maintient inchangé à 8% le niveau des fonds propres réglementaires couvrant les risques encourus. En revanche, un calibrage du risque en fonction de sa qualité est exigé.

A cet effet, on introduit la prise en compte des risques opérationnels (fraudes et erreurs) en complément du risque de crédit ou de contrepartie et des risques de marché. Le Nouvel Accord affine donc l'accord de 1988 et impose aux établissements financiers de détenir un niveau de fonds propres adéquat avec les risques encourus.

Cette exigence fait passer d'un ratio Cooke à un ratio McDonough donné par formule suivante :

$$\frac{\text{fonds propres de la banque}}{(\text{risque de crédit} + \text{risque de marché} + \text{risque opérationnel})} \geq 8\%$$

❖ **Pilier II : la procédure de surveillance prudentielle**

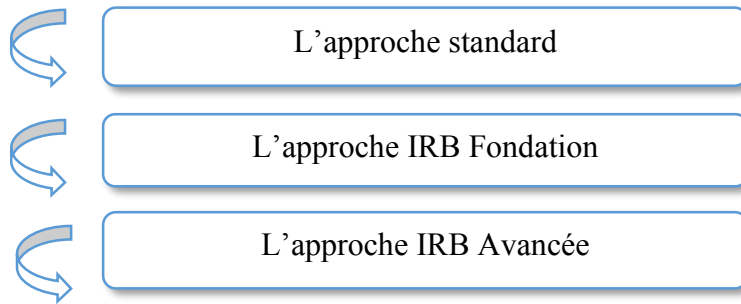
Le pilier II a pour objectif, d'une part, inciter les banques à faire appel à de nouveaux outils et de nouvelles méthodes de gestion de risques et, d'autre part, permettre aux autorités de régulation de majorer les exigences de capital réglementaire en cas de nécessité.

❖ **Pilier III : la discipline de marché**

Le pilier III indique que l'amélioration de la communication financière permet de renforcer la discipline de marché, perçue comme un complément à l'action des autorités de contrôle. L'information sur les actifs, les risques et leur gestion est mise à la disposition du public et les pratiques doivent être transparentes et uniformisées.

3. Les approches proposées pour la mesure du risque crédit :

Trois approches sont proposées pour calculer le risque de crédit :



❖ L'approche standard

La méthode dite standard, consiste à recourir à des analyses effectuées par des tiers, notamment les agences de rating, qui une fois validées par la réglementation prudentielle, serviront de base à la mise en œuvre des exigences de fonds propres dans le cas d'un concours bancaire accordé à l'emprunteur faisant l'objet de l'évaluation. Le recours aux travaux des évaluateurs externes est fondé sur plusieurs arguments logiques :

- Les agences sont mieux dotées des données historiques et importantes sur les grands emprunteurs.
- Ces agences disposant d'un large accès à l'information en provenance des émetteurs ce qui rend l'approche standard plus rationnelle et moins coûteuse que celle consistant à évaluer séparément par chaque établissement prêteur.
- Si applicable, l'approche présente l'avantage de fournir des éléments quantifiés sur le risque des emprunteurs selon des bases comparables.

Les engagements sont répartis en différentes catégories d'actifs, lesquelles sont rangées dans des classes de risque sur la base des notations fournies par les agences de notation externes.

On prend par exemple les notations de l'agence du rating Standars & Poor's :

Tableau 01 : La notation des titres par les principales agences de « Rating »

Standard & poor's	Moody's	commentaire
AAA	Aaa	Excellente qualité
AA	Aa	Très bonne qualité
A	A	Bonne qualité
BBB	Baa	Qualité assez bonne
BB	Ba	Qualité moyenne
B	B	Faible qualité
CCC	Caa	Risque très important
CC	Ca	Très proche de la faillite
Non noté	D	En défaut

Source : http://www.abcbourse.com/apprendre/1_agences_notation.html

❖ **L'approche des notations internes⁸ : Internal based rating(IRB)**

Le principe de l'approche des notations internes se base sur l'appréciation, par les banques elles-mêmes, de leur risque de crédit. Cette appréciation ne permet pas à ces banques de déterminer les propres exigences en fonds propres. Elle classe les expositions en cinq : entreprises, souverains, banques, détail et action.

Les paramètres d'appréciation du risque de crédit sont la probabilité de défaillance (PD), la perte en cas de défaillance (LGD), l'exposition à la défaillance (EAD) et la maturité(M). Toutefois, une fonction de calcul des pondérations est donnée par le Comité de Bâle intégrant tous les paramètres.

L'approche IRB « internal ratings-based approach » porte principalement sur le système de notation interne. L'expression "système de notation " recouvre l'ensemble des processus, méthodes, contrôles ainsi que les systèmes de collecte et d'information qui permettent d'évaluer le risque de crédit, d'attribuer des notations internes et de quantifier les estimations de défaut et de pertes.

⁸ Mme FEKIH Zohra, thèse « *Etude de la relation Crédit – Information* », 2014, P25

Signification des paramètres :

- *La Probabilité de défaut* : La PD est Exprimée en pourcentage, elle correspond à la probabilité qu'une contrepartie soit défaillante sur un horizon de douze mois. Un emprunteur dispose d'une PD unique quels que soient les produits souscrits.
- *La Perte en cas de défaut* : La LGD est Exprimée en pourcentage, elle correspond au taux de perte constaté en cas de défaillance. Elle se définit comme une perte économique, qui se mesure en prenant en compte tous les coûts directs et indirects liés au recouvrement.
- *Exposition en cas de défaut EAD* : elle correspond à l'exposition en cas de défaillance. Elle englobe les encours bilanciels ainsi qu'une quote-part des engagements hors bilan
- *La maturité M* : elle correspond à la durée du crédit. C'est le temps nécessaire à l'emprunteur pour honorer ses engagements.

Les variables définies permettent de quantifier la perte moyenne attendue sur un engagement et à un horizon donné comme suit :

$$\text{Perte en cas de défaut} = \text{EAD} \times \text{PD} \times \text{LGD}$$

Ces paramètres sont ensuite pris en considération dans les fonctions utilisées pour le calcul des actifs pondérés du risque, fonctions prescrites par les dispositions réglementaires et variant selon la catégorie du débiteur.

$$\text{Besoin en fonds propres} = [\sum f(\text{PD}, \text{LGD}, \text{M}) \times \text{EAD}] \times 8 \%$$

✓ **Approche IRB Fondation (FIRB) :**

Les banques déterminent seulement les probabilités de défaut associées à chaque créance tandis que les estimations des autres paramètres sont fournies par les autorités de contrôle prudentiel.

En effet, d'après le comité de Bâle II, la valeur de la LGD est fixée à 50% ou 75% et la maturité n'est pas prise en compte. Dans ce cas, la maturité moyenne des expositions est supposée égale à 3 ans.

✓ **Approche IRB avancée (AIRB) :**

Cette approche diffère sensiblement de la méthode simple. En effet, les valeurs de perte en cas de défaillance ne sont plus fixées par le régulateur, mais sont estimées par la banque. Ensuite, la maturité est explicitement prise en compte.

SECTION 3 : LA GESTION INTERNE DU RISUQUE DE CREDIT

1. La gestion traditionnelle du risque de crédit :

1. Le diagnostic financier :

« L'analyse crédit traditionnelle, telle qu'elle est entreprise par les agences de notation et les investisseurs crédit, s'appuie généralement sur un classement des risques en deux grandes catégories. La première, le risque opérationnel, englobe un éventail de facteurs couvrant généralement le risque industriel (notamment les risques réglementaires), le risque de marché, le risque de gestion, le risque de gouvernance (qui pourrait correspondre à un sous-ensemble du risque de gestion), le risque de concurrence, le risque de substitution, ainsi que d'autres catégories similaires. La seconde, le risque financier, se rapporte plus spécifiquement à des événements ayant des incidences potentielles sur les bilans, les comptes de résultat ou les flux de trésorerie, ou, plus généralement, pouvant affecter la capacité d'une entreprise à servir sa dette et ses autres obligations »⁹. En effet, L'analyse du risque financier est une démarche qui s'appuie sur l'examen critique de l'information comptable et financière fournie par une entreprise dans le but d'apprécier sa performance ainsi que sa solidité financière. Ainsi, avant d'accorder un financement ou un prêt à une entreprise, l'exploitant bancaire examine les états financiers de l'entreprise en question.

Il ne prend sa décision qu'après l'examen de quelques critères financiers relevés de ses états de synthèse. Ces éléments serviront ensuite à calculer quelques indicateurs ou ratios qui contribueront à déceler les forces et les faiblesses d'une entreprise en ce qui concerne son autonomie et son indépendance financière, ses performances, sa solvabilité et sa trésorerie. Le diagnostic financier est parallèlement renforcé par :

- L'étude technique : elle concerne les études afférentes : aux locaux d'exploitation de l'entreprise localisation, superficie couverte et non couverte, capacité d'extension et de stockage, aux matériels de production, aux caractéristiques des produits fabriqués ou vendus.

⁹ BOB BUHR, « Elargir le champ d'analyse de la gestion du risque de crédit », revue économique financière, P 138

- L'étude commerciale : qui confère à la banque le moyen d'apprécier le développement de l'activité d'une affaire et de mieux comprendre les besoins de financement de celle-ci.
- L'étude de l'activité : l'analyste doit choisir les comptes revêtant une importance particulière selon l'information qu'il cherche à savoir. Les principaux comptes que le banquier doit analyser sont :
 - Chiffre d'affaires (CA) ;
 - Valeur ajoutée (VA) ;
 - L'excédent brut d'exploitation (EBE) ;
 - Résultat net de l'exercice (RN) ;
 - Capacité d'autofinancement (CAF) ;
- L'étude de l'équilibre financier : L'équilibre financier de l'entreprise s'apprécie à travers l'étude des agrégats financiers suivants : Le Fonds de Roulement (FR), le Besoin en Fonds de Roulement (BFR) et la Trésorerie Nette.
- Analyse des ratios : Les ratios sont un bon moyen d'évaluer le rendement d'une entreprise et de repérer les problèmes dont elle fait face. Ils permettent de mesurer certains facteurs comme la rentabilité et la solvabilité de l'entreprise. La pertinence d'un diagnostic financier par la méthode des ratios n'est pas fonction du nombre des ratios définis et calculés mais davantage de la qualité d'homogénéisation, de l'indépendance et de la complémentarité des ratios retenus.

2. Les garanties

Les prêts bancaires sont accordés en fonction de la capacité de remboursement des emprunteurs Plus le crédit est long, plus le risque est difficile à évaluer.

Afin de limiter son risque, le banquier exige des garanties. Ces dernières ont un double rôle. Elles entraînent une diminution de la probabilité de défaut de l'emprunteur liée au risque d'aléa moral (l'emprunteur fera tous les efforts pour rembourser la banque afin de ne pas perdre la garantie fournie) et servira de compensation de la perte en cas de réalisation du risque.

❖ **Les garanties personnelles :**

Appelées aussi « Sûreté personnelle ». Ces garanties sont constituées par l'engagement d'une ou plusieurs personnes de rembourser le créancier en cas de défaillance du débiteur principal. Elles se réalisent sous formes juridiques de cautionnement et de l'aval.

- **La caution :**

C'est le contrat par lequel une personne, promet au créancier d'une obligation d'y satisfaire si le débiteur ne la remplit pas lui-même.

Selon l'article 644 du code civil, le cautionnement est défini comme étant « *un contrat par lequel une personne garantit l'exécution d'une Obligation en s'engageant envers les créanciers à satisfaire à cette Obligation, si le débiteur n'y satisfait pas lui-même* »

- Dans le cas de cautionnement, il y a lieu de vérifier la capacité juridique de la personne physique qui représente la caution.
- La caution doit toujours être limitée dans son étendu et sa durée.

La personne qui se propose pour être caution doit avoir la capacité de jouissance qui se définit comme étant l'aptitude d'avoir des droits, des obligations et la capacité d'exercer le pouvoir de mettre en œuvre soit même ses droits et ses obligations.

- **L'aval :**

L'aval est l'engagement cambiaire donné par une personne appelée avaliste ou donneur d'aval, qui garantit l'exécution de l'engagement d'un débiteur de la lettre de change.

❖ **Les garanties réelles :**

« *Dans le cadre des sûretés réelles, le paiement des créances est garanti par des droits permettant à leurs titulaires d'être payé par priorité. La sûreté est dite réelle car elle porte sur un bien meuble ou immeuble donnant aux créanciers des droits particuliers. Elles portent sur les biens de l'emprunteur sont les plus anciennes et les plus appréciées par la banque* »¹⁰.

¹⁰ André NEVEU, « Financer l'agriculture : quels systèmes bancaires pour quelles agricultures ? ». Edition Charles Léopold Mayer, Vanves, 2001, P141

- ***Le nantissement :***

Il peut être défini comme étant un contrat réel de garantie, par lequel le débiteur remet à un créancier pour sûreté de sa dette, la possession effective d'un bien (mobilier) ainsi remis tout en s'assurant que la valeur liquide de la garantie peut couvrir le montant de la dette.

Il existe plusieurs types de nantissement :

- Nantissement de fonds de commerce.
- Nantissement du matériel et outillage.
- Nantissement sur véhicule automobile.

- ***L'hypothèque :***

« *L'hypothèque est un contrat par lequel un débiteur accorde à son créancier un droit sur ses biens immeubles (terrains, bateaux, immeubles...)* »¹¹ Si l'emprunteur ne rembourse pas le prêt, les biens seront saisis et vendus par le prêteur.

L'hypothèque est la sûreté réelle immobilière par excellence. L'hypothèque est une garantie coûteuse comparativement au nantissement. Elle est essentiellement sollicitée en couverture de crédits d'investissements. Mais elle peut être exigée pour couvrir globalement les crédits d'exploitations lorsque le total de ces concours atteint un montant important.

Néanmoins, la banque désormais ne peut compter uniquement sur l'analyse traditionnelle et bilancielle des entreprises afin de gérer le risque de crédit vu la multitude des événements hautement risqués et moins attendus pouvant toucher l'activité de l'entreprise.

Pour cela, la banque doit disposer d'autres moyens plus sophistiqués et alertant au mieux au profit de la banque, le maximum des événements risqués probables auxquels elle pourrait être confrontée.

¹¹ Franck Olivier MEYE, « *Évaluation de la rentabilité des projets d'investissement : méthodologie pratique* », Edition l'Harmattan, Paris, 2007, P89

2. La gestion moderne du risque de crédit :

La mesure du risque de crédit est devenue aujourd'hui l'une des priorités des institutions financières, puisque ce risque de défaut est considéré comme une notion plus complexe qui recouvre des réalités différentes, selon le contexte dans lequel il est évoqué et selon le caractère négociable des engagements considérés.

