



ROYAUME DU MAROC

HAUT COMMISSARIAT AU PLAN

INSTITUT NATIONAL DE STATISTIQUE ET D'ECONOMIE APPLIQUEE
INSEA



Projet de Fin d'Etudes

*_*_*_*_*_*_*

MODELISATION DE LA PROBABILITÉ DE DÉFAUT SOUS LA NORME IFRS9

Préparé par :

- M. BELHAJ Imad
- Mme. MERZOUQ Manal

Sous la direction de :

- M.SAID Khalil
- M.Harmouch Jamal

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

Filière : Actuariat-Finance

Devant le jury composé de :

-
-
- M. SAID Khalil (INSEA)
 - M. MARRI Fouad (INSEA)
 - M. HARMOUCH Jamal (Fidaroc Grant Thornton)
-
-

Année Universitaire : 2019/2020



Résumé

La crise financière de 2008 a mis au jour les coûts systémiques d'une reconnaissance tardive des pertes de crédit de la part des banques et d'autres prêteurs. Avant la crise, les banques ne faisaient pas un provisionnement adéquat des pertes de crédit que les risques émergeant étaient susceptibles d'entraîner. En effet, la méthode de dépréciation dans IAS 39 reposait sur une approche à posteriori, ne reflétant dans les états financiers que les conséquences d'évènements déjà produits dans le passé.

À la suite de cette crise, les instances de normalisation comptable, notamment l'International Accounting Standards Board (IASB), ont publié en 2014 le document IFRS 9 Financial Instruments. IFRS 9 exige que les prêts consentis par les banques et autres établissements fassent l'objet de provisions fondées sur les pertes de crédit attendues. Les modèles sont impactés par des facteurs de positionnement dans le cycle économique et des paramètres de sensibilité macroéconomique estimés via des modèles économétriques pour la considération de l'information future dite Forward Looking.

Notre travail repose sur l'estimation de certains paramètres qui permettent de calculer la provision pour le risque de crédit qui sont entre autre la probabilité de défaut (PD) et le taux de perte en cas de défaut (LGD) au sens IFRS 9.

Nous proposons l'utilisation de la régression logistique. Ils'agit d'une technique demodélisation qui, dans sa version la plus répandue, vise à prédire les valeurs d'une variable catégorielle binaire Y (variable à prédire) à partir d'une collection de variables X continues ou catégorielles (variables prédictives). Elle fait partie des méthodes d'apprentissage supervisé ; elle peut s'inscrire dans le cadre de la régression linéaire généralisée. Nous utilisons aussi le modèles de survie comme celui de «Cox» qui permet de prédire le temps de survie du crédit en d'autre terme le temps où le client va tomber en défaut.

Mots clés : IFRS 9, Forward Looking, Perte attendue, Machine learning , Modèle de régression logistique , PD , Cox

Remerciements

Au terme de ce travail, nous remercions tout d'abord M.SAID Khalil, pour nous avoir fait l'honneur de nous encadrer, pour son implication, pour ses conseils avisés tout au long de notre stage.

Nos remerciements s'adressent également à notre encadrante externe, M. HARMOUCH Jamal pour sa constante disponibilité, ses efforts et son encouragement qui ont fait de notre stage une expérience enrichissante.

Nous tenons à remercier M. MARRI Fouad, professeur l'INSEA pour avoir accepté d'évaluer ce travail.

Nous adressons nos remerciements à l'ensemble du corps professoral de l'INSEA et notre Directeur SKALLI Ahmed pour leurs efforts continus afin de nous offrir la meilleure formation, ainsi que toute autre personne ayant participé, la réussite de notre projet de fin d'études.

Dédicace Imad

Je dédie ce travail, à mes chers parents Aicha Assou et Mohamed Belhaj. Un travail qui sans leur aide, leur patience, et leurs sacrifices, ne sera réussi.

A mes meilleurs amis pour leur tendresse, leur soutien moral, et leurs encouragements ;

A mon frère Ayoub Belhaj pour son amour, son appui et son temps ;

A mes enseignants pour tout le savoir qu'ils nous ont transmis, et à qui j'espère être à la hauteur de leur espérance.

Merci infiniment.

Dédicace Manal

Je dédie ce travail, à mes chers parents Nouzha Benaoui et Rachid Merzouq qui ont toujours été là pour moi.

A mes meilleurs amis pour leur soutien moral, et leurs encouragements ;

A mes chères sœurs Hind et Meriem qui ont toujours su me remonter le moral ;

A mes enseignants pour tout le savoir qu'ils nous ont transmis, et à qui j'espère être à la hauteur de leur espérance.

Merci infiniment

Table des matières

Résumé	3
Remerciements	4
Dédicace Imad	5
Dédicace Manal	6
1 Organisme d'accueil et contexte général du PFE	12
1.1 Présentation de l'organisme d'accueil	13
1.2 Le comité des associés de Grant Thornton	14
1.3 Les différents services offerts par l'organisme	15
1.4 Contexte général du stage	15
2 L'accord de Norme IFRS 9 et Risque de Crédit	16
2.1 La Norme IAS 39 : « "Instruments financiers : comptabilisation et évaluation" . . .	19
2.1.1 La norme IAS 39 : Définition	19
2.1.2 La norme IAS 39 : Enjeux, Objectifs et Champs d'application	20
2.2 Norme IFRS 9 (International Financial Reporting Standards 9)	21
2.2.1 Chronologie :	21
2.3 Risque de Crédit :	28
2.4 les composants du risque crédit :	28
2.5 Mesure du risque crédit :	30
2.6 Analyse du risque crédit en pratique :	31
3 Modèle monopériodiques de probabilité de défaut PD	33
3.1 Analyse et traitement de la base de donnée	34
3.1.1 Les variables propres au client	34
3.1.2 Analyse descriptive	35
3.1.3 Vérification des valeurs manquantes	37
3.1.4 Détection des valeurs aberrantes (BOXPLOT)	38
3.1.5 Test d'indépendance : KHI2	40
3.1.6 Mesure des corrélations entre les variables quantitatives	40
3.2 Modélisation de la probabilité de défaut PD	41

3.2.1	Méthode de Régression logistique	44
3.2.2	Méthode de Machine Learning : Arbre de décision et Forêts aléatoires	54
3.2.3	Choix du meilleur modèle	62
4	Modèle multipériodiques de probabilité de défaut PD	64
4.1	Modèles de risque en temps discret	65
4.1.1	L'événement par défaut	65
4.1.2	Les modèles non lineaire : Probit	65
4.1.3	Application notre base de donnée	66
4.2	Modèles de risque en temps continu	69
4.2.1	Modèle de survie : COX	69
4.2.2	Application à notre portefeuille	72
	Conclusion générale	79
A	Annexe	81
	Bibliographie	84

Liste des figures

1.1	Le comité des associés de Grant Thornton	14
2.1	Les 3 niveaux de comptabilisation de la provision risque crédit sous les normes IFRS.9	22
2.2	Les trois composantes de la norme IFRS 9	24
3.1	Tableau des effectifs des bons et mauvais emprunteurs	35
3.2	Table de contingence entre Job de client et défaut de remboursement	35
3.3	La distribution de la variable Job par rapport la variable Defaut	36
3.4	Statistiques élémentaires des variables quantitatives	36
3.5	Détection des valeurs manquantes(variables qualitatives)	37
3.6	Détection des valeurs manquantes (variables quantitatives)	37
3.7	Statistiques élémentaires des variables quantitatives	38
3.8	Boxplot de la variable "Age"	38
3.9	Boxplot de la variable "Duration"	39
3.10	Boxplot de la variable "Amount"	39
3.11	Test KHI2 : défaut*variables catégorielles	40
3.12	Détection de corrélation entre les variables quantitatives	41
3.13	Présentation des résultats du modèle LOGIT pour les particuliers.	50
3.14	Modèle logit : Analyse des effets Type 3.	51
3.15	Modèle logit : Estimation par l'analyse du maximum de vraisemblance.	51
3.16	Modèle logit : Estimation des rapports de cotes.	52
3.17	Modèle logit : AUC et Indice GINI.	53
3.18	Modèle logit : Courbe ROC pour toutes les étapes de création de modèles.	53
3.19	Modèle logit : Adéquation du modèle.	54
3.20	Arbres de décision	59
3.21	Arbres de décision : Courbe ROC	60
3.22	Arbres de décision : index ROC et coefficient de Gini	60
3.23	Forêts aléatoires : Selection des variables	61
3.24	Forêts aléatoires : Courbe ROC	61
3.25	Forêts aléatoires :index ROC et coefficient de Gini	62