A cet égard plusieurs méthodes, ou modèles sont mis en place afin d'évaluer le risque lié au contrepartie d'une manière plus représentative, du fait que plusieurs recherches théoriques ainsi que des tests empiriques sont au centre de développements des modèles pour évaluation exacte du risque, parmi les modèles les plus connus on peut citer mis à part la méthode scoring qu'on traitera isolément dans le chapitre suivant : le crédit Metrics, les modèle KMV, le credit risk+.

1. Le Credit Metrics :

« *Le credit Metrics, est considéré comme le premier modèle de portefeuille destinée à évaluer le risque de crédit* »¹², il a été mis en place par J. P. Morgan (1997), qui le considère comme un modèle de référence pour les années avenir .

Ce modèle permet une première appréciation individualisée des grands risques d'un portefeuille de crédit, du fait que les variations liées à la qualité du dossier de crédit dépendent non seulement des événements liés au défaut y compris la faillite, les retards de paiement dus aux difficultés financières mais aussi aux fluctuations (appréciation ou détérioration) de la cote du crédit de la firme.

Ainsi cette méthode se base sur l'estimation de la value at Risk du crédit à partir d'une modélisation du risque de spread associé au rating des actifs d'un portefeuille obtenu sur le marché ou à travers des données statistiques tel que celles publiées par les agences de notation (matrice de transition, taux de défaillance), c'est la méthode la plus opérationnelle.

D'après Gupton (1997), le Credit Metrics comporte certains bénéfices qui consistent à :

- Quantifier le risque de crédit agrégé, d'autre terme les pertes potentielles dues aux événements de crédit.

¹² MORGAN J.P, « *CreditMetrics* », document technique, 1997.

- Identifier les sources de risque et donc mesurer le risque spécifique à chaque situation de crédit.
- Calculer le risque supplémentaire apporté par une position au portefeuille total.
- Identifier le risque maximal acceptable, en d'autre terme mettre en place des limites de risque.
- Déterminer le montant du capital économique que l'institution financière doit conserver en cas des pertes non anticipées.

2. *Le modèle KMV¹³ :*

Afin de mesurer la fréquence de défaut espérée, la compagnie KMV offre une méthode spécialisée qui identifie de façon périodique les probabilités de défaut, ensuite comme méthode de gestion du risque de crédit d'un portefeuille. KMV offre le portfolio Manager qui sert à déterminer les caractéristiques du risque et du rendement des portefeuilles de titre à revenu fixe.

Ces modèles de type KMV, se basent sur des valeurs du marché et non pas des valeurs comptables, elles se procèdent ainsi d'une modélisation prospective de la valeur de la firme, alors que les autres méthodes se contentent d'établir des corrélations statistiques entre les différentes variables, et qui peuvent ne pas persister dans le futur.

Ainsi ces méthodes sont basées sur le modèle de Merton (1974), d'où d'une part ils relient la valeur de marché des capitaux propres de la société à celle de ses actionnaires, d'une autre part ils comparent cette dernière à la valeur de ses passifs. En effet la modélisation stochastique de la probabilité de défaut, s'est développée dans les années 90 par l'agence KMV, qui s'appuie sur des fondements théoriques reconnus, celles proposées par Merton « l'analyse de défaut de la firme ». Selon ce modèle, les actions d'une société endettée sont considérées comme des options d'achats (call) avec un niveau d'exercice égal à la valeur faciale de la dette.

De ce fait le KMV utilise le modèle d'évaluation d'option de black&choles afin d'arriver à des valeurs de marché individuelles des actifs pour chaque emprunteur, les relations entre ces valeurs sont ensuite réparties selon des facteurs spécifiques et communs.

En d'autre terme la modélisation stochastique de KMV, permet de déterminer à tout instant la différence entre la valeur du marché des actifs de l'entreprise notée V_0 , et le montant de la

¹³ I. ZORGUI, « *le risque de crédit* », mémoire, master banque finance, 2006. P 23.

dette à rembourser noté DPT_{Γ} qui représente la distance par rapport au défaut DD. On peut l'exprimer comme suit :

$$\frac{\ln \frac{V_0}{DPT_{\Gamma}} + (\mu - 0.5\sigma^2)}{\sqrt{T}\sigma}$$

Une fois (DD) obtenue, elle va être comparée par KMV avec d'autres entreprises ayant une DD égale et dont on connaît le taux de défaut historique afin d'obtenir PD (ou expected default frequencier pour KMV).

Un obstacle pratique dans l'utilisation de ces modèles se manifeste par la complexité de la structure du capital d'une société, puisqu'elle comprend plusieurs types de dette à échéance différente.

3. *Le modèle CreditRisk+ :*

Ce modèle a été créé par le crédit suisse First Boston (1997) dont son développement a été mené à partir des produits financiers de crédit suisse, il se diffère du Crédit Metrics de J. P. Morgan, puisqu'il utilise une approche actuarielle pour présenter dans les termes de probabilité les pertes de crédits de portefeuille, résultant des défauts de crédit.

Le creditrisk+ représente une approche moderne de mesure et de gestion du risque de crédit en permettant une meilleure compréhension du risque liée au portefeuille d'actifs, et cela en mesurant les pertes espérées et non espérées liées au crédit, ainsi que la détermination du niveau de capital requis afin de supporter le risque de contrepartie, et en essayant d'identifier les actifs qui contribuent le plus au risque du portefeuille.

D'après le crédit suisse First Boston (1997), « *le modèle Creditrisk+ est un modèle statistique de risque de défaut lié au crédit, il considère les taux liés au défaut comme des variables aléatoires continues et incorpore la volatilité des défauts afin de faire face à l'incertitude lié à ce dernier* ».

D'après la même source, « *le Creditrisk+ permet de déterminer la distribution des expositions de crédit au risque entre les différents secteurs, il est plutôt considéré comme le modèle le plus représentatif de risque crédit d'un portefeuille, ainsi la probabilité de défaut peut être estimée à partir des données historiques, ou auprès des institutions statistiques* ».

Le Creditrisk+ se base sur des modèles statistiques, du fait que deux événements qui peuvent exister simultanément pour chaque contrepartie, l'une se présente par défaut avec une probabilité PA, l'autre de non défaut avec une probabilité (1-PA). En effet pour un portefeuille de N engagements, le nombre de défaut sur une période donnée suit une loi de poisson :

$$\frac{e^{-\mu} \mu^n}{n!}$$

« En dépit des avantages du modèles KMV, il présente quelques limites, ils sont les même concernant la méthodologie des crédits Metrics et KMV, qui assument la non prise en compte du risque de marché. Le CreditRisk+ ignore notamment la migration du risque, et ne dépend pas des changements éventuels dans la qualité des crédits, de même pour la variabilité des taux d'intérêt futurs. En d'autre terme les taux de défaut qui sont basés sur des données historiques constantes et négligent les changements que peuvent subir les facteurs liés au marché du crédit (les évolutions et les perturbations économiques (récession/dépression). D'une manière générale ce modèle est caractérisé par la complexité des modélisations et la difficulté à évaluer l'impact des approximations mathématiques, qui rendent délicat l'utilisation du modèle, opacité du calcul, faible capacité à expliquer les résultats. »¹⁴

CONCLUSION :

A travers ce premier chapitre, on a pu constater que le risque de crédit est l'un des risques majeurs auquel l'activité de la banque est confrontée. En effet, ce risque n'est pas uniquement relié à la défaillance de la contrepartie mais aussi à plus d'autres facteurs et risques notamment le risque de marché et le risque opérationnel. Cette liaison entre les risques pousse la banque à renfoncer sa réglementation prudentielle et développer des mesures de gestion de ce dernier autres que les mesures classiques basées sur le diagnostic financier afin d'affiner le mieux possible ses différents risques notamment le risque de crédit.

¹⁴ I. ZORGUI, op.cit, P 25.

CHAPITRE II

LE SCORING COMME UN OUTIL DE PREVENTION DU RISQUE CREDIT

INTRODUCTION :

La gestion du risque de crédit connaît de plus en plus au milieu bancaire une évolution sophistiquée caractérisée par des outils et des méthodes modernes y compris celles abordées dans le premier chapitre, conçues pour prévoir, détecter et apprécier le niveau de risque permettant ainsi de mettre en œuvre des solutions adéquates pour l'atténuer.

En effet, parmi ces outils nous allons mettre l'accent sur la méthode scoring à but préventif du risque de crédit. Etant donné que cette méthode n'est jusqu'à présent pas bien exploitée au sein des banques algériennes, nous allons présenter dans un premier temps les différents types de modèles dans le cadre du crédit scoring pour enfin se focaliser sur l'une d'elles « la méthode logistique » qu'on abordera d'une façon plus approfondie.

SECTION 1 : PRESENTATION DE LA METHODE SCORING

1. Définition et objectif du Scoring

Selon le langage courant, « le terme score peut signifier classement, résultat, marque etc. En statistique, c'est l'idée de « classement » qui est surtout retenue »¹⁵.

D'autres définitions sont retenues : « Le Scoring correspond à une méthode d'analyse financière qui tente à synthétiser un certain nombre de variables sous forme d'un seul indicateur susceptible de distinguer les entreprises saines des entreprises défailtantes. »¹⁶

« Le crédit-Scoring est une technique de sélection automatisée de la clientèle généralement basée sur l'analyse statistique. Il permet de déterminer à partir d'un échantillon d'étude la meilleure classification des clients en fonction de leur profil de risque en attribuant à chaque demandeur de crédit une note inversement proportionnelle à sa probabilité de défaut. Le score est une approche scientifique d'un processus décisionnel. Cette note attribuée à une demande de financement synthétise l'information contenue dans plusieurs ratios financiers et surtout ceux les plus prédictifs de la défaillance et les plus discriminants entre les deux classes des « bons » et « mauvais » clients et ceci en se basant sur une démarche statistique d'agrégation et de traitement des données ».¹⁷

Elle est utilisée généralement pour les crédits de faible montant destinés à des catégories professionnelles bien définies (PME, professions, libérales, agricoles ...).

❖ Objectif du crédit Scoring :

Le crédit-Scoring a pour objectif de déterminer une note, c'est-à-dire une numérisation de la capacité de remboursement d'une entreprise sollicitant un crédit. Le défi qu'essaie de relever la méthode scoring est de détecter les informations caractérisant le crédit expliquant le mieux sa solvabilité.

« La performance d'une fonction score est conditionnée à :

- L'homogénéité de comportement des emprunteurs afin que les critères décisionnels soient valables pour tous.

¹⁵ Hassen MATHLOUTHI, « Cours de méthodes de scoring », université de carthage, P10.

¹⁶ Azzouz ELHAMMA, « la gestion du risque de crédit par la méthode scoring », Rabat, 2009.

¹⁷ www.Détection faillite.fr

- *La comparabilité des risques encourus en choisissant des crédits qui présentent une certaine identité de montant, de durée ou d'objet.*

Ainsi, les méthodes de crédit-Scoring ont pour objet d'élaborer un indicateur permettant de juger rapidement la situation financière d'une entreprise. L'indicateur le plus efficace est celui qui déterminera, avec la meilleure probabilité, le classement d'une firme dans l'une ou l'autre de ces catégories. »¹⁸

❖ **Les paramètres d'un modèle Scoring :**

L'élaboration d'un modèle scoring se base sur le choix et la bonne sélection d'un ensemble de paramètres quantitatifs et qualitatifs qui l'expliquent au mieux.

✚ **Les paramètres quantitatifs :**

Les principaux paramètres quantitatifs que l'on rencontre sont généralement les ratios fournis par l'analyse financière des comptes de l'entreprise. On en distingue :

Les ratios de structure

Les ratios de liquidité

Les ratios de rentabilité

Les ratios d'activité

❖ **Les ratios de structure :**

C'est un ensemble de ratios qui renseignent sur la nature des financements de l'entreprise :

¹⁸ www.Détection faillite.fr

Tableau N° 03 : Les ratios de structure d'une PME

Ratio	Interprétation
Autonomie financière Capitaux propres / Total dettes	Il mesure la capacité d'endettement de l'entreprise
Autonomie financière Capitaux propres / Total actifs	Il mesure le degré d'indépendance de l'entreprise par rapport à ses prêteurs.
Couverture du BFR par le FR FR /BFR	Il mesure le degré de couverture du BFR par le FR
Capacité de remboursement CAF/DLMT	Il permet de mesurer le degré de l'entreprise à honorer ses engagements.
Financement des emplois stables Capitaux propres / Immobilisations	Il mesure la couverture des emplois stables par les ressources stables.

Source : LOTMANI N, « Introduction à l'analyse financière », édition enseignement, Alger 2008.

❖ **Les ratios de liquidité :**

« Ces ratios mesurent la capacité de l'entreprise à faire face à ses engagements à brèves échéances par la mise en œuvre du fonds de roulement et/ou la liquidation des éléments de l'actif circulant »¹⁹

Tableau N°04 : Les ratios de liquidité d'une PME

Ratio	Interprétation
Liquidité générale Actifs circulants / Dettes à court terme	Il mesure la capacité de l'entreprise à payer ses dettes à court terme en utilisant ses actifs à court terme
Liquidité restreinte (créances + disponibilités) / dettes à court terme	Il mesure la capacité de l'entreprise à payer ses dettes à court terme en utilisant les créances et les disponibilités
Liquidité immédiate disponibilités/dettes à court terme	Il mesure la capacité de l'entreprise à payer ses dettes à court terme en utilisant les disponibilités.

Source : LOTMANI N, « Introduction à l'analyse financière », édition enseignement, Alger 2008.

❖ **Les ratios de rentabilité :**

¹⁹ LOTMANI N, « Introduction à l'analyse financière », édition enseignement, Alger 2008, P71.

C'est un ensemble de ratios qui mesure la performance et l'efficacité de l'entreprise dans son exploitation.

Tableau N° 05 : Les ratios de rentabilité d'une PME

Ratio	Interprétation
Taux de marge nette Résultat net/ Chiffre d'affaires HT	Il mesure la capacité d'une entreprise à générer du résultat net à partir du chiffre d'affaires.
Taux de marge commerciale Marge commerciale /Ventes de marchandises HT	Ratio qui s'applique uniquement aux entreprises commerciales et permet de le comparer avec celui des entreprises du même secteur.
Taux de marge brute EBE/Chiffre d'affaires HT	Il mesure la capacité de l'entreprise à générer une rentabilité d'exploitation à partir du chiffre d'affaires.
Rentabilité financière Résultat net / Capitaux propres	Il mesure l'aptitude de l'entreprise à rentabiliser les fonds apportés par les associés
Taux de rentabilité économique EBE/Ressources stables	Il mesure la capacité de l'entreprise à rentabiliser les fonds apportés par les associés et les prêteurs

Source : LOTMANI N, « Introduction à l'analyse financière », édition enseignement, Alger 2008.

❖ **Les ratios d'activité :**

Un ensemble de ratios qui renseigne sur l'efficacité de la gestion de l'entreprise et sur l'évolution de son activité.

Tableau N° 06 : Les ratios d'activité d'une PME

Ratio	Interprétation
Taux d'intégration Valeur ajoutée/Chiffre d'affaires HT	Il mesure le taux d'intégration de l'entreprise dans le processus de production ainsi que le poids des charges externes.
Poids de l'endettement Charges d'intérêts / EBE	Il mesure le poids de l'endettement de l'entreprise
Frais personnel Frais personnel / VA	Il mesure la part de richesse qui sert à rémunérer le travail des salariés
Rémunération de l'Etat Impôts et taxes / VA	Il mesure la part que prend l'Etat de la VA.

Rotation des stocks (entreprise industrielle) (stock moyen matières premières *360)/coût d'achat matières premières consommées	Il mesure la durée d'écoulement des stocks.une augmentation de cette durée entraîne une augmentation du BFR
Durée moyenne du crédit fournisseurs (fournisseurs et comptes rattachés /achats de biens et services TTC)*360	Il mesure la durée moyenne en jours du crédit obtenu par l'entreprise de ses fournisseurs. Il doit être supérieur au ratio du crédit clients.
Durée moyenne du crédit clients (clients et créances rattachées/Chiffre d'affaires TTC) *360	Il mesure la durée moyenne en jours du crédit consenti par l'entreprise à ses clients.

Source : LOTMANI N, « Introduction à l'analyse financière », édition enseignement, Alger 2008.

On trouve également d'autres critères tels que :

- La taille du débiteur
- Sa part de chiffre d'affaires à l'exportation
- L'évolution de son chiffre d'affaires déflaté : cet indicateur informe aussi bien sur la santé présente que future de l'entreprise.

Les paramètres qualitatifs :

L'efficacité du crédit-scoring nécessite l'intégration d'un certain nombre de critères qualitatifs reflétant les aspects non financiers qui conditionnent la réussite d'une entreprise.

Parmi ces aspects on a :

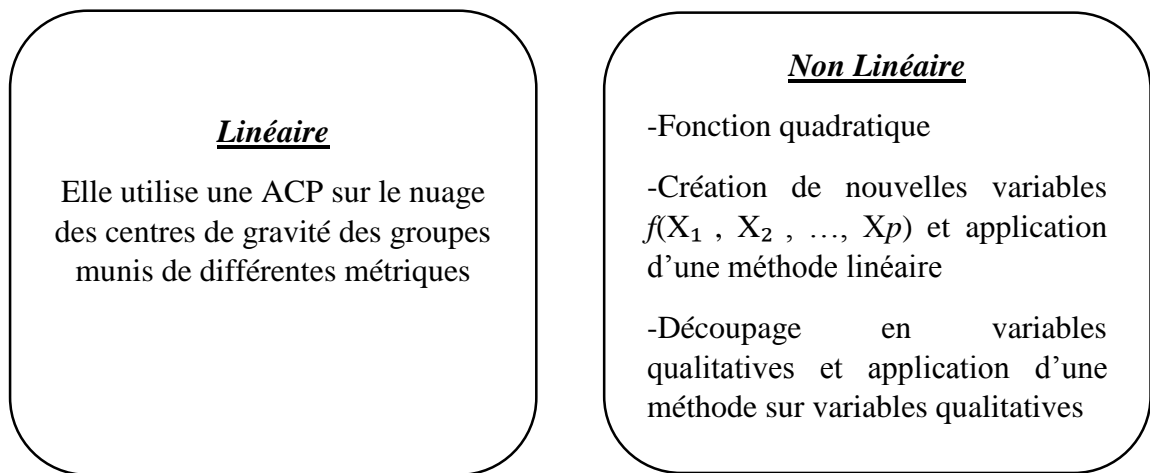
- *Le secteur d'activité ;*
- *L'âge du débiteur ;*
- *La qualité du climat social ;*
- *La forme juridique du débiteur.*
- *La relation du débiteur avec la banque*
- *Les incidents de paiement*

2. Les méthodes appliquées dans le cadre du crédit Scoring :

Il convient de distinguer deux types dans l'analyse discriminante :

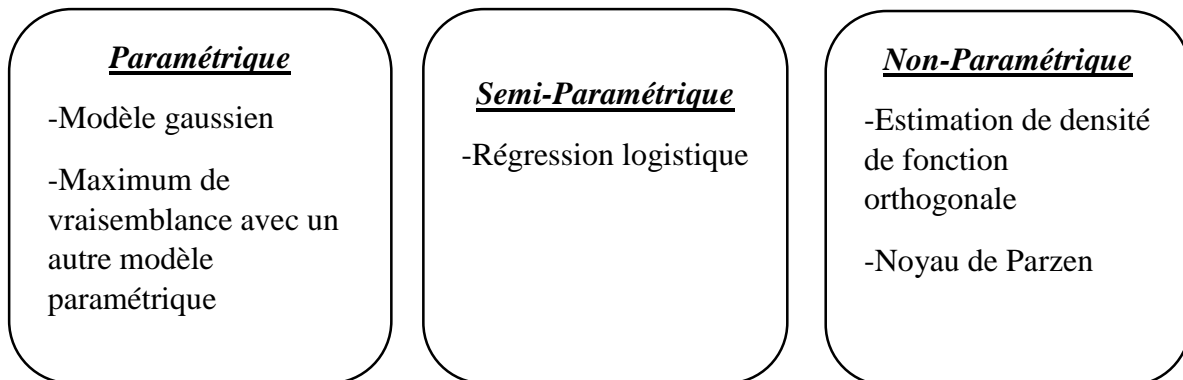
- ❖ **Les méthodes géométriques** : qui se basent sur des notions de distance et qui supposent uniquement l'égalité des matrices variances covariances. Elles optent à chercher la meilleure fonction discriminante $g(X_1, X_2, \dots, X_p)$ en adoptant deux approches selon Pierre-Louis GONZALEZ :

Figure N° 01 : les catégories des méthodes géométriques



- ❖ **Les méthodes probabilistes** : qui se basent sur l'estimation directe des probabilités d'appartenance aux groupes définis par Y et qui supposent la normalité des variables ainsi que l'équivalence des matrices de variances covariances. Selon toujours Pierre-Louis GONZALEZ on en distingue trois catégories :

Figure N° 02 : les catégories des méthodes probabilistes



- ❖ **Autres approches** : il existe d'autres approches en plus de ces deux grandes catégories notamment : les réseaux de neurones « *qui sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau*²⁰. »

Nous allons présenter néanmoins dans cette partie uniquement les deux fameuses méthodes paramétriques : la méthode logistique étant la méthode principale qui nous intéresse qu'on va aborder en détail et l'analyse discriminante linéaire ayant pour objet de comparaison avec l'autre méthode qu'on va aborder d'une manière abrégée.

1. L'analyse discriminante de Fisher :

1.1. Présentation :

« *Cette méthode est la plus ancienne des méthodes statistiques de classement. Remontant aux travaux de Fisher en 1936, elle permet de classer les individus d'une population entre différents groupes définis a priori au vue de données relatives à des variables quantitatives.*

Cette méthode se présente comme une application du classifieur de Bayes²¹ dans le cas où les descripteurs X sont des variables aléatoires continues et sont supposées suivre, conditionnellement à chaque groupe E_k , une loi normale de moyenne μ_k et une matrice de variances et covariances Σ_k . C'est donc une méthode où les probabilités conditionnelles à estimer sont supposées relever de lois de probabilités données mais dépendant néanmoins de paramètres inconnus à estimer à partir des données mises à disposition (estimation paramétrique) ».²²

Cette méthode est connue sous deux types qui sont complémentaires l'une pour l'autre :

❖ **L'analyse discriminante descriptive :**

Pour une population subdivisée en K groupes et décrites en J caractères, l'analyse discriminante descriptive a deux objectifs :

²⁰ Claude TOUZET, « *les réseaux de neurones artificiels* », document pratique, Collection de l'EERIE, 2006.

²¹ Le classifieur de Bayes est la règle de décision minimisant la perte moyenne parmi toutes les règles de décision possibles.

²² Hassen MATHLOUTHI, idem, P26.

- Objectif descriptif : Mettre en évidence les caractéristiques qui permettent de distinguer au mieux les groupes
- Objectif prédictif : Classer automatiquement un nouvel individu (l'affecter à un groupe) à partir de ses caractéristiques

❖ **L'analyse discriminante décisionnelle :**

« L'analyse discriminante décisionnelle pose le problème suivant : étant donné un nouvel individu sur lequel on a observé les p variables X_j mais pas la variable qualitative Y (l'appartenance à un groupe), comment décider de la modalité y_h de Y , c'est-à-dire du groupe auquel appartient cet individu. Pour cela nous allons définir des règles de décision et nous donner les moyens de les évaluer sur un seul individu. »²³

1.2. Le traitement des variables qualitatives par la méthode DisQual :

Etant donné que la méthode de discrimination linéaire est applicable que sur des variables quantitatives, l'introduction des variables qualitatives dans l'analyse n'est pas impossible mais nécessite une transformation numérique de ces dernières en passant par plusieurs étapes.

Etape 1 :

Soient p variables qualitatives X_1, X_2, \dots, X_p à m_1, m_2, \dots, m_p modalités.

La première étape consiste à appliquer une ACM (analyse des correspondances multiple) sur ces variables qualitatives.

Etape 2 :

« Remplacer les p variables qualitatives par les m modalités sur les axes factoriels en attribuant à chaque catégorie de chaque variable un score ou une valeur numérique. Cela revient à transformer chaque variable qualitative en une variable discrète à m valeurs »²⁴.

A ce niveau-là, l'analyse discriminante sur les variables qualitatives devient possible et répond aux hypothèses de travail de la méthode.

1.3. La fonction score et problème d'affectation :

²³ Laurent CARRARO, « Notions sur l'analyse discriminant », ENSM.SE, 2007, P19

²⁴ Pierre-Louis GONZALEZ, « l'analyse discriminante », document de formation, 2010.

Le score Z pour une entreprise i donnée s'exprime comme la différence entre la valeur de la fonction discriminante pour cette entreprise et sa valeur pour l'entreprise moyenne.

$$Z = \sum_{i=1}^P \alpha_i X_i + cte$$

P_i est la valeur moyenne de la variable X_i calculée sur l'ensemble de la population appelée parfois valeur pivot.

L'affectation à l'une ou l'autre des classes se fait sur la base du signe de Z :

- Si le signe est positif, l'entreprise est affectée au groupe dont les performances sont supérieures aux performances de l'entreprise moyenne
- Si le signe est négatif, l'entreprise est affectée au groupe le moins performant

En effet, divers critères d'affectation ont été proposés dont le plus fréquemment employés consistent à affecter l'entreprise qu'on veut classer au groupe dont le point moyen est le plus proche.

Soit i l'individu à classer. On calcule sa distance au point moyen de chaque groupe g_k :

$$d(i, g_k) = (i - g_k)'T^{-1}(i - g_k)$$

On décide d'affecter i au groupe g_0 tel que :

$$d(i, g_0) = (i - g_0)'T^{-1}(i - g_0) = \min d(i, g_k)$$

1.4. Pouvoir discriminant et performance de la discrimination :

Le pouvoir discriminant de la fonction discriminante est mesuré par la valeur propre associée à cet axe et qui représente la part de la variance totale de la forme linéaire expliquée par la variance inter-classe.

De tel, le pouvoir discriminant de la variable X_i est mesuré par :

$$\frac{\alpha^2 i \sigma^2 i}{\sum_i^P \alpha^2 i \sigma^2 i}$$

Le terme $\sigma^2 i$ désigne la variance totale de la variable X_i

En terme de coût de la fonction discriminante, est d'autant plus coûteuse (coût de collecte des données et coût du traitement informatique) que le nombre de variables que l'on prend en compte est plus élevé. Il est donc souhaitable de limiter le nombre de variables.

D'autre part, la fiabilité de la discrimination sera meilleure si l'on considère un petit nombre de variables. Ceci peut être effectué en passant par une analyse en composantes principales (ACP).

1.5. Test de performance du modèle : Q de press

C'est un test qui permet de vérifier la capacité prédictive de la fonction score élaborée par rapport au Hazard dont l'hypothèse nulle H_0 est « *le hazard classe les entreprises mieux que la fonction score* ».

La statistique du test qui suit la loi de khi-deux à 1 degré de liberté s'écrit comme suit :

$$Q = \frac{(n - n'k)^2}{n(k - 1)}$$

Rejeter l'hypothèse nulle du test revient à constater que le modèle score classe mieux que le hazard. Sachant que :

n : la taille de l'échantillon

n' : l'effectif des entreprises correctement classées

k : le nombre de groupes d'affectation

1.6. La validation croisée

La validation croisée sert à tester la capacité prédictive du modèle score sur un nouvel échantillon d'entreprises appelé l'échantillon de test. Cette étape permet de tester la performance du modèle à classer correctement les entreprises. Elle fournit les taux de bon classement, de mauvais classement ainsi que la sensibilité et la spécificité du modèle servant à déduire d'autres indicateurs de performance tel que la courbe ROC, AUC ...

1.7. Avantages et limites de l'analyse discriminante :

❖ Avantages²⁵ :

L'analyse discriminante linéaire est très répandue dans les logiciels de Statistique et d'apprentissage et est très utilisée. Les raisons de son succès sont les suivantes :

²⁵ Georges COLINCE, « *Evaluation statistique du risque crédit par la technique du scoring* », mémoire de master de statistique appliquée, 2007.

- Elle offre souvent un très bon compromis pertinence/complexité ; autrement dit, elle permet souvent de bien résoudre le dilemme biais-variance. Elle est ainsi souvent supérieure à l'analyse discriminante quadratique qui dépend d'un nombre notamment plus important de paramètres.
- Dans le cadre de l'analyse discriminante linéaire, la sélection de variables peut être réalisée de manière quasi optimale en utilisant une statistique F de Fisher. En fait, les critères classiques de sélection de variables supposent de manière sous-jacente les hypothèses gaussiennes de l'analyse discriminante gaussienne.
- L'analyse discriminante linéaire fournit des résultats stables (peu sujets aux fluctuations d'échantillonnage) et robustes (supportant bien des écarts assez importants aux hypothèses de normalité des groupes et d'égalité des matrices variances).

❖ ***Inconvénients***²⁶ :

Les méthodes statistiques de scoring et particulièrement l'analyse discriminante souffrent néanmoins de quelques insuffisances dont entre autres :

- La décision pouvant être prise suite à l'utilisation des méthodes de scoring est basée sur une probabilité et non sur une certitude
- Les méthodes statistiques de scoring supposent comme toute autre méthode statistique que le futur est identique au passé.
- Le risque est expliqué par les seules variables disponibles
- La mise en place d'un système de scoring dans une entreprise n'est pas toujours facile à réaliser du fait de la nécessité de son intégration informatique avec les autres systèmes d'information.

2. La méthode discriminante logistique

Etant donné que notre travail porte essentiellement sur l'analyse discriminante logistique, cela nécessite une présentation beaucoup plus approfondie de la méthode que celle menée par l'analyse discriminante de Fisher. Pour cela, nous allons dans un premier temps introduire des généralités de la méthode logistique standard, ses fondements et les points de divergences par rapport à l'analyse discriminante de Fisher et dans un deuxième temps, procéder à la

²⁶ Hassen MATHLOUTHI, idem, P8.

construction, l'évaluation et la validation d'un modèle LOGIT qui seront abordées dans la « section 2 » de ce chapitre.