3.26	Courbes ROC : comparaison des modèles	62
3.27	Indice Gini :Évaluation des modèles	63
4.1	Modèle probit	67
4.2	Modèle probit : PIT	68
4.3	Aperçu de la base de données initiale	72
4.4	Aperçu de la base de données après traitement des données	73
4.5	Aperçu de la nouvelle variable « duree »	73
4.6	Sortie de la fonction « PROC PHREG » montrant les coefficients du modèle	74
4.7	Graphe montrant la probabilité de survie en fonction de la durée du crédit	74
4.8	Code utilisé pour générer les deux variables définissant le début et la fin de chaque épisode	75
4.9	Modèle de survie CPH covariables variant dans le temps	76
4.10	aperçu des probabilités de survie du client 1	76
4.11	Aperçu des Probabilité de défaut pour des durées allant de 1 mois à 44 mois	77
4.12	Sortie des résidus de Schoenfeld de notre modèle	78
A.1	Information value : Age, Duration et Amount	81
A.2	Information value : Checkingstatus1, History, Property, Purpose et Savings	81
A.3	Tables de contingence entre défaut et Property, Purpose et Housings	82
A.4	Tables de contingence entre défaut et Checkingstatus1 et History	82
A.5	Code utilisé pour la modélisation LOGIT	82
A.6	Code utilisé pour la modélisation PROBIT	82
A.7	Code 1 utilisé pour la modélisation par durée de vie	83
A.8	Code 2 utilisé pour la modélisation par durée de vie	83
A.9	Code 3 utilisé pour la modélisation par durée de vie	84

Liste des tableaux

3.1	Les variables propres au client	34
3.2	interprétation de l'aire sous la courbe ROC	47

Chapitre 1

Organisme d'accueil et contexte général du PFE

Sommaire

1.1	Présentation de l'organisme d'accueil	13
1.2	Le comité des associés de Grant Thornton	14
1.3	Les différents services offerts par l'organisme	15
1.4	Contexte général du stage	15

1.1 Présentation de l'organisme d'accueil

Dans le cadre de notre stage PFE, on a intégré le cabinet d'audit de Fidaroc Grant Thornton basé à Casablanca.

Grant Thornton, qui est un groupe spécialisé en audit, conseil financier et en expertise conseil, appartient à une organisation mondiale déployée dans 147 pays. Il a cependant intégré le marché national depuis près de 26 ans en émanant de la Société Fiduciaire du Maroc en 1994, avant de rejoindre l'organisation mondiale Grant Thornton en 2005. Sans oublier que chaque membre du réseau Grant Thornton est indépendant au plan juridique, financier et managérial.

Vu comme l'une des principales organisations mondiales d'audit et de conseil, Grant Thornton jouit de 40 000 collaborateurs et d'un chiffre d'affaires de près de 4,7 milliards de dollars en 2014.

Ce dernier allie son savoir-faire mondial à une approche locale adaptée afin de délivrer des services d'audit et de conseil à forte valeur ajoutée.

Notons quelques dates clés dans le parcours de Grant Thornton au Maroc :

- 1929 : Création de la Société Fiduciaire du Maroc offrant les activités du Conseil Juridique, Fiscal, Social et l'expertise comptable
- 1994 : Création de Fidaroc, émanant de la SFM, pour couvrir de façon distincte les services d'audit et commissariat aux comptes
- 2005 : Fidaroc rejoint l'organisation mondiale Grant Thornton
- 2008 : Une division Conseil est mise en place.
- 2010 : Fidaroc Grant Thornton inaugure un bureau à Marrakech, impulsant le développement de secteurs d'activité phares
- 2011 : Grant Thornton Conseil, une entité entièrement dédiée aux métiers du Conseil, est mise en place par Fidaroc
- 2013 : Le dirigeant de Fidaroc Grant Thornton, est nommé par ses pairs à la Présidence du Hub Afrique Francophone (8 pays) au sein de Grant Thornton

Chiffres clés :

- 120 collaborateurs,
- 9 associés au Maroc,
- 2 bureaux : Casablanca et Marrakech

1.2 Le comité des associés de Grant Thornton



FIGURE 1.1 – Le comité des associés de Grant Thornton

1.3 Les différents services offerts par l'organisme

Grant Thornton offre une large gamme de services dans ce qui suit :

- Audit et conseil comptable,
- Le conseil,
- En ce qui est fiscal, juridique et social.

Et ceci pour les secteurs suivant :

- Agroalimentaire
- Hôtellerie et tourisme,
- Immobilier et construction,
- Services financiers.

1.4 Contexte général du stage

C'est dans la subdivision de l'ingénierie financière qu'on a effectué notre stage, cette dernière se caractérise par les missions suivantes :

- Elaboration des modèles d'optimisation de la rentabilité, quantification des risques, valorisation des portefeuilles à travers des analyses par marge, Embedded value, MCEV, PVFP....
- Analyse actif/passif diagnostic et assistance à la mise en place d'un modèle optimal d'allocation stratégique de l'actif en adéquation avec les engagements de l'entreprise.
- Accompagnement solvabilité 2.
- Formation avec SAS et VB.

Ainsi le sujet de notre stage reposera sur l'analyse du risque crédit sous la norme IFRS 9. On étudiera alors la probabilité de défaut en se basant sur les directives de cette norme comptable.

Chapitre 2

L'accord de Norme IFRS 9 et Risque de Crédit

Sommaire

2.1	La Norme IAS 39 : « "Instruments financiers : comptabilisation et évaluation"	19
2.1.1	La norme IAS 39 : Définition	19
2.1.2	La norme IAS 39 : Enjeux, Objectifs et Champs d'application	20
2.2	Norme IFRS 9 (International Financial Reporting Standards 9) . . .	21
2.2.1	Chronologie :	21
2.3	Risque de Crédit :	28
2.4	les composants du risque crédit :	28
2.5	Mesure du risque crédit :	30
2.6	Analyse du risque crédit en pratique :	31

La crise financière mondiale de 2007-2008 est une crise financière marquée par une crise de liquidité et parfois par des crises de solvabilité tant au niveau des banques que des États, et une raréfaction du crédit aux entreprises. Amorcée en juillet 2007, elle trouve son origine dans le dégonflement de bulles de prix (dont la bulle immobilière américaine des années 2000) et les pertes importantes des établissements financiers provoquées par la crise des subprimes.

Comment est née la crise financière ?

C'est du petit marché des subprimes, du nom de ces crédits hypothécaires américains risqués, que tout est parti. Parce que des banquiers ont octroyé des crédits immobiliers à des ménages aux revenus très modestes peu solvables en calculant leur capacité d'emprunt sur la valeur de la maison achetée. Tant que les prix immobiliers progressaient ce système fonctionnait. Mais lorsque l'immobilier a commencé à se replier aux Etats-Unis en 2007, l'effet pervers de cette mécanique s'est enclenché. Les ménages n'ont plus été capables de rembourser leurs emprunts, et les établissements de crédit qui les avait accordés se sont effondrés. La crise s'est ensuite diffusée à tout le système financier par le canal de la titrisation, cette technique née dans les années 1970 consistant à transformer des prêts bancaires en obligations achetées par les investisseurs du monde entier.

Remarque

Un crédit hypothécaire est un prêt qu'une banque accorde à un particulier en se réservant le droit de saisir le bien financé si l'emprunteur n'assure pas le paiement de la dette. Dans cette hypothèse, la revente de cet actif permettra au créancier de se rembourser.

International Accounting Standards Board (IASB) :

Le Bureau international des normes comptables. Organisme international chargé de l'élaboration des normes comptables internationales IAS/IFRS.

International Accounting Standards (IAS) :

Les normes IAS (International Accounting Standards) sont des normes comptables internationales qui ont été publiées avant 2001. Leur domaine est restreint aux documents de synthèse (bilan,

résultats et tableaux annexes). Les normes IAS ont pour but d'assurer une certaine comparabilité des états financiers d'entreprises et s'appliquent à toutes les sociétés cotées en Bourse qui doivent respecter les normes imposées.

International Financial Reporting Standards (IFRS) :

Les IFRS (International Financial Reporting Standards) sont des normes internationales qui ont été mises en place dans un souci de normalisation des règles comptables au niveau international. L'objectif est de rendre lisible pour l'ensemble des acteurs comptables et financiers les documents, bilan, compte de résultats et annexes, que font paraître l'ensemble des grandes entreprises.

International Financial Reporting Standards 9 (IFRS 9) :

Norme comptable qui porte sur les instruments financiers et qu'est venue remplacer en 2018 la norme IAS 39 jugée trop complexe et inadaptée.

Risque de crédit :

Risque de perte financière encouru par un créancier en cas de non respect du débiteur de ses engagements de paiement d'intérêts ou de remboursement de la dette contractée.

La juste valeur :

la juste valeur (ou « fair value ») est une norme comptable consistant à valoriser, à la date de clôture du bilan, les actifs et les passifs des grandes entreprises selon leur valeur de marché ou à défaut selon leur valeur calculée par la somme actualisée des flux de revenus attendus de leur utilisation.

Coût amorti :

montant auquel est évalué l'actif ou le passif financier lors de sa comptabilisation initiale, diminué des remboursements en principal, majoré ou diminué de l'amortissement cumulé calculé par la méthode du taux d'intérêt effectif, de toute différence entre ce montant initial et le montant à l'échéance, et diminué de toute réduction (opérée directement ou par le biais d'un compte de correction de valeur) pour dépréciation ou irrécouvrabilité.