2.1. Présentation

L'analyse discriminante logistique dans sa version standard est une autre méthode de classement aussi connue que la méthode d'analyse discriminante de Fisher. Tout en reposant sur le principe bayésien de classement, elle se distingue toutefois de cette dernière en procédant plutôt à modéliser les probabilités à posteriori et non les probabilités conditionnelles.

❖ *L'hypothèse fondamentale de la méthode logistique :*

« Les hypothèses d'un modèle discriminant pèsent sur la forme de la frontière induite pour distinguer les classes dans l'espace de représentation. La régression logistique par exemple produit un séparateur linéaire, c'est la principale information qu'il faut retenir »²⁷

Considérons d'abord la probabilité conditionnelle $P(Y = y_k / X)$:

$$P(Y = y_k / X) = \frac{P(Y = y_k) \times P(X / Y = y_k)}{P(x)}$$

Dans le cas à deux classes, nous devons comparer simplement $P(Y = 1 / X)$ et $P(Y = 0 / X)$. Tel que :

- (1) représente le groupe « mauvais »
- (2) représente le groupe « bon »

Ceci nous conduit à en former le rapport suivant :

$$\frac{P(Y = 2 / X)}{P(Y = 1 / X)} = \frac{P(Y = 2)}{P(Y = 1)} * \frac{P(X / Y = 2)}{P(X / Y = 1)}$$

La phase la plus importante réside dans l'estimation du rapport de probabilité $\frac{P(X / Y = 2)}{P(X / Y = 1)}$. La régression logistique introduit l'hypothèse fondamentale suivante :

$$\ln \left[\frac{P(X / Y = 2)}{P(X / Y = 1)} \right] = b_0 + b_1 X_1 + \dots + b_j X_j$$

²⁷ Ricco Rakotomalala, « *Pratique de la Régression Logistique* », Université Lumière Lyon, 2017, P75.

Cette hypothèse englobe plusieurs lois notamment la loi normale, exponentielle, beta, poisson, etc.

« Contrairement à l'Analyse Discriminante Linéaire, que l'on qualifie de méthode paramétrique car on émet une hypothèse sur les distributions respectives de $P(X/Y = 1)$ et $P(X/Y = 2)$ (loi normale), la régression logistique est une méthode semi-paramétrique car l'hypothèse porte uniquement sur le rapport de ces probabilités. Elle est moins restrictive. Son champ d'action est donc théoriquement plus large »²⁸

❖ **Le modèle LOGIT :**

L'expression suivante représente la transformation LOGIT de $\pi(i)$ pour une entreprise i :

$$\ln \left[\frac{\pi(i)}{1 - \pi(i)} \right] = a_0 + a_1 X_1 + \dots + a_j X_j$$

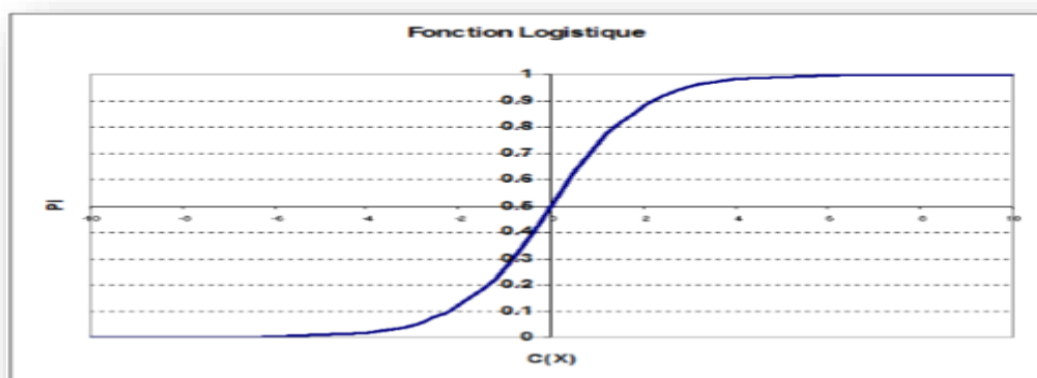
La quantité $\frac{\pi}{1-\pi} = \frac{P(Y=2/X)}{P(Y=1/X)}$ exprime un odds c-à-d un rapport de chances de la survenance du numérateur comparé au dénominateur.

Notons $B(X) = a_0 + a_1 X_1 + \dots + a_j X_j$, π étant représenté par la fonction logistique devient :

$$\pi = \frac{e^{B(X)}}{1 + e^{B(X)}}$$

Le LOGIT = $B(X)$ est théoriquement défini entre $-\infty$ et $+\infty$. En revanche, $0 \leq \pi \leq 1$ issue de la transformation de $B(X)$ représente une probabilité comme indiqué sur la figure suivant :

Figure N° 03 : la présentation graphique de la fonction logistique (<http://www.adscience.fr>)



²⁸ Ricco Rakotomalala, idem. P 78.

❖ *La règle d'affectation :*

La règle d'affectation peut être basée sur π de différentes manières :

- Si $\frac{\pi}{1-\pi} > 1$ Alors $Y = 2$
- Si $\pi > 0.5$ Alors $Y = 2$

On peut constater à ce stade et à travers ces généralités concernant la méthode discriminante logistique que :

- ✓ Poser que le score est linéaire équivaut donc à supposer que les probabilités à postériori sont logistiques
- ✓ L'analyse discriminante logistique consiste à estimer ces probabilités à partir des données, contrairement à l'analyse discriminante dont l'objectif est d'estimer les lois conditionnelles.
- ✓ On se base toujours sur le principe bayésien : l'individu est à affecter au groupe pour lequel la probabilité à postériori d'y appartenir est la plus élevée. Cela suppose au préalable que l'on dispose d'une estimation des a_j pour j allant de 0 à P , étape que nous allons appréhender en détail dans la « *section 2* ».

SECTION 02 : ELABORATION D'UN MODELE LOGIT DANS LE CADRE DU CREDIT SCORING

L'élaboration d'un modèle LOGIT consiste à passer par plusieurs étapes primordiales :

- 1- Le choix et la sélection des variables
- 2- La construction et la modélisation
- 3- L'évaluation et validation du modèle

1. Le choix et la sélection des variables :

Cette première étape consiste à collecter un ensemble préliminaire bien diversifié des variables X pouvant être en corrélation avec la variable à expliquer ($Y =$ bon crédit '2' ou mauvais '1'). Cette phase de présélection des variables candidates a été automatisée par programmation des tests statistiques dont les plus connus sont khi-deux, Kendall, Spearman, Kruskal Wallis et par sélection automatique des variables candidates ayant une forte liaison à la variable endogène.

« La seconde étape consiste à analyser la corrélation entre les variables candidates et d'éliminer celles qui sont fortement corrélées entre elles-mêmes, ceci évitera de tomber dans la redondance ainsi que l'utilisation d'un grand nombre de variables qui ne portent pas autant d'informations pertinentes à la variable d'intérêt. Cette étape se fait à travers la matrice de corrélation qui met en liaison toutes les variables candidates choisies dont l'élimination se base sur le choix d'un seuil qui, selon les auteurs et la littérature actuelle, n'a pas été exactement fixé mais considèrent que deux variables sont fortement corrélées lorsque leur coefficient de corrélation est supérieur au coefficient de corrélation moyen observé dans la matrice de corrélation »²⁹.

La troisième étape quant à elle consiste à étudier l'effet conjoint des variables retenues dans la deuxième étape sur la variable à expliquer à travers une régression logistique qui indiquera sur la significativité et le pouvoir explicatif des variables exogènes afin de sélectionner définitivement les variables explicatives du modèle LOGIT à construire. Cette étape sera bien détaillée dans le chapitre III.

La construction et la modélisation :

Après avoir déterminé l'ensemble des variables à retenir dans la construction du modèle LOGIT, il convient dès lors et avant de passer à l'estimation des coefficients du modèle, de subdiviser notre échantillon en deux sous échantillons « Test-Apprentissage » (pour plus détails sur ce point, voir la section 1 du CHAPITRE III).

1.1. Estimation des paramètres :

L'estimation des paramètres d'un modèle LOGIT se fait par la maximisation de la vraisemblance noté L. Néanmoins, il convient d'abord de déterminer la loi distribution de P(Y/X).

Pour une entreprise i , on modélise la probabilité à l'aide de la loi binomiale $B(1, \pi)$, avec :

$$P[Y(i)/X(i)] = \pi(i)^{y(i)} * (1 - \pi(i))^{(1-y(i))}$$

Ceci conduit à déduire la vraisemblance du modèle donnée comme suit :

$$\prod_i \pi(i)^{y(i)} * (1 - \pi(i))^{(1-y(i))}$$

²⁹ Pascal Legrand, « Le choix des variables explicatives dans les modèles de régression logistique », papier de conférence, 2015.

Cette méthode de vraisemblance consiste à optimiser \hat{a} de telle sorte à rendre la vraisemblance maximum (on rappelle que \hat{a} est sans biais, de variance minimale et il est asymptotiquement gaussien).

Quoique, pour des raisons de simplicité, il est préférable d'utiliser le logarithme de la vraisemblance noté LL, ce qui donne une nouvelle expression :

$$LL = \sum_i y \ln \pi + (1 - y) \ln(1 - \pi)$$

1.2. Evaluation préliminaire du modèle :

Cette première évaluation consiste à dire si le modèle LOGIT construit à première vue est intéressant ou pas. Plusieurs indicateurs peuvent nous renseigner sur la significativité globale du modèle dont les plus pertinents sont :

▪ *Le pseudo-R² ou coefficient de McFadden :*

C'est un coefficient d'ajustement qui est défini comme suit :

$$R^2 = 1 - \frac{L(a)}{L(0)}$$

Où L(a) et L(0) sont respectivement le log-vraisemblance du modèle estimé à P variables et le log-vraisemblance du modèle construit uniquement de la constante.

- Quand le R² est proche de zéro, cela veut dire que la régression est inutile et que les variables ne sont pas pertinentes
- Quand le R² est proche de 1, cela veut dire que L(a) est plus grande que L(0) ce qui signifie un bon pouvoir explicatif des variables exogènes.

▪ *Le test du rapport de vraisemblance :*

Un autre test renseignant sur la significativité globale du modèle (test de nullité des tous les coefficients) est élaboré et établi à partir des deux log-vraisemblance noté LRT qui suit une loi de khi-deux à P degrés de liberté :

$$LRT = -2[L(a) - L(0)]$$

Ce test mesure la qualité générale d'ajustement du modèle en mesurant si le modèle construit fournit une meilleure explication de la variable Y que le modèle réduit uniquement à la constante. Généralement, on rejette l'hypothèse nulle du test au seuil de 5% pour confirmer la bonne qualité du modèle.

▪ ***Le test de Wald : (test de nullité globale des coefficients)***

Le test de Wald dans cette première version, permet d'évaluer la significativité globale du modèle, c.à.d tester la nullité simultanée des coefficients relatifs aux variables explicatives du modèle. Il repose sur l'hypothèse que le vecteur \hat{a} est asymptotiquement normal. L'hypothèse nulle de test exclut la constante et est donnée par :

$$H_0 : a_1 = a_2 = \dots = a_p = 0$$

La statistique de Wald suit la loi de khi-deux à P degrés de liberté et est donnée par :

$$w_j = \hat{a}' \cdot \sum_j^{-1} \cdot \hat{a}$$

Tel que : \sum Est la matrice variance-covariance des coefficients (pour j allant de 1 à P)

Commentaires :

- Le test du rapport de vraisemblance est plus puissant. Il détecte mieux l'hypothèse alternative lorsque cela est justifié. Il est en revanche plus gourmand en ressources. Ce problème ne se pose véritablement que lorsque nous avons à traiter une grande base de données.
- Le test de Wald est moins puissant, car il favorise l'hypothèse nulle H_0 ce qui falsifie les résultats de l'analyse et conduit par la suite à supprimer à tort des variables importantes du modèle. Il en revanche peu gourmand en ressources.

▪ ***La significativité partielle des paramètres du modèle :***

Ce test s'appuie également sur la statistique de Wald notée w_j qui, sous H_0 , suit une loi de khi-deux à 1 degrés de liberté :

$$w_j = \frac{\hat{a}_j^2}{\sigma_j^2}$$

Où σ_j^2 est la variance estimée du coefficient $\hat{\beta}_j^2$, lue sur la diagonale principale de la matrice de variance covariance des coefficients.

2. Evaluation et validation du modèle :

Une fois que le modèle LOGIT ait été construit et prédit, il convient ensuite de l'évaluer, c'est à dire :

- Comparer les vraies valeurs observées à celles prédites par le modèle ;
- Comparer les vraies valeurs de π (probabilité de défaut) avec celles fournies par le modèle.

Comme précédemment dit, l'évaluation ainsi que la validation du modèle logistique doit se faire sur un échantillon autre que celui qui a servi pour la construction du modèle (échantillon test).

Plusieurs techniques et indicateurs sont mis en œuvre pour répondre à ce besoin dont les plus pertinents sont :

2.1. La validation croisée :

« C'est une matrice qui confronte les valeurs observées de la variable dépendante avec celles qui sont prédites, puis comptabilise les bonnes et les mauvaises prédictions. Son intérêt est qu'elle permet à la fois d'appréhender la quantité de l'erreur (le taux d'erreur) et de rendre compte de la structure de l'erreur (la manière de se tromper du modèle) ». ³⁰

Tableau N° 07 : la matrice de confusion

$Y_x \hat{Y}$	1	0	Total
1	a	b	a+b
0	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

La matrice de confusion présente un outil très utile qui permet d'extraire plus d'informations telles que :

- *Le taux d'erreur* : qui représente le taux de mauvais classement par rapport au total des effectifs : $\text{taux d'erreur} = \frac{b+c}{n}$

³⁰ Ricco Rakotomalala, op.cit, P138.

- *Le taux de succès* : qui représente le taux de bon classement rapporté au total des effectifs : $\text{taux de succès} = \frac{a+d}{n}$
- *La sensibilité* : qui indique la capacité du modèle à retrouver les vrais positifs $\frac{a}{a+b}$
- *La précision* : indique les vrais positifs parmi ceux qui ont été classés positifs $\frac{a}{a+c}$
- *La spécificité* : c'est l'inverse de la sensibilité $\frac{d}{c+d}$
- *Le rapport de vraisemblance* : il indique les chances que les positifs soient classés positifs : $\frac{\text{sensibilité}}{1-\text{spécificité}}$, Plus grande est sa valeur, meilleur sera le modèle.

2.2. Le test de Hosmer-Lemeshow :

C'est un test qui permet de comparer les vraies valeurs de π (probabilité du score) avec celles prédites par le modèle. Il se base sur la subdivision des données en G groupes en calculant pour chaque groupe la moyenne des scores observés.