Taux d'intérêt effectif :

le taux qui actualise exactement les décaissements ou encaissements de trésorerie futurs sur la durée de vie prévue de l'instrument financier ou selon les cas sur une période plus courte de manière à obtenir la valeur comptable nette de l'actif ou du passif financier.

2.1 La Norme IAS 39 : « "Instruments financiers : comptabilisation et évaluation" »

2.1.1 La norme IAS 39 : Définition

La norme IAS 39 couvre un large panel d'instruments, y compris tous les instruments dérivés, les placements dans les instruments de dette et de capitaux propres, les actifs et passifs détenus à des fins de transaction et l'endettement propre de l'entreprise. Tous les actifs et passifs financiers (y compris les dérivés) doivent être inscrits au bilan.

Utilisant de plus en plus d'instruments financiers complexes pour gérer les risques financiers ainsi que les mouvements de certaines variables (taux d'intérêt ou de change, cours des actions, cours des matières et produits de base par exemple) qui font peser sur leur activités et devant faire face à une demande d'information bien plus complexe de la part des investisseurs, les sociétés cotées ont besoin d'un référentiel comptable offrant un degré élevé de transparence et de comparabilité des résultats, ainsi que d'un marché unifié et n'obéissant qu'à une seule législation .

C'est donc dans l'optique de répondre aux besoins précités que sont nées les normes IAS. Au nombre de quarante et une, ces normes s'appliquent aux comptabilités des pays européens et autres. Une fois mises en œuvre, ces dispositions ont permis de répondre à trois grands principes :

L'Exhaustivité : Les états financiers retranscrivent l'activité de l'entreprise et favorisent la disparition des informations hors-bilan.

La comparabilité : La comparabilité : Les états financiers sont normés et identiques à toutes les entreprises.

La neutralité : Les normes ne laissent pas de marge de manœuvre aux entreprises dans le pilotage des comptes.

Parmi ces normes, se trouve l'IFRS 9 qui aujourd'hui vient remplacer la norme IAS 39 traitant de la comptabilisation et de l'évaluation des instruments financiers (Actifs, passifs, dérivés), des opérations de couverture et des règles de sortie de bilan des actifs et passifs financiers.

2.1.2 La norme IAS 39 : Enjeux, Objectifs et Champs d'application

En 2002, l'Union européenne a adopté une réglementation qui exige des sociétés européennes cotées qu'elles appliquent les normes internationales d'information financière (les IFRS, et par conséquent **La norme IAS 39**) à compter de 2005 pour l'établissement de leurs états financiers consolidés. Il s'agit de la plus importante évolution en matière de communication financière depuis 25 ans, ce qui rend la transition aux normes IFRS majeure pour près de 7 000 sociétés cotées qui établissaient auparavant leurs comptes selon les principes comptables généralement admis dans leur pays.

La norme IAS 39 en particulier, qui porte sur les instruments financiers (substantiellement révisée en 2003), est une révolution. Pour les entreprises qui investissent, empruntent ou utilisent les instruments dérivés de façon significative, la norme soulève des questions plus vastes que toute autre norme antérieure dans tous les domaines de la comptabilité.

L'objectif de **La norme IAS 39** est d'établir les principes de comptabilisation et d'évaluation des actifs financiers, des passifs financiers et de certains contrats d'achat ou de vente d'éléments non financiers.

La norme s'appliquait, depuis la date de son entrée en vigueur, à toutes les entreprises qui investissent, empruntent ou utilisent les instruments dérivés de façon significative, quelles que soient leur forme, qu'elles soient cotées ou non et quelque soit le secteur d'activité auquel elles appartiennent (industrie et commerce, banque, assurance...).

Elle couvrait également tous les instruments financiers, à l'exception notamment des titres de participation, des contrats de location et d'assurance, des plans d'épargne salariaux et des instruments de capitaux propres (actions) émis par l'entreprise. La norme s'appliquait également aux contrats

sur marchandises sauf si la marchandise fait partie des actifs concernés par l'activité de l'entreprise.

2.2 Norme IFRS 9 (International Financial Reporting Standards 9)

2.2.1 Chronologie :

Publié par IASB, cette norme a été finalisée en juillet 2014 afin de remplacer la norme IAS-39.

Passant par une longue procédure dont les dates clés sont :

- 12/11/2009 : l'I.A.S.B. publie la 1ère version de la norme IFRS-9, instruments financiers : qui porte exclusivement sur la classification et l'évolution des actifs financiers.
- 28/10/2010 : une 2ème version apparaît, venant compléter la première version en y incluant les passifs financiers.
- 19/11/2013 : publication d'un nouveau chapitre de la norme IFRS-9, portant sur la comptabilité de couverture ainsi que des amendements à IFRS-9, IFRS-7 « Instruments financiers : informations à fournir » et à IAS-39 « Instruments financiers : comptabilisation et évaluation ».
- 24/07/2014 : Version finale de l'IFRS-9, regroupant les 3 phases qui ont constitué le projet :
 - Classification et évaluation
 - Dépréciation,
 - Comptabilité de couverture.
- En 2019 : l'I.A.S.B. a publié des amendements à la norme IFRS-9 intitulés « Caractéristiques de remboursement anticipées avec rémunération négative ».
- 26/09/2019 : l'I.A.S.B. a publié des amendements aux normes IAS 39, IFRS7 et IFRS9 dans le cadre de la réforme des taux d'intérêt de référence.

Ces amendements, conçus pour permettre aux entités de fournir des informations financières utiles pendant la période d'incertitude liée à la réforme de l'IBOR, modifient certaines dispositions en matière de comptabilité de couverture.

IFRS 9, la nouvelle norme de comptabilisation des instruments financiers, est d'application obligatoire à compter du 1er janvier 2018.

Elle modifie en profondeur les règles actuelles de classement et d'évaluation des actifs financiers, ainsi que le modèle de dépréciation des actifs financiers. Elle aligne également davantage la comptabilité de couverture et la gestion des risques des entreprises.

Une nouvelle méthode du mode de provisionnement des créances entrera en vigueur le 1er janvier 2019 au Maroc. Le changement touche le mode de provisions qui passera des pertes de crédit encourues à pertes de crédit attendues. Habituellement, la provision n'est constatée qu'après trois échéances impayées, l'IFRS 9 révisée impose le provisionnement dès l'octroi du crédit de façon à prévenir une éventuelle insolvabilité du client. En effet, le renforcement des provisions pour risques induit un changement considérable pour les groupes bancaires. Celui-ci n'affectera pas seulement les fonctions finance et risque, mais aussi l'organisation.

Le niveau supérieur des provisions de pertes induit par l'IFRS 9 va contraindre les groupes bancaires à revoir les exigences en capital. La gamme de produit et le business model devront aussi être réévalués. (Figure ci-dessous)

Niveau 1	Niveau 2	Niveau 3
Risque de crédit <i>Lors de l'entrée au bilan et en l'absence d'une augmentation significative du risque de crédit</i>	Risque de crédit <i>Augmentation significative du risque de crédit</i>	Risque de crédit <i>Présence d'une indication objective suggérant la dépréciation</i>
Comptabilisation de la provision pour risque <i>Dépréciation à hauteur de la perte de crédit attendue sur 12 mois</i>	Comptabilisation de la provision pour risque <i>Dépréciation à hauteur de la perte de crédit attendue sur la durée de vie</i>	Comptabilisation de la provision pour risque <i>Dépréciation à hauteur de la perte de crédit attendue sur la durée de vie</i>
Intérêt perçu <i>Sur la base de la valeur comptable brute</i>	Intérêt perçu <i>Sur la base de la valeur comptable brute</i>	Intérêt perçu <i>Sur la base de la valeur comptable nette</i>

FIGURE 2.1 – Les 3 niveaux de comptabilisation de la provision risque crédit sous les normes IFRS.9

Si après la comparaison avec la comptabilisation initiale, il s'avère que le risque de crédit a sensiblement augmenté, l'instrument financier est classé en niveau 2. Là, l'entreprise doit comptabiliser une dépréciation à hauteur des pertes attendues sur la durée de vie de l'actif financier (lifetime expected credit losses) dans le compte de résultat.

S'il existe, le jour du bilan, des indications objectives suggérant une dépréciation, l'instrument financier est classé en niveau 3. Le calcul de la dépréciation se fait alors de manière analogue au modèle de dépréciation actuel, sur la base de la valeur comptable nette. Pour appliquer le modèle des pertes attendues, les entreprises ont besoin d'autres informations que pour l'actuel modèle de dépréciation. L'introduction de ce modèle demande beaucoup de travail et d'importantes adaptations du système.

IFRS 9 propose dans certaines conditions un modèle simplifié, par exemple pour les créances résultant de livraisons et de prestations. Ce modèle permet de comptabiliser une dépréciation à hauteur des pertes attendues sur la durée de vie, et ce, indépendamment de la qualité du crédit.