L'hypothèse nulle de ce test est H_0 : le modèle est de bonne qualité et sa statistique qui suit la loi de khi-deux à G-2 degrés de liberté est donnée par :

$$\sum_g \frac{(mg1 - \dot{m}g1)^2}{\dot{m}g1(1 - \text{moy}(\pi g1))}$$

Tel que :

- $mg1$: est le nombre de positifs dans chaque groupe
- $\pi g1$: est la moyenne des scores observés dans le groupe g

2.3. Le test Mann-Whitney :

C'est un test non paramétrique qui permet de comparer graphiquement les distributions des scores conditionnellement aux groupes d'appartenance. Le modèle est jugé bon quand les deux distributions conditionnelles sont séparées et mauvais quand elles sont confondues. C'est un test qui permet de dire si les scores positifs sont significativement plus élevés que les négatifs.

Sous H_0 « les distributions sont confondues », la statistique du test notée Z suit une loi normale centrée réduite $N(0,1)$ et est donnée par :

$$Z = \frac{U - n_1n_2}{\sqrt{\frac{1}{12}(n^1 + n^2 + 1)n_1n_2}}$$

Tel que : $U = \text{Min} (U_1 , U_2)$ tout en sachant que :

$$U_1 = r_1 - \frac{n_1(n_1+1)}{2} \qquad U_2 = r_2 - \frac{n_2(n_2+1)}{2}$$

NB : r_1 et r_0 étant les sommes conditionnelles de rang pour chaque groupe.

2.4. La courbe ROC³¹ :

La courbe ROC est un outil très intéressant pour l'évaluation et la comparaison de la performance d'un modèle LOGIT. Il permet de :

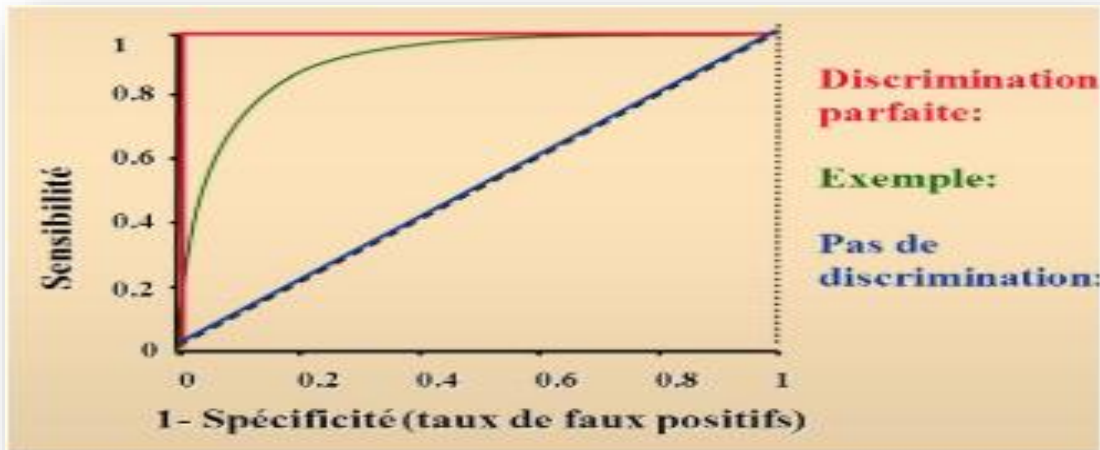
- Elle propose un outil graphique qui permet d'évaluer et de comparer globalement le comportement des classifieurs.
- Elle est indépendante des coûts de mauvaise affectation. Elle permet par exemple de déterminer si un classifieur surpasse un autre, quelle que soit la combinaison de coûts utilisée.
- On peut lui associer un indicateur synthétique, le critère AUC qu'on abordera ultérieurement.

La courbe ROC met en relation le taux de vrais positifs (la sensibilité) et le taux de faux positifs (1 - Spécificité) dans un graphique nuage de points. Habituellement, nous comparons $\pi(i)$ estimés à un seuil $s = 0.5$ pour effectuer une prédiction de $y(i)$.

Deux situations extrêmes peuvent survenir. La discrimination est parfaite. Tous les positifs sont situés devant les négatifs, la courbe ROC est collée aux extrémités Ouest et Nord du repère. Les scores sont totalement inopérants, le classifieur attribue des valeurs au hasard, dans ce cas les positifs et les négatifs sont mélangés. La courbe ROC se confond avec la première bissectrice.

³¹ Ricco Rakotomalala, op.cit, P80.

Figure N° 04 : Les formes possibles de la courbe ROC (<http://www.adscience.fr>)



2.5. Le critère AUC :

Il représente la surface située sous la courbe ROC, c'est la version numérique de cette courbe. En effet, il exprime la probabilité de placer un bon individu devant un mauvais individu, le tableau suivant résume l'interprétation globale de ce critère :

Tableau N° 08 : l'interprétation des valeurs de l'AUC

Valeur de l'AUC	Commentaire
AUC = 0.5	Pas de discrimination
$0.7 < \text{AUC} < 0.8$	Discrimination acceptable
$0.8 < \text{AUC} < 0.9$	Discrimination excellente
AUC > 0.9	Discrimination exceptionnelle

Source : R. Rakotomalala, « pratique de la régression logistique »

2.6. Analyse des résidus :

L'analyse des résidus vient en dernière étape de la validation du modèle LOGIT construit, elle permet de diagnostiquer la qualité de la régression en déterminant les points qui ont été mal modélisés (mal expliqué par le modèle). Parmi les outils utilisés lors de l'analyse des résidus on en retient le plus fameux :

❖ Résidus de Pearson :

Cela consiste à modéliser la variable endogène Y en fonction de l'erreur et de la probabilité du score comme suit :

$$Y(i) = \pi(i) + \varepsilon(i)$$

Où $\varepsilon(i)$ est l'erreur de modélisation, avec $\varepsilon(i) = Y(i) - \pi(i)$, elle peut prendre deux valeurs possibles :

$$\left\{ \begin{array}{l} \varepsilon(i) = 1 - \pi(i) \text{ avec la probabilité } \pi(i) \\ \varepsilon(i) = -\pi(i) \text{ avec la probabilité } 1 - \pi(i) \\ \text{Avec : } E(\varepsilon) = 0 \text{ et } V(\varepsilon) = \pi(1 - \pi) \end{array} \right.$$

Pour un individu ω , le résidu de Pearson permet d'identifier les points mal modélisés :

$$r(i) = \frac{y(i) - \pi^{\wedge}(i)}{\sqrt{\pi^{\wedge}(i)(1 - \pi^{\wedge}(i))}}$$

Ce résidu suit approximativement une loi gaussienne $N(0,1)$. Ainsi, pour un intervalle de confiance de 95%, tout point dépassant plus ou moins 2 est considéré mal expliqué. La lecture des résultats peut être facilitée à travers la représentation graphique des nuages de points de la variable Y en fonction des variables explicatives choisies qui indique les points inusuels par rapport aux autres.

SECTION 03 : AVANTAGES ET LIMITES DE LA METHODE LOGISTQUE

1. Avantages :

- En proposant une appréciation synthétique de la situation d'une entreprise, la méthode des scores permet, d'anticiper le risque de défaillance de l'entreprise et de diminuer par conséquent les impayés, aussi parce qu'il fondé sur une appréciation objective des critères de risque, l'utilisation des scores permet à l'établissement de crédit de disposer en fonction de sa sensibilité aux risques le niveau d'impayés qu'il tolère ;
- « *Les modèles de score par rapport aux autres méthodes traditionnelles permettent, grâce à la rapidité de décision qu'ils présentent, un traitement de masse de populations nombreuses d'emprunteurs et leur usage réduit de manière significative la durée du traitement des dossiers de crédit (de 15 jours à quelques heures, pour la plupart des crédits standard)* »³².
- Contrairement à la regression linéaire classique où la variables à expliquer doit être de nature quantitative et continue et les variables explicatives de nature quantitatives ou binaires, la regression logistique présente l'avantage d'étudier l'effet de variables explicatives quantitatives et/ou qualitatives ;
- Les variables explicatives de la regression logistiques ne doivent pas forcément respecter la contrainte de normalité de leur distribution (méthode non paramétrique), contrairement à la regression classique ;
- La variable à expliquer dans la regression logistique, contrairement à la regression classique, peut ne pas être en forme linéaire avec les variables explicatives.

2. Limites :

En dépit des avantages de la méthode logistique, le modèle trouve ses limites dans le fait que :

- Les variables explicatives, comme les coefficients, sont extrêmement dépendants du contexte économique-financier et de la population elle-même ayant servi de base au calcul de la fonction de score.

³² BENSAT Samir, « *Gestion du risque de crédit par la méthode crédit scoring* », mémoire de master, 2013, P68.

- Le modèle suppose une répétition dans le futur de comportements passés. En effet, les modèles de score capturent mal les changements de toute nature ce qui modifient l'attitude des emprunteurs par rapport au défaut ;
- Enfin, l'étude est fondée sur des éléments quantitatifs qui connaît quelques failles notamment au mode de calcul des grandeurs comptables. Il en est de même pour les éléments qualitatifs et les informations collectées par les banques auprès des entreprises où parfois semblent incohérentes, très optimistes et non fiables pour mener une étude réaliste et significative de risque de crédit.
- Les modèles de score sont des outils statistiques. Ils comportent deux types d'erreurs, l'erreur de type : l'erreur de type I qui consiste à classer comme sain un emprunteur dont la probabilité de défaut est en réalité élevée et l'erreur de type II qui consiste à classer en défaut des emprunteurs sains.³³
- Au niveau des banques algériennes, la mise en place d'un tel outil semble toujours difficile à atteindre du fait du manque de systèmes experts et de personnel compétent en la matière afin d'implanter un outil puissant et rigide dans le temps et dans l'espace qui permet de faire face au risque de crédit.

CONCLUSION :

A travers ce chapitre, nous avons présenté d'abord la méthode scoring en général, son objectif et sa méthodologie et ensuite avons explicité les deux modèles majeurs illustrant cette méthode étant l'analyse discriminante de Fisher qui se base sur le calcul des lois conditionnelles et les principes des lois paramétriques et la méthode logistique qui fait l'objet de notre étude qui se base quant à elle sur les principes des lois non paramétriques et sur le calcul des probabilités à postériori.

En outre, nous avons évoqué toutes les étapes nécessaires à la construction d'un modèle LOGIT dans le cadre de la prévention de la défaillance des PME allant de la sélection des variables jusqu'à la validation du modèle.

Dans le troisième prochain chapitre, nous allons mettre sur le terrain toutes ces notions et toutes ces étapes mise en œuvre à l'élaboration d'un modèle scoring apte à prévenir, à un certain niveau de confiance, correctement la défaillance des entreprises de la CNEP-banque qui sera notre champ d'application.

³³ Patrick BOISSELIER, « *scoring et anticipation de défaillance des entreprises* », article, 2011.

CHAPITRE III

APPLICATION DE LA METHODE SCORING SUR LE CAS DE LA CNEP- BANQUE

INTRODUCTION :

Nous avons présenté précédemment dans la partie théorique les deux méthodes empiriques utilisées dans la prévention du risque de crédit. Dans ce dernier chapitre, on va s'intéresser d'abord à mettre en œuvre la méthode logistique sur les données de PME de la CNEP-banque, élaborer le modèle LOGIT et tester sa robustesse. Dans un deuxième lieu, on va appliquer sur les mêmes données la méthode de discrimination linéaire, montrer son intérêt dans la prévention de la défaillance. Enfin, comparer les résultats de la méthode de discrimination linéaire à ceux de la méthode logistique dans le but de mettre en lumière les atouts de cette dernière par rapport à l'autre méthode.

SECTION 01 : LES RAISONS DE DEFAILLANCE DES PME

« Depuis le pic de faillites qui a frappé l'ensemble des pays de l'OCDE (L'Organisation de coopération et de développement économiques) en 2009, le nombre d'entreprises qui disparaissent annuellement reste à un niveau élevé. Tous les secteurs et toutes les tailles d'entreprises sont concernés par ce risque de disparition accru.

Les disparitions de groupes industriels qui ont défrayé les chroniques constituent la partie la plus visible de ce phénomène mais, derrière ces affaires retentissantes, ce sont tous les ans plusieurs milliers de PME et, avec elles, des dizaines de milliers d'emplois qui disparaissent.

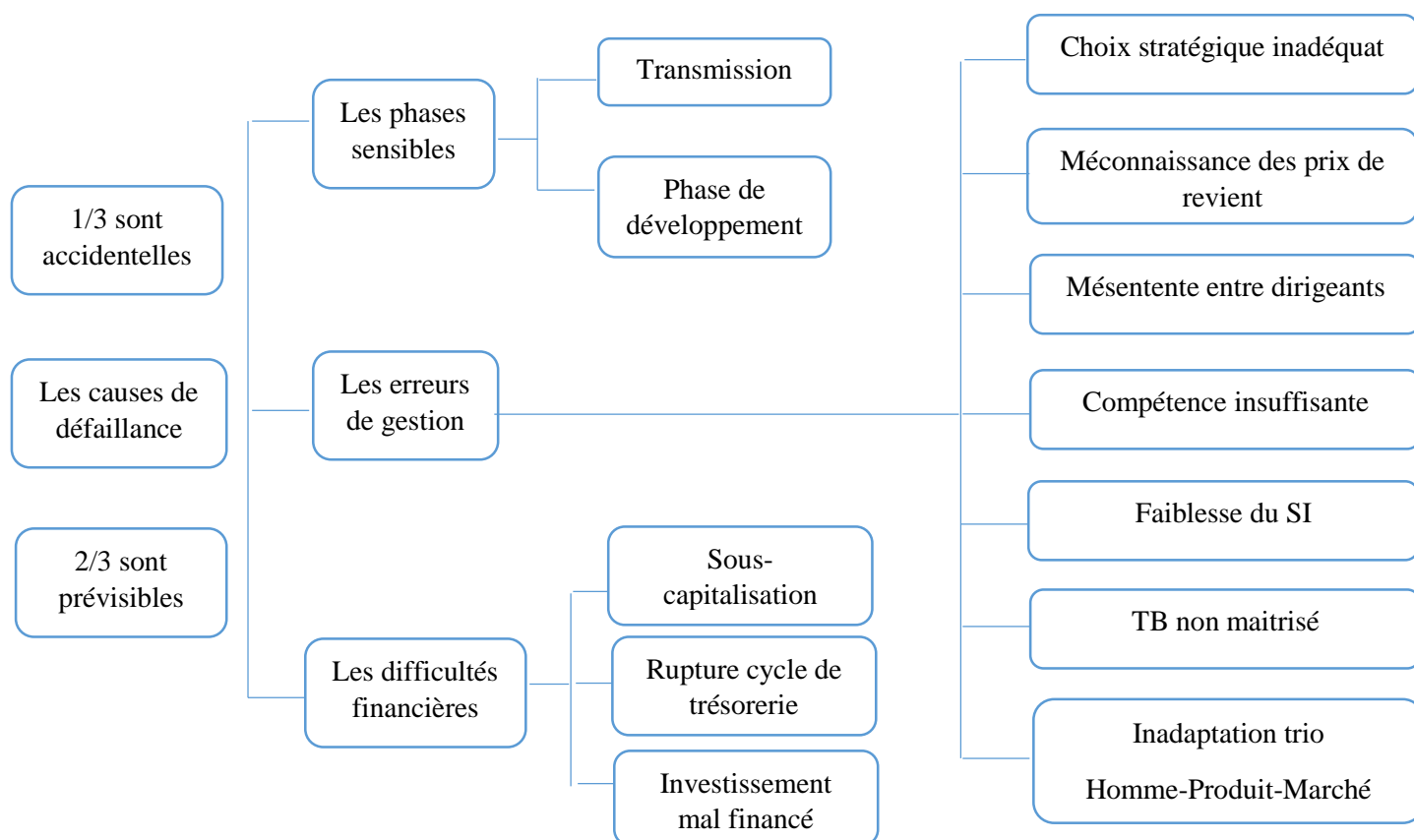
S'il est difficile d'estimer précisément le nombre d'emplois détruits par les liquidations judiciaires et les cessions d'éléments d'actif, les analystes s'accordent à considérer que le coût économique, social et personnel de la fermeture d'entreprises est exceptionnellement élevé et mérite d'être pris en considération pour mieux être traité ».³⁴

Selon BLAZY (2000), « *la défaillance correspond à un évènement qui modifie profondément le cadre juridique dans lequel évolue l'entreprise. Elle va cependant au-delà des seuls aspects techniques et financiers pour impacter la vie personnelle et sociale des entrepreneurs qui la subissent* ».