Cette nouvelle norme définit de nouvelles règles en matière de classement et d'évaluation des instruments financiers, de dépréciation du risque de crédit des actifs financiers et de comptabilité de couverture, hors opérations de macro-couverture. En effet la norme :

- rassemble dans une norme unique les différentes phases du projet de remplacement de la norme IAS 39
- Apporte une approche logique et unique pour la classification et l'évaluation des actifs financiers qui reflète le modèle économique dans le cadre duquel ils sont gérés ainsi que leurs flux de trésorerie contractuels.
- Garantit l'intégration d'un nouveau modèle de dépréciation des actifs financiers,
- Vient avec une approche sensiblement simplifiée de la comptabilité de couverture.

Cette norme est constituée de 3 phases qui ont permis de structurer les projets au sein des établissements financiers :

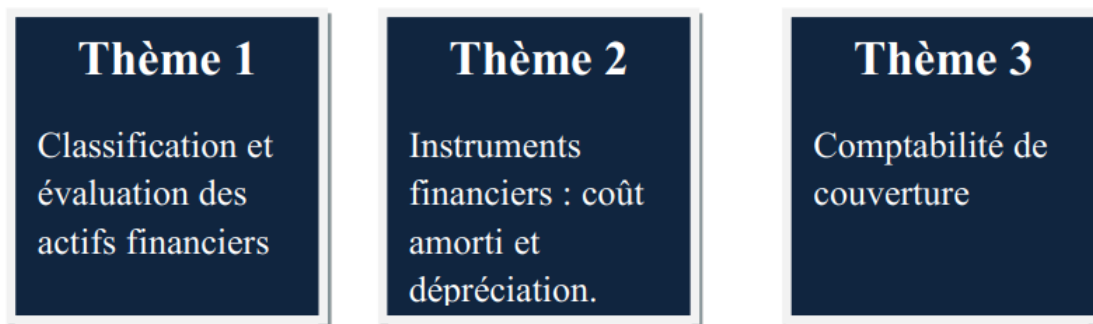


FIGURE 2.2 – Les trois composantes de la norme IFRS 9

Classification et évaluation :

L'IASB a publié les dispositions de la norme IFRS 9 relatives à la classification et l'évaluation des actifs financiers et des passifs financiers, respectivement en novembre 2009 (actifs financiers) et en octobre 2010 (passifs financiers).

Actifs financiers : IAS 39 comprend différentes catégories de classification et différents modèles de dépréciation qui y sont associés. Un bon nombre de problèmes d'application soulevés par IAS 39 est lié à la classification et l'évaluation des actifs financiers. La classification détermine la manière dont les actifs financiers sont comptabilisés dans les états financiers, et en particulier, la manière dont ils sont évalués de façon continue. Les dispositions relatives à la classification et l'évaluation constituent le fondement de la comptabilité des instruments financiers. Les dispositions relatives à la dépréciation et à la comptabilité de couverture sont basées sur cette classification.

IFRS 9 introduit une approche logique et unique de classification pour tous les actifs financiers, soit au coût amorti, soit à la juste valeur, y compris pour les actifs financiers qui comportent un dérivé. Dans ce cas de figure, l'actif financier est classé dans son intégralité plutôt que d'être soumis à des règles complexes de décomposition. L'approche est fondée sur des principes plutôt que sur des règles comme dans IAS 39, jugées complexes et difficiles à appliquer. Deux critères doivent être utilisés pour déterminer comment les actifs financiers doivent être classifiés et mesurés :

- Le business model de l'entité pour la gestion des actifs financiers, et
- Les caractéristiques des flux de trésorerie contractuels de l'actif financier.

Pour un actif financier dans le champ d'IFRS 9, il y a 3 types de modèles économiques :

- L'objectif du modèle économique est uniquement de détenir des actifs financiers pour encaisser des flux de trésorerie contractuels : l'actif financier est évalué au coût amorti.
- L'objectif du modèle économique est à la fois de détenir des actifs financiers pour encaisser des flux de trésorerie contractuels et de vendre des actifs financiers : l'actif financier est évalué à la juste valeur par le biais des autres éléments du résultat global. Ce modèle économique a été ajouté par l'IASB en juillet 2014 dans la version finale de la norme IFRS9.
- Les actifs financiers qui ne sont détenus dans le cadre d'aucun des deux modèles économiques ci-dessus sont évalués à la juste valeur par le résultat.

IFRS 9 requiert qu'un actif financier soit reclassé d'une catégorie à une autre, si et seulement si le modèle économique de l'entité pour gérer les actifs financiers est modifié, ce qui en conséquence devrait se produire peu souvent. Dans ce cas de figure, des informations sur le reclassement devront être fournies en application de la norme « IFRS 7 Instruments financiers : informations à fournir ».

Passifs financiers : Lors de l'élaboration de la norme IFRS 9, la plupart des répondants ont estimé que IAS 39 fonctionnait plutôt correctement. En conséquence, les dispositions contenues dans IAS 39 demeurent pour l'essentiel inchangées dans IFRS 9 : la plupart des passifs financiers continueront donc à être évalués au coût amorti.

IFRS 9 inclut la même option que dans IAS 39, permettant aux entités d'évaluer leurs passifs financiers à la juste valeur par le résultat si des critères spécifiques sont remplis.

Le seul vrai point auquel il a été demandé à l'IASB de remédier urgemment concerne la volatilité du compte de résultat liée aux variations de juste valeur résultant du risque de crédit propre lorsqu'une entité a opté pour la juste valeur dont le résultat est contre-intuitif. Lorsqu'une entreprise a émis des instruments et que son propre crédit se détériore, la diminution de la juste valeur de son passif se traduit par un profit. Au contraire, le rétablissement de son crédit se traduit par une perte. Selon IFRS 9, les variations de juste valeur seront désormais comptabilisées dans les

autres éléments du résultat global tandis que le bilan continuera à enregistrer la juste valeur.

IFRS 9 permet aux entités d'appliquer par anticipation la disposition relative à la comptabilisation des variations de juste valeur liées au risque de crédit en autres éléments du résultat global pour les passifs financiers comptabilisés à la juste valeur par le résultat, sans adopter IFRS 9 dans son intégralité.

La dépréciation :

En comptabilité, la dépréciation est le fait de comptabiliser la potentielle future moins-valeur d'un actif (soit la perte de valeur de l'actif). Dans notre contexte, la dépréciation d'un actif est représentée par sa provision, et cette provision est déterminée par une estimation des pertes liées à un crédit (on associe à une ligne de crédit, un montant estimé qui devrait permettre de couvrir le risque engendré par la contraction de ce crédit). C'est dans cette partie que le modèle de dépréciation IAS 39 des pertes encourues est remplacé par le modèle IFRS 9 des pertes attendues. Ce modèle identifie 3 groupes dans lesquels sont classés les actifs en fonction de leur qualité de crédit, c'est-à-dire le niveau de risque qu'ils représentent. Ces groupes sont appelés « Bucket » ou « Stage », et l'estimation de la perte liée à un crédit dépendra de son classement dans un groupe ou dans l'autre :

- Stage 1 : l'actif vient d'être émis ou acheté et est considéré comme peu risqué, ou sa qualité de crédit ne s'est pas significativement détérioré. La dépréciation représentera alors la perte de crédit attendue sur les 12 prochains mois (l'ECL à un an). Si la qualité de crédit ne change pas, ce calcul à un an est appliqué chaque année jusqu'à l'échéance du prêt (jusqu'au dernier remboursement).
- Stage 2 : dès l'origine l'actif est considéré comme risqué, ou sa qualité de crédit s'est significativement détériorée, sans qu'aucune perte de crédit n'ait cependant encore été observée. Par exemple un crédit avec un impayé (l'impayé représente ici l'événement déclencheur du passage en Stage 2. En revanche il n'y a pas de perte observée, le crédit n'est pas en défaut). La dépréciation représentera alors la perte de crédit attendue sur la durée de vie totale du contrat (jusqu'à sa date d'échéance).

- Stage 3 : la qualité du crédit s'est significativement détériorée, et une perte est observée. Par exemple un contrat en défaut, ou au contentieux. Comme pour le stage 2, la dépréciation est estimée « à maturité », c'est-à-dire jusqu'à la date de clôture du contrat.

En Stage 2 et 3, l'estimation de la perte attendue est donc projetée jusqu'à la dernière échéance du prêt, contrairement au stage 1 où elle est projetée « seulement » à un an. En moyenne, les provisions comptabilisées en Stage 2 et 3 seront donc plus importantes qu'en Stage 1 (excepté pour les contrats arrivants ayant une échéance à moins d'un an, et pour lesquels le calcul en Stage 1, 2 ou 3 donnera donc le même ECL. Par exemple un contrat ayant une date de fin dans 6 mois aura une ECL identique en Stage 1, 2 ou 3). Néanmoins, la très grande majorité des portefeuilles des banques est classée en Stage 1.