Pour mieux cerner donc ce phénomène d'échec des entreprises, il est primordial de creuser profondément dans les raisons directes et indirectes susceptibles de causer ou d'aggraver ce phénomène. Parmi les études déjà faites, Christophe GODDYN, entre autres, a pu élucider les principales causes de la défaillance des entreprises qu'on a résumé dans le schéma suivant :

³⁴ Nadine LEVRATTO, « la défaillance et l'échec des PME, revue internationale PME, 2018.

Figure N 05 : les principales raisons de défaillance des PME



Source : C. GODDYN, « Gestion de la PME, Guide pratique du chef d'entreprise et de son conseil »

❖ **La répartition des créances aux entreprises en relation avec la CNEP-banque :**

Selon les rapports de la centrale des risques de la CNEP-banque liés à la défaillance des entreprises, le critère de défaillance choisi pour juger classer les entreprises est celui de leur insolvabilité dont les principales raisons causant ce phénomène ont été jugées d'ordre prévisible comprenant les erreurs de gestion et les difficultés financières indiquées dans le schéma en haut.

En effet, les entreprises en relation avec la CNEP-banque ont tendance, après l'octroi et l'obtention du crédit, à négliger certains indicateurs de gestion et de performance notamment la non maîtrise de certains postes de bilan liés principalement à leurs cycles d'exploitation et à leurs trésoreries qui mènent à bouleverser l'équilibre financier de l'entreprise ainsi que sa stratégie financière et d'investissement et sa politique de développement.

La centrale des risques de la banque explique notamment que les raisons de défaut des entreprises sont liées parfois à la non transparence et à l'aléa moral de ces dernières vis-à-vis de la banque concernant certaines informations surtout managériales et stratégiques menés par les dirigeants de l'entreprise qui s'avèrent souvent non rigides et très sensibles à la conjoncture économique et concurrentielle du marché.

Voici quelques statistiques extraites des rapports de la centrale des risques de la CNEP-banque expliquant au mieux l'évolution de l'octroi du crédit aux entreprises ainsi que la répartition des créances classées constatées au cours des années (2014-2015-2016).

❖ **La répartition des engagements au cours des trois années :**

Les engagements de la CNEP-banque envers les entreprises on atteint pendant les trois années concernées les 1 877 143 173.39DA. En effet, l'ensemble de ses engagements a été repartitionné en deux catégories à savoir :

- Les créances courantes dont le règlement des échéances se fait d'une manière régulière et sans retard de paiement
- Les créances classées qui représentent la somme des impayés du au retard de paiement des échéances aux dates convenues.

Le tableau ci-dessous résume la répartition des engagements (en volume) de la banque pendant les trois années :

Tableau N° 09 : la répartition des engagements pendant les trois années

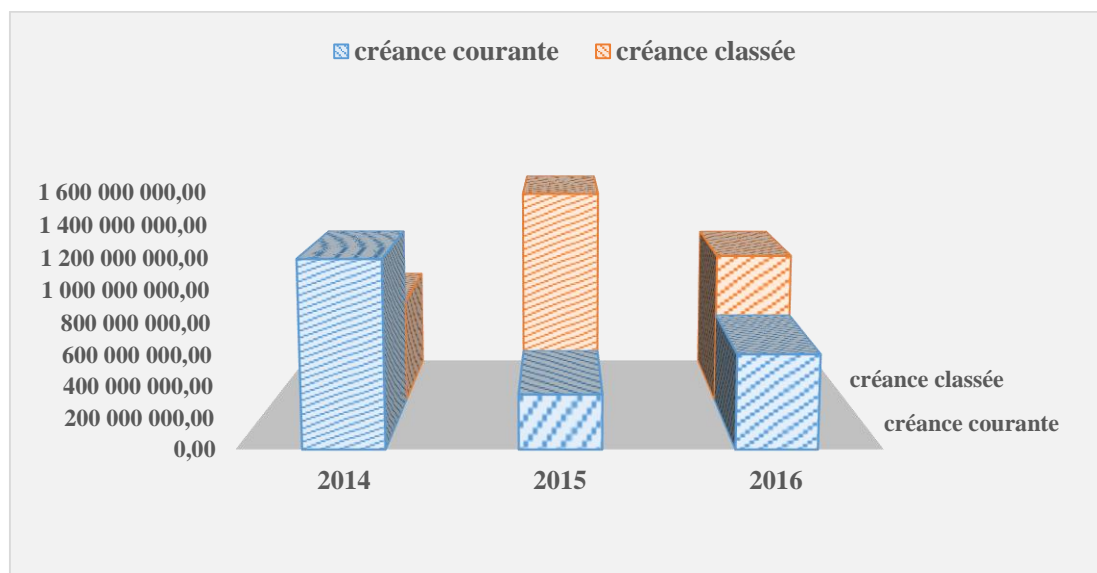
Les créances	2014	2015	2016
Créances courantes	1187944 991,61	347426494,79	600 254 264,20
Créances classées	688 198 181,78	1 446 984 312,24	1017564636,34

Source : la CNEP-banque

A travers le schéma ci-dessous qui résume les données du tableau, on constate que les créances classées prennent de l'ampleur pendant la période de 2014-2016 accompagné évidemment d'une dégradation du volume des créances courantes en dépit d'une légère amélioration pendant l'année 2016. Cette dégradation doit être fortement prise en compte gérée par la banque, du fait que toute augmentation au niveau des créances classées implique une augmentation au niveau des provisions qui sert à couvrir la perte attendue découlant du

portefeuille de la banque et donc une fragilisation par la suite de la performance des fonds propres de la banque.

Figure N° 06 : répartition graphique des engagements de la banque



❖ **Répartition des créances classées :**

Le tableau suivant résume la répartition des créances classées (en volume) au cours des deux années 2015 et 2016 :

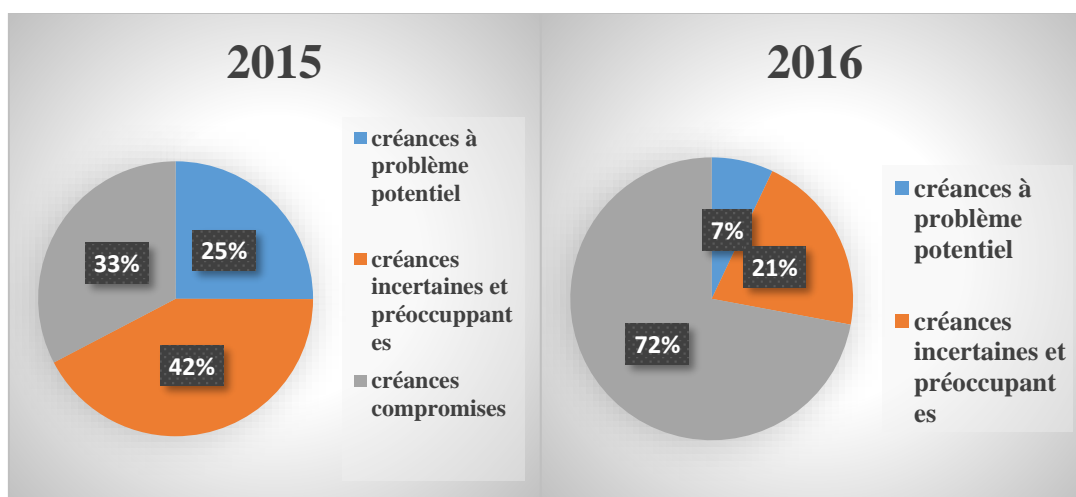
Tableau N° 10 : La répartition des créances classées pendant les deux années

Créances classées	2015	2016
Créances à problème potentiel	362304270,67	71 841 549,81
Créances incertaines et préoccupantes	611942963,22	212 375 045,82
Créances compromises	472737078,35	733 348 040,71

Source : la CNEP-banque

Le schéma ci-dessous montre une transition importante d'une bonne partie des créances courantes et des créances très risquées (incertaines et préoccupantes) de l'année 2015 vers les créances compromises, voire une augmentation de ces dernières de 30% pendant l'année 2016.

Figure N° 07 : répartition graphique des créances classées



En conclusion de cette courte analyse du portefeuille des entreprises de la banque, on constate que le taux de défaillance des entreprises enregistre une augmentation importante que la banque doit gérer par les moyens dont elles disposent et par d'autres moyens plus aptes à faire face au risque de contrepartie, étape que nous allons aborder dans la deuxième section de ce chapitre.

SECTION 2: ELABORATION ET VALIDATION D'UN MODELE LOGIT

1. Etude exploratoire

❖ Population ciblée :

Notre échantillon est constitué d'une base de données de 100 entreprises de type PME issue de la direction du risque crédit de la CNEP-banque. Ces entreprises proviennent de différents secteurs d'activité à savoir :

Tableau N° 11 : La répartition des entreprises par secteur d'activité

Secteur d'activité	Effectif
Industrie	13
BTP	53
Services	13
Autres	21

Plusieurs données quantitatives et qualitatives nécessaires à notre analyse ont été extraites des documents comptables et financiers fiables de l'année 2016, c'est-à-dire N-1 avant la constatation de la défaillance.

On signale que l'accès à ce type de données au sein de la banque reste toujours relativement peu accessible du fait des règles de confidentialité professionnelles liées aux informations collectées auprès de la clientèle.

❖ **L'échantillonnage :**

La construction de notre échantillon a été faite sur la base d'une sélection purement indépendante et donc aléatoires mais équilibrée, c'est-à-dire, le nombre des bonnes et mauvaises entreprises est le même.

Afin de respecter les étapes de l'analyse, nous avons subdivisé notre échantillon de 100 entreprises en deux sous échantillons à savoir :

- L'échantillon apprentissage : c'est celui qui va servir de la construction du modèle LOGIT et est lui-même constitué de 40 bonnes entreprises et 40 mauvaises entreprises
- L'échantillon test : c'est celui qui va servir de la validation du modèle construit à partir de l'échantillon apprentissage afin de tester sa robustesse et sa capacité prédictive sur de nouveaux clients. Il est de même constitué de 20 bonnes entreprises et 20 mauvaises entreprises.

Afin de faciliter l'analyse sur le logiciel, nous avons codifié les modalités de la variable binaire relative à la défaillance notée Y (*classe*) comme suit :

1 : si l'entreprise est « *mauvaise* »

2 : si l'entreprise est « *bonne* »

❖ **Présentation des variables :**

Comme précédemment présenté dans la section 1 du chapitre II, nous avons opté pour un ensemble de variables quantitatives comprenant 20 ratios financiers essentiels pour juger l'activité d'exploitation et financière de l'entreprise et des variables qualitatives comprenant 8 critères extracomptables nécessaires pour renforcer les résultats de l'analyse et s'approcher le plus possible de la réalité. La codification des variables était comme suit :

- *Variables quantitatives* : **Ri** ; *i* allant de 1 jusqu'à 20 financiers. Pour plus de détails sur ces ratios financiers, voir section 2 du chapitre II
- *Variables qualitatives* :
 - Statut juridique
 - Secteur (public, privé)
 - Secteur d'activité (industrie, BTP, services, autres)
 - Type relation (l'entreprises est ancienne ou nouvellement créée)
 - Type crédit (court terme CMT, long terme CLT, promotion immobilière, autres)
 - Climat social (favorable ou défavorable par rapport à la concurrence et à la conjoncture économique)
 - Age débiteur (le dirigeant ou le chef d'entreprise)
 - Classe risque (créance courante, à problème potentiel, incertaine, préoccupante)

❖ **Filtrage et pouvoir discriminant des variables :**

Cette étape consiste à sélectionner les variables les plus discriminantes de la variable Y et étudier leur impact sur elle. Dans une première étape, on va étudier l'effet de chaque variable (quantitative et qualitative) sur la variable dépendante et dans une deuxième étape la corrélation entre les variables exogènes afin de vérifier l'indépendance de ces dernières qui une condition nécessaire pour appliquer la méthode logistique.

Etape 1 :

Pour la première étape, deux tests différents sont appliqués :

- Le test d'Anova à 1 facteur : qui sert à analyser la corrélation qui existe entre une variable quantitative et une variable qualitative de type nominale qui est dans notre cas la variable Y
- Le test de khi-deux : qui sert à analyser la corrélation qui existe entre deux variables qualitatives nominales

Les résultats du test d'Anova fournis par le logiciel SPSS ont donné les constatations suivantes :

Tableau N° 12 : les résultats du test d'ANOVA

Ratios	moyenne des carrés	F	sig
R1	27,198	4,903	0.03
R4	28,262	7,133	0.009
R5	35,562	5,021	0.06
R16	47,332	5,506	0.017

Dans le tableau ci-dessus on a retenu que les ratios qui sont significativement en relation avec la variable « classe » et donc nous avons rejeté évidemment tous les autres ratios qui ne le sont pas. En effet, le test d'Anova permet de comparer les moyennes des deux échantillons « bon » et « mauvais » dont l'hypothèse nulle H_0 est « *les deux moyennes sont égales* » au niveau de 5%. Le seuil de signification de chacun des ratios indiqués dans le tableau est inférieur à 5%, cela revient donc à rejeter H_0 et donc d'affirmer l'existence d'une relation significative entre eux et la variable dépendante.