Le modèle IAS 39 ne démarrait la comptabilisation d'une perte de crédit qu'à l'apparition d'un événement déclencheur comme par exemple un retard de paiement de 30 jours. Par rapport à cette ancienne norme, la principale évolution d'IFRS 9 est donc le provisionnement des encours sains dits « non sensibles » (Stage 1), qui engendre une augmentation considérable des pertes comptabilisées.

Du côté du FASB, l'approche est plus conservatrice, et plus simple à comprendre. Pas de stages, mais une méthode de comptabilisation similaire pour chaque deal : calcul de l'ECL sur l'ensemble de la durée du contrat dès son émission. Les provisions calculées via cette norme seront donc par construction plus importantes.

La comptabilité de couverture :

IFRS 9 introduit un modèle sensiblement réformé pour la comptabilité de couverture, avec des informations accrues sur l'activité de gestion des risques. Le nouveau modèle représente une révision importante de la comptabilité de couverture qui aligne le traitement comptable sur les activités de gestion des risques, permettant aux entités de mieux rendre compte de ces activités dans leurs états financiers. De plus, grâce à ces modifications, les utilisateurs des états financiers bénéficieront d'une meilleure information sur la gestion des risques et sur l'effet de la comptabilité

de couverture sur les états financiers.

2.3 Risque de Crédit :

La crise de 2008 a mis en évidence l'échec des banques dans la gestion du risque crédit, tel qu'on parle de reconnaissance tardive des pertes de crédit de la part des banques.

Ainsi le comité de Bâle définit le risque comme un aléa et une perte potentielle. Si cet aléa porte sur des scénarios positifs, il n'est pas considéré comme risque, de même si la perte est certaine, ce n'est pas non plus une risque.

Donc pour résumer le risque crédit est l'aléa qui pèse sur le respect d'un engagement de régler une dette. Si on généralise, c'est l'incertitude affectant les montants et les dates auxquels les paiements du débiteur seront effectués. On distingue alors 3 risque :

- Un risque de dégradation (downgrading risk)
- Un risque de défaut (default risk)
- Un risque de faillite (failure risk)

Et le risque crédit recouvre deux éléments distincts : le risque de perte potentielle (downgrading) et le risque de perte extrême (risque de défaut).

L'examen du risque de crédit, permet aux créanciers de décider :

- De l'acceptation de son principe (accord de crédit),
- Son montant, sa maturité et son mode de remboursement,
- Les conditions de taux et security package (garanties et covenants).

2.4 les composants du risque crédit :

Généralement, on admet que le risque de crédit se compose :

- Du défaut : vu comme accident de crédit, tel que c'est le fait que l'emprunteur n'honore pas ses engagements à l'échéance.
- De l'exposition à la date de défaut : c'est le capital restant dû.

- De la perte en cas de défaut : c'est la fraction de l'exposition qui ne sera pas remboursée (cette notion dépend fortement du taux de recouvrement en cas de défaut).
- De l'horizon du défaut : le moment futur où le défaut peut se produire.

Pour un crédit de maturité T , et de montant C , on a un taux d'intérêt i :

En cas de non défaut, la valeur de ce crédit sera : $V_T = C * (1 + i)^T$

Si on a défaut, on ne récupérera qu'une fraction R du crédit :

$$V_T = C * (1 + i)^T * R$$

On pose p : la probabilité de défaut, on sait que R et p sont aléatoire :

$$V_T = C * (1 + i)^T * R * p + V_T + C * (1 + i)^T * (1 - p)$$

$$V_T = C * (1 + i)^T * (1 - p * (1 - R))$$

De cette formule, on conclut que le risque crédit dépend de :

a- Probabilité de défaut “Probability of default , PD” :

Probabilité de survenance d'un cas de défaillance pendant une durée déterminée.

Exemple : une probabilité de défaillance de 15% signifie que la probabilité de survenance d'un défaut de paiement est de 15%. IFRS 9 fait la distinction entre la probabilité de défaillance pour les 12 mois à venir (12-month PD) et la probabilité de défaillance pour la durée de vie résiduelle totale (lifetime PD).

b- Perte en cas de défaillance “loss given default, LGD ” :

Pourcentage de perte en cas de survenance d'un cas de défaillance.

Exemple : un taux de perte de 60% signifie qu'en cas de survenance d'un cas de défaillance, 60% de la créance sera perdue au moment de la survenance de cet événement. Les 40% restants de la créance peuvent être recouvrés.

c- Exposition en cas de défaillance “Exposure at default, EAD ” :

montant de la créance à recouvrer au moment de la survenance du cas de défaillance.

Forward Looking :

C'est l'approche prospective pour l'estimation des PD. Celle-ci sert à résumer dans les ECL l'information contenue dans les événements passés, la conjoncture actuelle et les prévisions de la conjoncture économique future.

Cette approche représente le plus de la norme d'IFRS 9, et pour répondre à ses exigences : le calcul de l'ECL doit se faire par l'une des deux méthodes :

- La moyenne pondérée des différentes pertes crédit déterminées pour chacun des scénarios choisis, la pondération de fait par la probabilité d'occurrence de ces scénarios.
- L'ajustement des probabilités de défaut calculées sous le scénario central de manière à tenir compte des différentes non-linéarités qui peuvent exister entre ces PD et les scénarios macro-économiques adverses.

2.5 Mesure du risque crédit :

L'analyse de ce dernier se fait pour plusieurs objectifs :

- Evaluer le niveau de risque présenté par un débiteur pour son créancier à l'instant « t » et anticiper son évolution probable (Forward).
- Exprimer le risque de manière claire et intelligible sur une échelle par un système de notation au sens large, qualitative et quantitative, permettant à la fois de l'exprimer clairement et de le situer dans l'absolu et de la relativiser dans le temps et dans l'espace.
- Examiner sa causalité et ses composantes afin de pouvoir expliquer le niveau de risque.
- Chercher à mesurer une probabilité mathématique de défaut ou de défaillance.

Le modèle de mesure de risque élaboré doit être testé (Backtesting) afin de démontrer sa capacité prédictive.

Le risque doit faire l'objet d'un monitoring afin de rester informer des variations de niveau de

risque

- **Un scoring** : résultat d'un traitement statistique, il correspond directement à une probabilité mathématique de survenance d'un événement. Il est un véritable système expert, souvent utilisé dans l'environnement des entreprises de taille intermédiaire (ETI) et les petites et moyennes entreprises (PME). Cette analyse n'est pas réalisée par les entreprises elles-mêmes mais par des institutions extérieures à ces organisations. Les banques se servent aussi beaucoup de cette méthode en interne grâce à leurs bases de statistiques.
- **Un rating** : il s'agit d'une note (quantitative ou qualitative) sur une échelle fermée s'interprétant directement en termes de risque (une classe de risque).
- **Un ranking** : système de classement qui hiérarchise les risques sur une échelle (classes) souvent qualitative.

2.6 Analyse du risque crédit en pratique :

En pratique cette analyse permet de mettre en place un système de notation de risque, certaines approches adoptées aboutissent directement à fournir une probabilité de défaut, alors que pour d'autres on se retrouve dans la nécessité de traduire la note en probabilité de défaut. On a pour cela 2 systèmes envisageables :

- La méthode at the point in time : PIT,
- La méthode trough the cycle : TTC.

Cependant, il faut savoir que tout système de mesure du risque de crédit peut entraîner des erreurs de classement de l'entreprise étudiée, on note les 2 types d'erreurs :

- Erreur de type 1 : non-détection du risque de faillite quand on considère une entreprise comme non risquée alors qu'elle l'est : elle conduit à un risque de perte sur le crédit,
- Erreur de type 2 : classer une entreprise en risque de faillite alors qu'elle ne l'est pas : elle expose à une perte d'opportunité.

Finalement grâce la crise 2008, plusieurs changements ont été mis en place, on note :

- BIS11 (The Barratt Impulsiveness Scale) a introduit de nouvelles contraintes à l'activité

bancaire.

- Plusieurs ratios ont été renforcés par la mise en place d'un ensemble de règles cohérent.
- Introduction du ratio de levier.
- Introduction du ratio de liquidité.
- L'utilisation du « stress testing » comme outil clé pour évaluer des risques potentiel.