Nous allons cette fois-ci analyser les corrélations qui existent entre la variable « classe » et chacune des variables qualitatives nominales. Le tableau ci-après indique que les variables dont la corrélation a été affirmée :

Tableau N° 13 : les résultats du test khi-deux

Variabes	valeur	ddl	Sig
Secteur	7,671	1	0.006
Classe risque	13,348	3	0.04
Type crédit	13,340	3	0.000
Type relation	17.802	1	0.000
Climat social	8,996	2	0.011

D'après ce tableau, le test de khi-deux appliqué sur l'ensemble des variables qualitatives dont l'hypothèse nulle H_0 étant « *les deux variables sont indépendantes* » montre que, mis à part les deux variables « *statut juridique* » et « *âge débiteur* », toutes les autres variables présentent une relation significative au seuil de 5% avec la variable « classe ».

Toutefois, on ne peut juger le pouvoir discriminant final de l'ensemble de ces variables sur notre variable Y que si l'on vérifie d'indépendance entre ces variables-là et éliminer le problème

de colinéarité et de redondance, étant donné qu'elle représente une étape cruciale pour notre étude afin d'éviter la falsification des résultats.

Etape 2 :

Dans cette deuxième étape, nous allons étudier la liaison deux à deux qui existe entre chacune des variables exogènes et éliminer toutes celles qui sont fortement liées.

➤ ***La corrélation entre les variables qualitatives :***

A travers toujours le test de khi-deux, on a pu constater comme l'indique le tableau ci-dessous l'indépendance entre les variables « *secteur-type relation-type crédit-climat social* » vu que probabilité de signification est supérieure à 5%.

Néanmoins, il existe une dépendance significative entre « *secteur d'activité-climat social* » et « *secteur d'activité et type crédit* » vu que la probabilité de signification est inférieure à 5%.

Tableau N° 14 : les résultats du test khi-deux

variables	valeur	ddl	sig
Secteur-type crédit	1.544	3	0.672
Secteur-climat social	1.138	2	0.535
Type relation-type crédit	1.835	3	0.607
Type relation-climat social	1.909	2	0.385

➤ ***La matrice de corrélation des variables quantitatives :***

A travers la matrice de corrélation présentée en annexe (*Annexe 01*), nous avons pu conclure les informations suivantes :

- Tous les quatre ratios financiers manifestent une corrélation bivariée très proche de zéro et qui n'est pas significative du fait que leur probabilité est supérieure au seuil de signification 5%, cela revient à considérer une dépendance très négligée entre ces variables.

Tableau N° 15 : les résultats de la matrice de corrélation

		R1	R4	R5	R16
R1	Corrélation de Pearson	1	,0002	,0001	-,0000
	Sig. (bilatérale)		,177	,591	,474
	N	80	80	80	80
R4	Corrélation de Pearson	,0002	1	-,007	-,005
	Sig. (bilatérale)	,177		,532	,269
	N	80	80	80	80
R5	Corrélation de Pearson	,0001	-,007	1	-,003
	Sig. (bilatérale)	,591	,532		,592
	N	80	80	80	80
R16	Corrélation de Pearson	-,0000	-,005	-,003	1
	Sig. (bilatérale)	,474	,269	,592	
	N	80	80	80	80

A partir de l'analyse de corrélation et le filtrage fait sur l'ensemble des variables quantitatives et qualitatives, nous allons intégrer toutes les variables retenues et les intégrer dans la régression logistique en étudiant leur effet simultané sur la variable « classe ».

1. Fonction score et interprétation des résultats :

La régression logistique binaire fait ressortir le tableau suivant qui indique les coefficients estimés ainsi que leur influence sur la probabilité que l'entreprise soit bonne.

Tableau N° 16 : l'estimation des coefficients de la fonction score

variables	C	SE	wald	sig	Exp(C)
R1	,116	,252	,211	,046	1,122
R4	,756	,305	1,790	,018	2,130
R5	,229	,293	,611	,043	,795
R16	-,424	,453	,876	,034	,654
secteur					
privé	-34,528	5,217	1,020	,008	,000
Type relation					
ancienne	12,075	1,005	6,128	,023	175430 ;8102
nouvelle	4,049	0,545	4,868	,009	57,354
Type crédit					
CMT	-2,652	1,841	1,076	,015	,071
CLT	,415	1,610	,066	,007	1,514

promotion	-3,683	1,928	1,649	,026	,025
Climat social					
Favorable	15,268	6,971	1,090	,030	4275630,634
Défavorable	2,114	1,075	1,863	,049	8,279
Classe risque					
Créance courante	22,389	3,493	3,925	,004	5,291E9
Pb potentiel	19,802	7,493	1,502	,013	3,978E8
incertaine	-18,963	5,493	2,760	,048	1,720E8
Constante	13,623	2,163	3,001	,011	824767,625

❖ **Interprétation financière des coefficients :**

Autonomie financière R1 : elle représente la capacité de l'entreprise à financer ses projets par ses propres fonds, son coefficient est de signe positif, cela veut dire que plus l'autonomie financière est grande moins l'entreprise sera dépendante de ses dettes et moins la probabilité de défaut surviendra.

Capacité de remboursement R4 : elle représente le degré de l'entreprise à honorer ses engagements à moyen et long terme à partir de sa capacité d'autofinancement mesurée par son résultat net dégagée au cours de son exploitation, ajouté aux dotations aux amortissement. Le signe positif indique que plus la capacité de remboursement est grande, plus le score de l'entreprise sera élevé et donc plus sa probabilité d'être « bonne » sera élevée

Financement des emplois stables R5 : ce ratio permet de mesurer l'excès des capitaux propres d'une entreprise après avoir financé son actif immobilisé. Le signe positif indique plus cet excès de capitaux est élevé, plus l'entreprise sera apte à financer son cycle d'exploitation et mener normalement son activité pour pouvoir rembourser ses dettes aux échéances fixées. Cela donc conduira à augmenter sa probabilité d'être classée « bonne ».

Frais de personnel R16 : ce ratio présente un signe négatif, ce qui est tout à fait logique vu que les frais de personne représente une charge pour l'entreprise et son augmentation conduira à baisser son excédent brut d'exploitation et donc baisser ses cash flows dont une partie servira de rembourser ses dettes. Cela va donc conduire à baisser son score et donc sa probabilité d'être classée « bonne ».

Secteur : comme déjà vu à partir de l'analyse de la corrélation, le type de secteur est une variable discriminante qui peut représenter un facteur important pour classer une entreprise. En effet, le secteur privé comme indiqué sur le tableau présente un signe négatif du fait qu'une entreprise privée présente plus de risque de défaut qu'une entreprise publique qui est couverte

par l'Etat. Comme la grande majorité des entreprises proviennent du secteur privé, le traitement de cette variable a été mis en priorité par rapport au secteur public dont son coefficient a été négligé.

Type relation : le type de relation ici se veut dire que l'entreprise est une ancienne création qui dispose déjà historiquement de documents comptables et bilans de son activité ou bien une nouvelle création qui vient de lancer son activité. Ici la variable « nouvelle » présente un signe positif mais nettement inférieur en poids à celui de la variable « ancienne » qui représente son double, ce qui signifie qu'une entreprise ancienne a plus de chance d'être classé « bonne » qu'une entreprise nouvellement créée.

Type crédit : on distingue trois catégories de types de crédit, le CMT qui est un crédit à court terme, le CLT qui présente le crédit à long terme et la promotion immobilière qui est un type assez spécifique à la CNEP-banque. Le signe des crédits à court terme et la promotion est négatif, cela peut être expliqué par fait que ce type de crédit possède des échéances courtes et donc un appel rapide aux entreprises pour honorer leurs échéances. Toutefois, les entreprises passent des fois durant leur cycle d'exploitation par des crises d'activité temporaires ou bien durables, ce qui les empêchent de respecter l'échéancier fixé par la banque. Seules quelques entreprises arrivent à s'en remettre et rééquilibrent leurs tableaux de remboursement. Contrairement aux crédits à long terme qui donne plus de flexibilité aux entreprises pour honorer leurs échéances.

Climat social : l'activité des entreprises est fortement liée au climat social dans lequel elles l'exercent ; cela comprend la dynamique du secteur d'activité, la concurrence, la conjoncture économique qui influe souvent sur le prix de vente du produit ou du service. Ici, il est clair que le climat favorable influe positivement sur la santé de l'entreprise et donc sur sa capacité à rembourser ses dettes envers la banque d'où le signe positif et le poids porté par cette variable. Quand le climat est défavorable, le signe reste positif mais en faible poids.

Classe risque : elle représente la classe liée au retard de paiement des échéances. Le tableau indique un signe positif jusqu'à un certain seuil de retard de paiement. Le premier signe positif pour la classe courante, ce qui est logique du fait qu'une entreprise qui n'a fait aucun retard de paiement reflète un bon signe en vis-à-vis du remboursement de sa créance et une bonne santé de cette entreprise. Le deuxième signe positif pour la classe à problème potentiel dont le retard de paiement est inférieur à 90 jours. Néanmoins, elle reflète une santé temporairement perturbée de l'entreprise. Enfin, la classe incertaine qui elle présente un signe négatif, cela dit que le

modèle considère que tout un retard de paiement supérieur à 90 jours présente un mauvais signe pour la santé de l'entreprise et donc accroît sa probabilité de défaut.

A ce niveau-là, on peut écrire la fonction score du modèle LOGIT estimé comme ceci :

$$\ln(Y) = 13.623 + 0.116 \mathbf{R1} + 0.756 \mathbf{R4} + 0.229 \mathbf{R5} - 0.424 \mathbf{R16} - 34.528 \mathbf{privé} + 12.075 \mathbf{ancienne} + 4.049 \mathbf{nouvelle} - 2.652 \mathbf{CMT} + 0.415 \mathbf{CLT} - 3.683 \mathbf{promotion} + 15.268 \mathbf{favorable} + 2.114 \mathbf{défavorable} + 22.389 \mathbf{courante} + 19.802 \mathbf{pb\ potentiel} - 18.968 \mathbf{préoccupante}$$

2. Evaluation et validité du modèle :

❖ La signification des coefficients par le test de Wald :

Comme précédemment dit, dans la partie théorique, le test de wald permet de vérifier la signification des coefficients estimés de chaque variable au niveau de confiance choisi. D'après le tableau ci-dessus, on a pu constater que les variables filtrées par le modèle possèdent des coefficients à une probabilité inférieure à 5%, ce qui revient à rejeter l'hypothèse nulle du test qui stipule la nullité du coefficient en question. Cela revient à dire que l'ensemble de ces variables sont significativement discriminantes de la variable « classe ».

❖ Qualité globale du modèle :

• Le test du rapport de vraisemblance :

La statistique du test khi-deux, fournit une probabilité inférieure à 5%, cela veut dire que le modèle avec variables permet de mieux expliquer la variable « classe » que le modèle sans variables et donc un pouvoir discriminant significatif des variables introduites dans notre modèle.

Tableau N° 17 : le test du rapport de vraisemblance

Test	Valeur khi-deux	ddl	sig
modèle	55.289	15	0.000

• Le coefficient de McFadden :

Par ailleurs, on a aussi d'autres indicateurs d'évaluation du modèle résumé dans le tableau ci-dessous qui indique qui taux satisfaisant de McFadden à 55.1% et une relation intense entre les variables explicatives et la variable à expliquer mesurée par le R² de Nagelkerke.

Tableau N° 18 : les indicateurs d'évaluation du modèle

Cox et Snell	0.501
Nagelkerke	0.668
McFadden	0.551

- **Statistique de Hosmer Lemeshow :**

Un autre test permettant de juger la qualité du modèle en mesurant la distance entre les valeurs observées et mesurées par le modèle notée D.

D'après le tableau ci-dessous, on a la probabilité de la statistique de khi-deux est supérieure à 5% ce qui revient à accepter l'hypothèse nulle du test qui stipule que la distance D est petite et donc bon ajustement aux données.

Tableau N° 19 : le test de Hosmer Lemeshow

Etape	Khi-deux	ddl	Sig
1	3.321	8	0.913

- ❖ **Validation et performance du modèle :**

La validation du modèle représente une étape très importante dans la régression logistique, elle implique la capacité du modèle à prédire les bonnes entreprises parmi les mauvaises entreprises. Le tableau de classement dans l'échantillon d'apprentissage ci-dessous permet d'analyser ceci :

Tableau N° 20 : la matrice de classement dans l'échantillon d'apprentissage

Niveau Observé	Niveau Prédit		
	MAUVAIS	BON	Pourcentage correct
MAUVAIS	32	8	80,0%
BON	10	30	75,0%
Pourcentage global			77,5%

D'après cette matrice de confusion on peut constater les informations suivantes :

- Taux de bon classement : $(32+30) / 80 = 77.5\%$
- Taux de mauvais classement : $(8 + 10) / 80 = 22.5\%$
- La sensibilité : $30 / (30+10) = 75\%$
- La précision : $30 / (30+8) = 78.94\%$
- La spécificité : $32 / (32+8) = 80\%$

D'après les informations constatées, le modèle est capable de classer 77.5% des entreprises dans leur bonne classe. Ainsi, une entreprise saine a 75% de chance de rembourser totalement son crédit et une entreprise défaillante a 80% de chance de faire défaut.

- **Validation croisée dans l'échantillon test :**

La validation croisée dans l'échantillon test donne les résultats suivants :

Tableau N° 21 : la matrice de classement dans l'échantillon de test

Observations		Prévisions		
		classe		Pourcentage correct
		MAUVAIS	BON	
classe	MAUVAIS	9	1	90,0
	BON	1	9	90,0
Pourcentage global				90,0

- Taux de bon classement : $(9+9) / 20 = 90\%$
- Taux de mauvais classement : $(1+1) / 20 = 10\%$

Le modèle construit, permet de classer correctement de nouvelles entreprises à 90% dans leur classe, ce qui est très satisfaisant comme taux de succès et indique une bonne performance du modèle.