Chapitre 3

Modèle monopériodiques de probabilité de défaut PD

Sommaire

3.1	Analyse et traitement de la base de donnée	34
3.1.1	Les variables propres au client	34
3.1.2	Analyse descriptive	35
3.1.3	Vérification des valeurs manquantes	37
3.1.4	Détection des valeurs aberrantes (BOXPLOT)	38
3.1.5	Test d'indépendance : KHI2	40
3.1.6	Mesure des corrélations entre les variables quantitatives	40
3.2	Modélisation de la probabilité de défaut PD	41
3.2.1	Méthode de Régression logistique	44
3.2.2	Méthode de Machine Learning : Arbre de décision et Forêts aléatoires	54
3.2.3	Choix du meilleur modèle	62

3.1 Analyse et traitement de la base de donnée

3.1.1 Les variables propres au client

Notre bases de données contient des variables propres au client c-à-d personel.elle contient 1 000 observations et inclut comme champs :

La variable	Libellé	Type
checkingstatus1	Statut du compte courant existant	qualitative
Duration	La durée en mois	numérique
History	Histoire de crédit	qualitative
Purpose	l'objectif de crédit	qualitative
Amount	Montant du crédit	numérique
Savings	Compte d'épargne / obligations	qualitative
Employ	la durée passée dans l'emploi actuel	qualitative
Installment	Taux de versement en pourcentage du revenu disponible	numérique
Status	Statut personnel et sexe	qualitative
Other	Autres débiteurs / garants	qualitative
Residence	la durée passée dans residence actuelle	numérique
Property	Propriété(immobilier, voiture....)	qualitative
Age	Age en années	numérique
Otherplans	Autres plans de versement(Banque, magasin, aucun)	qualitative
Housing	Logement (loyer, propre, gratuit)	qualitative
Cards	Nombre de crédits existants dans cette banque	numérique
Job	Emploi de client	qualitative
Liabe	Nombre de personnes susceptibles d'assurer la maintenance	numérique
Tele	Téléphone (aucun, oui : enregistré sous le nom du client)	qualitative
foreign	travailleur étranger(oui, non)	qualitative

TABLE 3.1 – Les variables propres au client

3.1.2 Analyse descriptive

-Le tableau ci-dessous montre que les bons emprunteurs représentent 70% de l'échantillon contre 30% de mauvais emprunteurs :

Default	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	700	70.00	700	70.00
1	300	30.00	1000	100.00

FIGURE 3.1 – Tableau des effectifs des bons et mauvais emprunteurs

-Le tableau ci-dessous montre que les chomeurs(A171) représentent 0,022% de l'échantillon. parmi eux 0.007% ont eu un défaut de remboursement contre 0.0 15% qui ont toujours bien remboursés leurs prêts.La base de donnée contient 630 employés (A173) ; 0.186% d'entre eux ont fait défaut contre 0.444% non défaut.

Les indépendants(A174) constituent 0.148% dont 0.051% ont fait défaut et 0.097% qui ont toujours bien remboursés leurs prêts . Pour les non qualifiés(A172) qui font le défaut, constituent 0.056%.

Le statut d'employés etre lié'a un haut risque de crédit alors celui du Chomeurs peut etre associé a un faible risque de crédit.

Table of Default by job					
Default	job				
	A171	A172	A173	A174	Total
0	15	144	444	97	700
1	7	56	186	51	300
Total	22	200	630	148	1000

FIGURE 3.2 – Table de contingence entre Job de client et défaut de remboursement

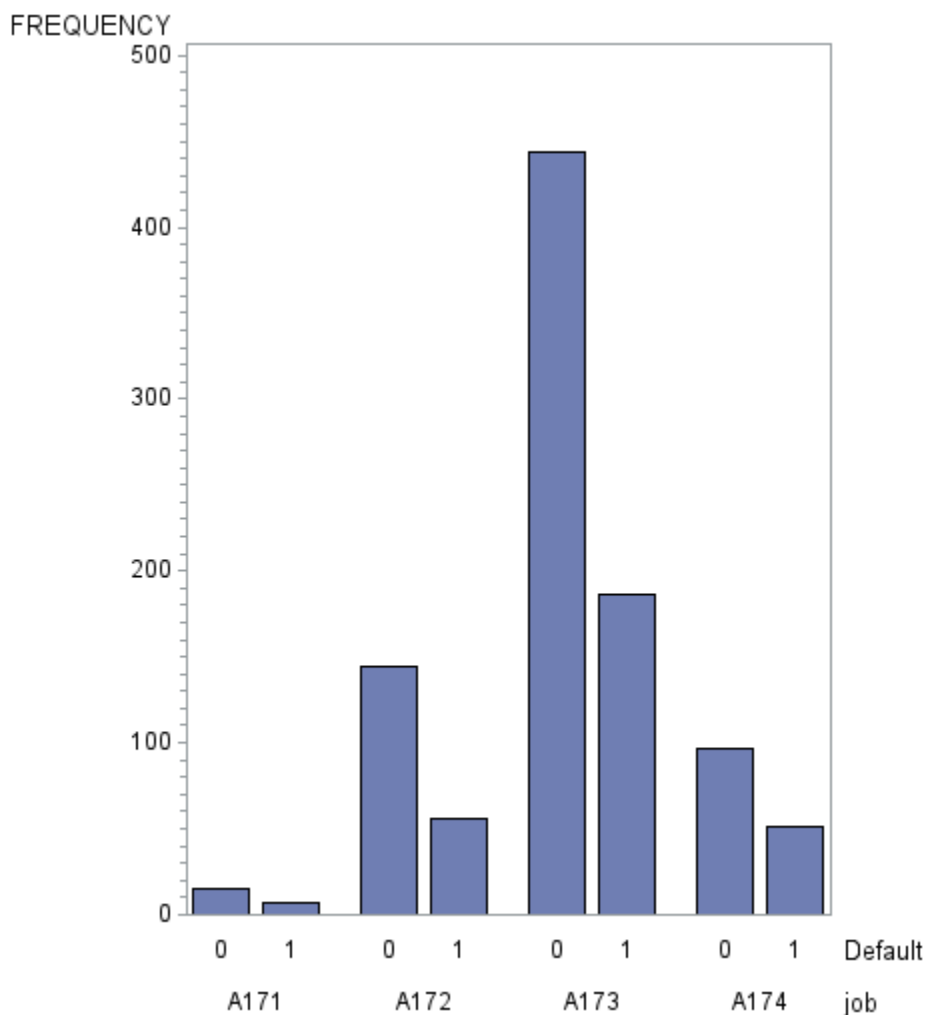


FIGURE 3.3 – La distribution de la variable Job par rapport la variable Default

-Le tableau ci-dessous montre les statistiques élémentaires liées aux variables quantitatives de notre base de données.

Variable	N	Mean	Std Dev	Minimum	Maximum
duration	1000	20.9030000	12.0588145	4.0000000	72.0000000
amount	1000	3271.26	2822.74	250.0000000	18424.00
installment	1000	2.9730000	1.1187147	1.0000000	4.0000000
residence	1000	2.8450000	1.1037179	1.0000000	4.0000000
age	1000	35.5460000	11.3754686	19.0000000	75.0000000
cards	1000	1.4070000	0.5776545	1.0000000	4.0000000
liable	1000	1.1550000	0.3620858	1.0000000	2.0000000

FIGURE 3.4 – Statistiques élémentaires des variables quantitatives

On constate que le moyen du montant de crédit est 3271.26 Dhs. Nous observons aussi que le nombre de credit existant dans cette banque entre 1 et 4 et l'age moyen égale 35.54.

3.1.3 Vérification des valeurs manquantes

En statistique, les données manquantes, ou valeurs manquantes, apparaissent lorsqu'une observation d'une variable n'a pas de valeur. Les données manquantes sont fréquentes et peuvent avoir un effet significatif sur les conclusions qui peuvent être tirées à partir des données

Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
checkingstatus1	INPUT	4	0	A14	39.40	A11	27.40
employ	INPUT	5	0	A73	33.90	A75	25.30
foreign	INPUT	2	0	A201	96.30	A202	3.70
history	INPUT	5	0	A32	53.00	A34	29.30
housing	INPUT	3	0	A152	71.30	A151	17.90
job	INPUT	4	0	A173	63.00	A172	20.00
otherplans	INPUT	3	0	A143	81.40	A141	13.90
others	INPUT	3	0	A101	90.70	A103	5.20
property	INPUT	4	0	A123	33.20	A121	28.20
purpose	INPUT	9	0	A43	28.00	A40	23.40
savings	INPUT	5	0	A61	60.30	A65	18.30
status	INPUT	4	0	A93	54.80	A92	31.00
tele	INPUT	2	0	A191	59.60	A192	40.40
Default	TARGET	2	0	0	70.00	1	30.00

FIGURE 3.5 – Détection des valeurs manquantes(variables qualitatives)

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum
age	INPUT	35.546	11.37547	1000	0	19
amount	INPUT	3271.258	2822.737	1000	0	250
cards	INPUT	1.407	0.577654	1000	0	1
duration	INPUT	20.903	12.05881	1000	0	4
installment	INPUT	2.973	1.118715	1000	0	1
liable	INPUT	1.155	0.362086	1000	0	1
residence	INPUT	2.845	1.103718	1000	0	1

FIGURE 3.6 – Détection des valeurs manquantes (variables quantitatives)

En vérifiant ceci avec **SAS Miner** dans notre base de données, on remarque qu'il n'y a heureusement aucune valeur manquante détectée.

3.1.4 Détection des valeurs aberrantes (BOXPLOT)

En statistique, une donnée aberrante est une valeur ou une observation qui est « distante » des autres observations effectuées sur le même phénomène, c'est-à-dire qu'elle s'écarte fortement des valeurs des autres observations, anormalement faible ou élevée. Une donnée aberrante peut être due à la variabilité inhérente au phénomène observé ou bien elle peut aussi indiquer une erreur expérimentale.

Les Box Plot (Boîtes à Moustaches) permettent de visualiser la distribution d'une seule variable. On peut mettre en évidence les points extrêmes en utilisant une règle simple. Nous calculons le 1er quartile Q_1 et le 3ème quartile Q_3 , nous en déduisons l'intervalle interquartile $IQ = Q_3 - Q_1$. Une valeur est considérée comme aberrante si elle est en deçà de $LIF = Q_1 - 1.5 * IQ$ ou au delà de $UIF = Q_3 + 1.5 * IQ$.

Variable	N	Mean	Std Dev	Minimum	Maximum
age	1000	35.5460000	11.3754686	19.0000000	75.0000000
duration	1000	20.9030000	12.0588145	4.0000000	72.0000000
amount	1000	3271.26	2822.74	250.0000000	18424.00

FIGURE 3.7 – Statistiques élémentaires des variables quantitatives

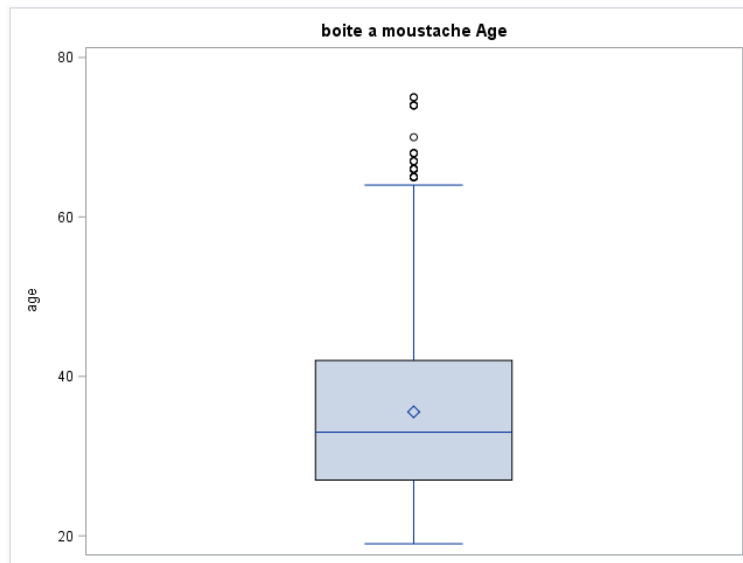


FIGURE 3.8 – Boxplot de la variable "Age"

On remarque à partir du Boxplot "Age" que la valeur 75 est clairement une valeur aberrante . Nous

remarquons aussi qu'il y a des valeurs extrêmes qui varient hors 4.5 et 64.5 soit 23 observations. Ces observations présentent 0.023% de la base totale.

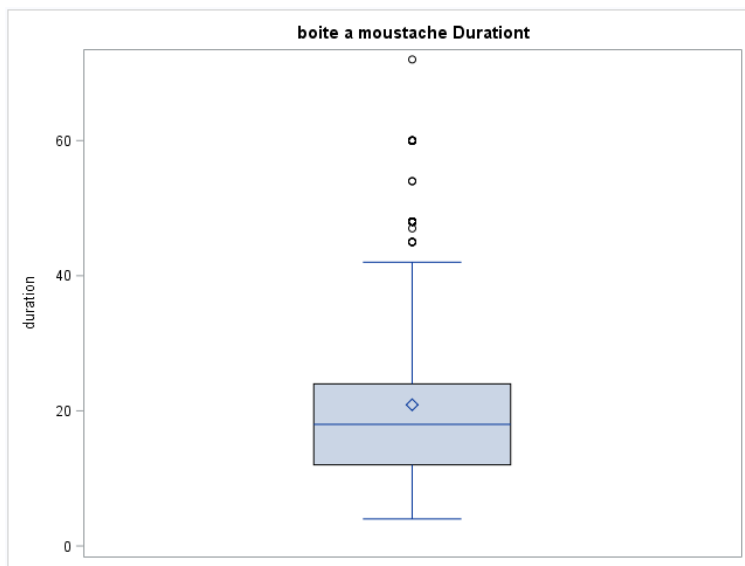


FIGURE 3.9 – Boxplot de la variable "Duration"

On constate à partir du Boxplot "Duration" que les valeurs qui ont supérieur 42 sont clairement des valeur aberrantes . Nous remarquons aussi qu'il y a des valeurs extrêmes qui varient hors -6 et 42 soit 70 observations. Ces observations présentent 0.07% de la base totale. Ce qui est largement négligeable.

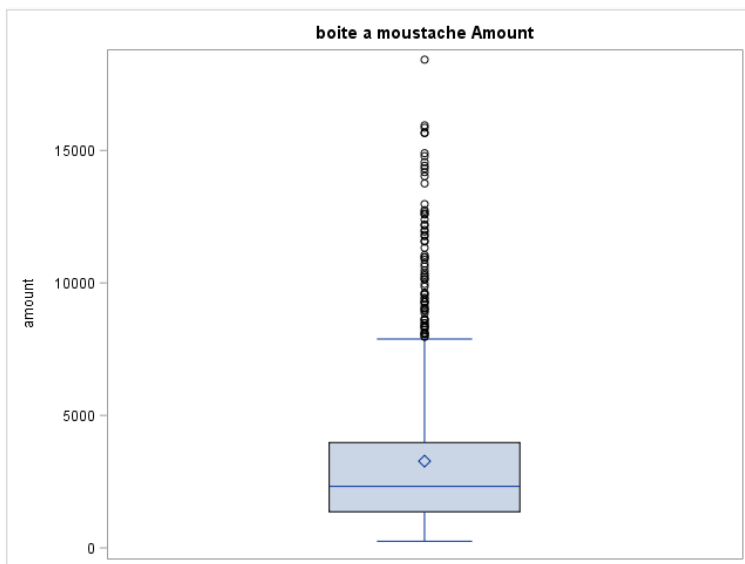


FIGURE 3.10 – Boxplot de la variable "Amount"

On constate à partir du Boxplot "Amount" que les valeurs qui ont supérieur 7883.75 sont clairement des valeur aberrantes . Nous remarquons aussi qu'il y a des valeurs extrêmes qui varient hors -2546.25 et 7883.75 soit 70 observations. Ces observations présentent 0.072% de la base totale. Ce qui est largement négligeable.

3.1.5 Test d'indépendance : KHI2

Le test de Chi-deux est utilisé pour tester l'hypothèse nulle d'absence de relation entre deux variables catégorielles.



FIGURE 3.11 – Test KHI2 : défaut*variables catégorielles

Dans notre cas on va étudier s'il y a dépendance entre notre variable à expliquer Défaut et nos variables catégorielles. Il ressort de notre figure que la variable défaut dépend des variables catégorielles **Tele**, **History** et **checkingstatus1**, par contre relation faible avec les variables **Job** et **Employ**, **foreign**, **housing**, **otherplans**, **property**, **Others**, **Status**, **savings** et **purpose**

3.1.6 Mesure des corrélations entre les variables quantitatives

La corrélation permet de détecter la présence ou l'absence d'une relation linéaire entre deux caractères quantitatifs continus.

Pearson Correlation Coefficients, N = 1000 Prob > r under H0: Rho=0							
	duration	amount	installment	residence	age	cards	liable
duration	1.00000	0.62498 <.0001	0.07475 0.0181	0.03407 0.2818	-0.03614 0.2536	-0.01128 0.7216	-0.02383 0.4515
amount	0.62498 <.0001	1.00000	-0.27132 <.0001	0.02893 0.3608	0.03272 0.3013	0.02079 0.5113	0.01714 0.5882
installment	0.07475 0.0181	-0.27132 <.0001	1.00000	0.04930 0.1192	0.05827 0.0655	0.02167 0.4937	-0.07121 0.0243
residence	0.03407 0.2818	0.02893 0.3608	0.04930 0.1192	1.00000	0.26642 <.0001	0.08963 0.0046	0.04264 0.1778
age	-0.03614 0.2536	0.03272 0.3013	0.05827 0.0655	0.26642 <.0001	1.00000	0.14925 <.0001	0.11820 0.0002
cards	-0.01128 0.7216	0.02079 0.5113	0.02167 0.4937	0.08963 0.0046	0.14925 <.0001	1.00000	0.10967 0.0005
liable	-0.02383 0.4515	0.01714 0.5882	-0.07121 0.0243	0.04264 0.1778	0.11820 0.0002	0.10967 0.0005	1.00000

FIGURE 3.12 – Détection de corrélation entre les variables quantitatives

D'après notre tableau , on remarque que les variables **Amount** et **Duration** sont corrélées, donc On va se limiter d'une seule variable. À part ce cas, on peut assurer que les autres variables ne sont pas significativement corrélées.