- **Le résidu de Pearson :**

C'est un test qui permet d'expliciter tous les points qui ont été mal modélisés dont la statistique suit la loi normale notée Z. en effet, toute observation dont le Zrésidu dépasse plus ou moins 2 est considérée comme mal classée par le test :

Tableau N° 22 : le test du résidu de pearson

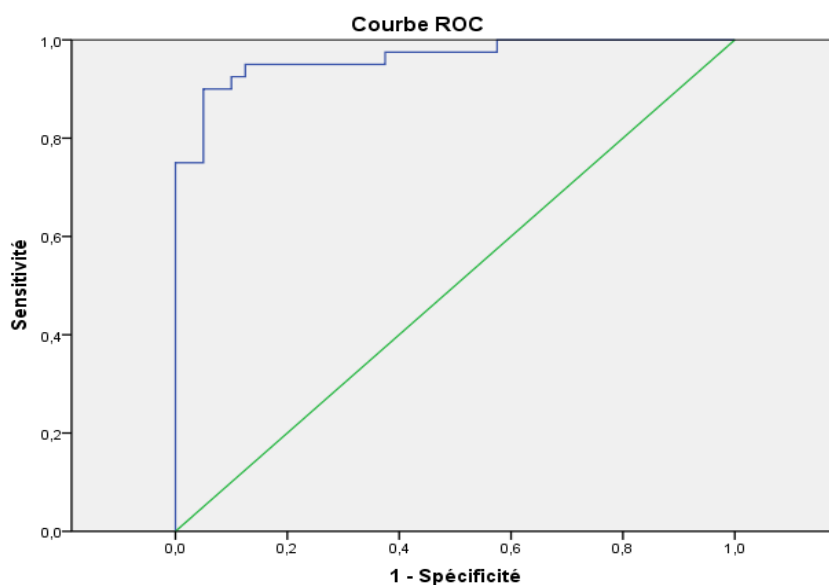
Observation	Etat sélectionné	Observations	Prévisions	Groupe prédit	Variable temporaire	
		classe			Resid	ZResid
13	S	B	,037	M	,963	5,113

D'après le tableau, on remarque que la seule observation mal classée selon le test de résidu de Pearson dans l'échantillon test est l'observation n° 13 car son Zrésidu est supérieur à 2, ce qui reflète une erreur de modélisation réduite.

- **La courbe ROC :**

La courbe ROC du modèle est représentée comme ceci, elle indique une surface largement supérieure à la moyenne d'une valeur égale à $AUC = 0.96$ qui très proche de 1, ce qui indique encore une fois une très bonne performance du modèle.

Figure N° 08 : la courbe ROC du modèle



SECTION 3 : APPLICATION DE LA METHODE LINEAIRE ET COMPARAISON DES DEUX METHODES

Dans cette section, nous allons construire à partir des mêmes variables introduites dans la régression logistique afin d'élaborer la fonction score résultant de la méthode linéaire.

1. Transformation numérique des variables qualitatives par Disqual :

Comme a été précédemment expliqué dans la partie théorique, le modèle issu de l'analyse discriminante linéaire ne peut contenir des variables qualitatives, de ce fait, on va transformer l'ensemble de ces variables en des variables numériques avant, de les introduire dans la modélisation, à travers la méthode Disqual (pour plus de détails, voir la section 1 du chapitre II).

Cette transformation aboutit au final à avoir non pas neuf variables comme dans le cas de la régression logistique mais 18 variables numériques y compris les variables qualitatives transformées.

2. La sélection des variables et pouvoir discriminant :

Le filtrage des variables les plus discriminantes se fait par la méthode pas à pas (stepwise) à travers le test de Lambda de Wilks qui suit une loi de Fisher. Le principe de la sélection se base sur la minimisation de cette statistique en introduisant dans le modèle que les variables qui minimisent cette statistique étape par étape.

Une variable est acceptée quand le seuil de signification est inférieur à 5% et elle est rejetée quand ce seuil dépasse 10%.

Le tableau ci-dessous fournit les résultats de ce test :

Tableau N° 23 : le pouvoir discriminant des variables (lambda de wilks)

	Lambda de Wilks	F	Signification
R16	,881	4,903	,030
FAVORABLE	,766	7,133	,009
PROMOTION	,730	5,906	,017
ANCIENNE	,704	8,273	,005
PRIVE	,677	22,325	,000
NOUVELLE	,657	22,325	,000
CMT	,650	4,119	,046
CLT	,625	11,175	,001
DEFAVORABLE	,607	10,894	,001
PB POTENTIEL	,584	9,270	,003
R4	,544	9,270	,003
R1	,535	6,751	,011
PREOCCUPANT	,523	6,478	,013

D'après le tableau, on constate que la statistique Lambda de Wilks dont la valeur est très grande a été écartée pour les variables « *R5, autres, courant, public et incertain* ». En effet, toutes les variables présentent un pouvoir discriminant significatif qui implique leur introduction à l'étape de la modélisation puisque la probabilité de signification est inférieure à 5%. Elles sont classées par ordre du moins discriminant au plus discriminant, cela fournit une information en plus quant à l'importance de chacune de ces variables.

3. La fonction score et interprétation des résultats :

A partir de l'estimation des coefficients des variables, on a abouti à la fonction score notée Z suivante :

$$Z = -0.286 + 0.173 \mathbf{R1} + 0.295 \mathbf{R4} - 0.162\mathbf{R16} - 0.335 \mathbf{PRIVE} + 0.601\mathbf{ANCIENNE} + 0.201\mathbf{NOUVELLE} - 0.372\mathbf{CMT} + 0.092 \mathbf{CLT} - 0.586 \mathbf{PROMOTION} + 0.405 \mathbf{FAVORABLE} + 0.102\mathbf{DEFAVORABLE} - 0.443\mathbf{PREOCCUPANT} + 0.253 \mathbf{PB POTENTIEL}$$

On remarque que le signe des coefficients des variables sont compatibles avec ceux obtenus par la régression logistique, d'où l'interprétation financière de chaque coefficient reste la même.

A partir de cette fonction score, on a pu calculer la moyenne des scores pour chacun des deux groupes appelé les barycentres :

Tableau N° 24 : les barycentres de la fonction score

Groupe	Score moyen
1	S1 = -1.241
2	S2 = 1.241

Ces deux scores moyens permettent de calculer le seuil d'affectation des entreprises dont la formule est :

$$\text{Seuil d'affectation} = \frac{S1 + S2}{2} = 0$$

A ce niveau, la règle d'affectation devient simple, l'entreprise ayant un score inférieur à zéro est considérée mauvaise (défaillante) et celle ayant un score supérieur à zéro est considérée comme bonne (saine)

4. L'évaluation et la validation du modèle :

L'évaluation de la fonction score se fait à travers différents indicateurs, notamment le test de lambda de Wilks, la valeur propre et la corrélation canonique.

❖ Le test de lambda de Wilks :

C'est un test qui permet de vérifier s'il existe une égalité de moyennes entre les deux groupes d'affectation, il permet ainsi de renseigner mieux sur le pouvoir discriminant des variables du modèle. Son hypothèse nulle H_0 est « *les moyennes sont égales* ».

D'après le tableau ci-dessous, on remarque que le seuil de signification est inférieur à 5%, cela veut dire que H_0 est rejetée et donc une différence des deux moyennes.

Tableau N° 25 : le test de Lambda de Wilks

Lambda de Wilks	Khi-deux	ddl	Signification
0,388	67,745	13	0,000

❖ **La corrélation canonique :**

La corrélation canonique renseigne sur la bonne séparation des deux groupes d'affectation, elle est de 78% ce qui indique une bonne qualité du modèle et du pouvoir discriminant de la fonction score élaborée.

Tableau N° 26 : la corrélation canonique

Fonction	Corrélation canonique
1	0,782

❖ **La matrice de bon classement du modèle :**

La matrice de confusion du modèle est représentée comme ceci :

Tableau N° 27 : la matrice de classement du modèle

CLASSE	Classe(s) d'affectation prévue(s)		Total
	1	2	
1	33	7	0.825
2	6	34	0.85
Taux de bon classement			83.75%

D'après cette matrice de confusion on peut constater les informations suivantes :

- Taux de bon classement : $(33+34) / 80 = 83.75\%$
- Taux de mauvais classement : $(6+ 7) / 80 = 16.25\%$
- La sensibilité : $34 / (34+6) = 85\%$
- La précision : $34 / (34+7) = 83.92\%$
- La spécificité : $33 / (33+7) = 82.5\%$

D'après les informations constatées, le modèle est capable de classer 83.75% des entreprises dans leur bonne classe. Ainsi, une entreprise saine a 85% de chance de rembourser totalement son crédit et une entreprise défaillante a 82.5% de chance de faire défaut.

❖ **La validation croisée :**

La validation croisée sur l'échantillon test donne la matrice de confusion suivante :

Tableau N° 28 : la matrice de classement dans l'échantillon de test

CLASSE		Classe(s) d'affectation prévue(s)		Total
		1	2	
2	1	10	0	100%
	2	1	9	90%
Taux de bon classement				95%

- Taux de bon classement : $(10+9) / 20 = 95\%$
- Taux de mauvais classement : $(1+0) / 20 = 5\%$

D'après les résultats de la matrice de confusion, on constate que le modèle score élaboré permet de prédire correctement la classe de nouvelles entreprises hors l'échantillon de construction à hauteur de 95% ce qui est un résultat très satisfaisant.

❖ **Le test de performance de la fonction score : Q press**

Ce test permet de juger la performance du modèle par rapport au Hazard, il vérifie si les résultats du modèle score élaboré sont significativement mieux expliqués que le Hazard.

La statistique de test suit une loi de khi-deux dont la valeur est de 36.45 qui est supérieure à la valeur tabulée de khi-deux à un 1 degré de liberté. On en constate que le modèle construit classe mieux que le hazard les entreprises.

Tableau N° 29 : le test de Q press

Q press	ddl	Sig
36.45	1	3.841

❖ **La courbe de performance : ROC**

Le calcul de la surface de la courbe de ROC située au-dessus de la première bissectrice donne une valeur de AUC égale à 0.98 qui est une valeur très proche de 1.

Cela signifie que le modèle score construit par l'analyse discriminante linéaire est assez performant pour expliquer le phénomène de défaillance des entreprises et les classer correctement.

5. La comparaison de résultats des deux méthodes :

A partir des résultats fournis par les deux méthodes empiriques, nous allons comparer les principaux indicateurs de performance de chacune d'elle dans le but de mettre en lumière les atouts de la méthode logistique par rapport à l'analyse discriminante linéaire et expliquer les différences de modélisation entre les deux méthodes.

❖ La sélection des variables :

Tableau N° 30 : la comparaison des coefficients des deux méthodes

Variables	Coefficients	
	ADL	RL
Constante	-0.286	13.623
Autonomie financière	0.173	0.116
Capacité de remboursement	0.295	0.756
Financement des emplois stables	-	0.229
Frais de personnel	-0.162	-0.424
Secteur privé	-0.335	-34.528
Relation ancienne	0.601	12.075
Relation nouvelle	0.201	4.049
Crédit court terme	-0.372	-2.652
Crédit long terme	0.092	0.415
Crédit Promotion immobilière	-0.586	-3.683
Climat favorable	0.405	15.268
Climat défavorable	0.102	2.114
Créance courante	-	22.389
Créance à problème potentiel	0.253	19.802
Créance préoccupante	-0.443	-18.968

D'après le tableau ci-dessus, on remarque que les deux méthodes paramétriques fournissent des coefficients du même signe pour chaque variable, ce qui affirme la compatibilité de la sélection des variables ainsi que le pouvoir discriminante des variables choisies.

En revanche, quoiqu'il semble que la méthode logistique accorde plus de poids à certaines variables par rapport à l'analyse discriminante linéaire, on ne peut comparer les valeurs absolues la valeur des coefficients estimés, du fait que chaque méthode adopte des hypothèses et des techniques d'estimation différente : l'ADL se base sur la fonction de Fisher et la RL se base sur la méthode de vraisemblance.

Cette différence va aboutir à deux seuils de discrimination différents et donc deux critères d'affectation différents, ce qui explique le fait que certaines variables ont été exclusivement sélectionnées par la RL et non par l'ADL.

❖ **La validation croisée et le taux de bon classement :**

Les résultats de la validation croisée des deux méthodes ont été résumés dans le tableau suivant :

Tableau N° 31 : la comparaison des taux de classement des deux méthodes

	ADL	RL
Le taux de bon classement	83.75%	77.5%
Le taux de mauvais classement	16.25%	22.5%
La sensibilité	85%	75%
La précision	83.92%	78.94%
La spécificité	82.5%	80%

D'après le tableau, on remarque que les deux méthodes fournissent des taux très satisfaisants quant à la classification des entreprises, quoique l'ADL semble fournir une meilleure classification.

En revanche, à travers l'analyse des deux méthodes, la RL présente l'avantage de la facilité d'exécution par rapport aux hypothèses de travail notamment la normalité de l'échantillon et la transformation numérique des variables qualitatives exigées dans l'ADL.

CONCLUSION :

Dans ce dernier chapitre, on a pu montrer l'intérêt que porte la méthode logistique dans la classification des entreprises, tout en explicitant les critères financiers et qualitatifs ayant une forte capacité de prédiction et un pouvoir discriminant très élevé dont le taux de bon classement était de 77.5%.

On a ensuite comparé la performance de cette méthode à celle de l'analyse discriminante linéaire, où on a pu constater que les deux méthodes possèdent un pouvoir discriminant très élevé, quoique la méthode logistique présente quelques avantages et atouts notamment particulièrement liés aux hypothèses d'application de la méthode.

CONCLUSION GENERALE

A la lumière de ce travail dont l'objectif était de développer un modèle de crédit scoring permettant de prédire la santé d'une nouvelle entreprise ainsi que sa probabilité de défaut, les résultats obtenus montrent qu'à travers la méthode logistique étant l'outil principal de notre analyse, le taux de bonne prédiction atteint 77.5%. Ce taux comparé à celui fourni par la méthode linéaire (étant de 83.75%) n'a fait que prouver que les deux outils possèdent un bon pouvoir prédictif.

Néanmoins, la méthode logistique présente plusieurs atouts, notamment la simplicité d'application ce qui n'est pas le cas pour la méthode linéaire qui requiert plusieurs conditions particulièrement la normalité des variables explicatives et le choix quantitatif des variables. En outre, la méthode logistique présente l'avantage dans le sens où les valeurs de la variable à expliquer peuvent être considérées comme une probabilité puisqu'elles se situent toutes entre 0 et 1 ce qui n'est pas valable pour l'autre méthode.

Cependant, plusieurs limites peuvent être attribuées à cette recherche qui empêcheront de mettre en œuvre le modèle score élaboré au sein de la banque, particulièrement celles liées à la taille de l'échantillon qui est assez petite pour généraliser les résultats de la recherche et aussi au manque d'informations relatives aux caractéristiques de la banque et sa stratégie d'optimisation d'octroi de crédit. En revanche, la méthode des scores reste un outil assez efficace pour faire face au risque de défaut des demandeurs de crédit qui a porté ses fruits dans plusieurs banques dans le monde.

Il faut signaler que cette méthode ne contrarie pas la méthode de gestion classique, elle est au contraire complémentaire et compatible avec les analyses par les ratios du fait que cette dernière constitue une étape primaire dans la modélisation scoring. Ainsi, dans une perspective d'introduire la méthode scoring dans la banque algérienne une question découle de cela :

« Faut-il ignorer l'effort des professionnels mis en œuvre dans la gestion classique du risque de crédit et se fier uniquement à une information chiffrée fournie par l'outil scoring ou bien trouver un seuil combinatoire entre les deux méthodes de gestion afin d'optimiser la qualité de la prise de décision ? »

Enfin, la méthode par les scores reste l'une des fréquentes techniques utilisées dans les banques dans le monde que la banque algérienne doit prendre en considération de l'intégrer dans son système de gestion des risques, quoiqu'il existe plusieurs autres techniques et c'est à la banque de choisir celle qui convient le mieux à sa stratégie.