3.2 Modélisation de la probabilité de défaut PD

La probabilité de défaut sur an est vue comme la version instantanée de la probabilité de défaut sur la durée de vie Deux raisons principales nous poussent à traiter les PD_{1an} et les $PD_{lifetime}$ séparément :

Les banques ont développé des modèles pour la PD_{1an} , pendant les deux dernières décennies pour les exigences réglementaire de Bâle 2.

Une structure en blocs de construction, divisée en PD_{1an} et $PD_{lifetime}$, facilité le processus d'apprentissage.

Pour mieux comprendre on va d'abord définir la notion de défaut :

D'après Bâle 2; un défaut s'est produit quand l'institution bancaire considère qu'il est peu probable qu'un débiteur rembourse intégralement ses obligations de crédit au groupe bancaire sans recours par l'institution bancaire des actions telles que la réalisation de suretés, ou quand le débiteur n'a pas respecter son échéance contractuel de remboursement et est en retard pour plus

de 90 jours sur toute obligation de crédit au groupe bancaire.

L'IFRS 9, lui, ne définit pas directement le défaut, mais exige plutôt aux institutions de se soumettre aux critères de la gestion interne du risque de crédit. Dans ce sens, ils disent que le défaut ne survient plus tard, que si un actif financier est en souffrance depuis 90 jours.

Dans l'ensemble les autorités de réglementation doivent s'efforcer d'aligner les exigences de fonds propres et les définitions comptables. En outre, les banques doivent améliorer la transparence et l'intelligibilité des rapports financiers.

On note alors des exemples d'éléments déclencheurs du défaut classés en éléments quantitatifs et éléments qualitatifs :

- **Indicateurs quantitatifs** : 90 jours de retard ou 3 mois d'arriérés sont des éléments déclencheurs très communs. Un compteur est mis en place pour tenir compte des retards ou des arriérés.

- **Indicateurs qualitatifs** : Les éléments suivants sont couramment utilisés dans la pratique comme déclencheurs par défaut :

La faillite : Cet événement peut être déclenché en ce qui concerne l'exposition spécifique d'une banque, ou en raison de son expositions vers d'autres institutions ;

L'échéance : le défaut peut être dû à un solde impayé après la date d'échéance ;

Autres indicateurs ; tel l'abstention : qui joue un rôle majeur dans le déclenchement du défaut.

Préparation de la base de données pour la modélisation d'une PD_{1an} :

Il faut savoir que pour le processus du développement du modèle, on a deux étapes clé :

- Estimation de Scorecard,
- Calibration du modèle

On aura besoin de déterminer les spécificités de ce qui suit :

- **Horizon temporel** : Une distinction apparaît entre le développement de la carte de performance et la calibration. Bâle 2 nécessite au moins un ensemble de données de 5 ans pour la modélisation PD. Et d'un point de vue comptable ; une estimation PIT nécessite de calibrer le modèle sur

l'horizon le plus récent.

- Les comptes dupliqués : Comme étape supplémentaire, il faut décider d'inclure ou non le même compte sur différentes périodes. Certains soulignent l'importance de conserver autant d'informations que possible, d'autres mettent en évidence un potentiel biais introduit en considérant le même compte plusieurs fois.

- L'échantillonnage de l'horizon temporel : Les plans d'échantillonnage s'appliquent généralement sur un horizon pluriannuel, on note des exemples où on décide d'échantillonner plus souvent dans les périodes récentes que dans les périodes éloignées.

Sans oublier que tout cela est influencé par la disponibilité des données.

Une importante étape consistera aux choix des variables explicatives. D'après le modèle de Merton, on aura besoin d'indicateurs comptables et financiers. Ainsi afin d'évaluer le risque de crédit qu'encourra une banque, on va exploiter des informations financières et les non-financières.

Altman (1968), un analyste reconnu pour son modèle basé sur l'info comptable, a soulevé 5 principaux ratios de 5 ordres différents et essentiels à la prédiction de défaut :

- Ratio de liquidité = Fonds de roulement / Actifs totaux
- Ratio de profitabilité = Bénéfices non répartis / Actifs totaux
- Ratio d'endettement = Bénéfices avant intérêts et impôts / Actifs totaux
- Ratio de solvabilité = Valeur de l'équité au marché / Passifs totaux
- Ratio d'activité = Ventes nettes / Actifs totaux

Ohlson (1980) était le premier à proposer, pour l'étude du défaut le modèle 'logit', en étudiant ainsi 4 principaux facteurs, on note : la taille de la firme, la structure du capital, la performance financière et les liquidités immédiates. Une étude vue comme un ajustement du modèle d'Altman. Dans notre étude on se basera sur deux méthodes, l'une basée sur les modèles linéaires généralisés et l'autre se fera grâce au Machine Learning.

3.2.1 Méthode de Régression logistique

On va se concentrer sur les GLM, tout d'abord vu leur caractère pratique et leur grande popularité dans le secteur financier, Sans oublier que les GLM fournissent un cadre utile pour d'autres méthodes.

Il faut souligner que les PD PIT estimées ne vont pas directement servir au calcul des ECL. En effet, il faut d'abord l'aligner avec les règles comptables, et ceci à l'aide du Forward Looking. L'étude se fera en 3 étapes :

1) Développement des ScoreCard du modèle linéaire généralisé :

Pour définir notre variable qu'est l'occurrence ou non du défaut, on pose comme convenue une variable binaire :

$$y_i = \begin{cases} 1 & \text{si on est en défaut;} \\ 0 & \text{sinon.} \end{cases}$$

Avec 'i' qui représente le compte sous le portefeuille étudié.

Pour le score du compte i, on utilise la relation logistique souvent vu en pratique :

$$SC_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 * X_{i1} + \dots + \beta_k * X_{ik})]}$$

Avec (X_{i1}, \dots, X_{ik}) vecteur des variables comportementales explicatives.

On a :

$$(\beta_0 + \beta_1 * X_{i1} + \dots + \beta_k * X_{ik}) = \ln\left(\frac{SC_i}{1 - SC_i}\right) = \text{Logit}(SC_i)$$

Où $SC = p(Y = 1/X_1, \dots, X_p)$.

* Estimation des paramètres β_j par la maximisation de la vraisemblance

-Méthode d'estimation :

Pour l'estimation des coefficients d'un modèle de régression logistique, la méthode généralement utilisée est celle du maximum de vraisemblance. On peut décrire les différentes étapes de l'estima-

tion des coefficients de ce modèle comme suit :

- Construction de la fonction de vraisemblance $L(\beta)$ qui présente la probabilité d'observer les résultats comme une fonction de $\beta = (\beta_1, \dots, \beta_p)$. Nous savons très bien que la distribution des probabilités d'un événement qui ne peut avoir que deux occurrences : un succès et un échec est une distribution binomiale. Pour un individu i donné, on modélise sa probabilité l'aide de la loi binomiale $B(1, SC)$ avec :

$$P(Y_i/X_i) = SC_i^{y_i} (1 - SC_i)^{1-y_i}$$

Où $SC = p(Y = 1/X_1, \dots, X_p)$.

La vraisemblance pour un échantillon de n observations aléatoires peut s'écrire alors sous cette forme :

$$L = \prod_{i=1}^n SC_i^{y_i} (1 - SC_i)^{1-y_i}$$

- Transformation par le logarithme de la fonction $L(\beta)$ pour donner la fonction $Ln(L(\beta))$ noté LL :

$$LL = \sum y_i \ln(SC_i) (1 - y_i) \ln(1 - SC_i)$$

- Déviation de la fonction LL par rapport aux différents β_j et annulation des dérivées pour la recherche des maximums.

Cette procédure conduit aux équations de vraisemblance.

$$\frac{\partial LL(\beta)}{\partial \beta_j} = 0 \quad \text{pour } j = 0, \dots, p$$

- Résolution du système d'équations pour retrouver les valeurs qui maximisent la fonction $LL(\beta)$. Le système obtenu est complexe et non linéaire. Il est difficile de le résoudre analytiquement. L'estimation ne peut être obtenue qu'en utilisant un algorithme d'optimisation numérique. Les algorithmes utilisés dans les principaux logiciels d'économétrie sont généralement construits selon la méthode Newton-Raphson.

-Test de significativité des coefficients : test de Wald :

Soit β_j le coefficient associé à la variable exogène X_j

Pour tester :

$$\begin{cases} H_0 : \beta_j = 0; \\ H_1 : \beta_i = 0. \end{cases}$$

On utilise le test statistique de Wald sous H_0 :

$$W = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

avec SE son erreur standard de la variable X_j

la statistique W suit sous H_0 la loi de X^2 à 1 ddl.

Si la valeur observée de $W_{obs} > X^2(1, 1\alpha)$, on rejette H_0 au seuil α .

On dit dans ce cas que le test est significatif, que la variable X_j a un effet sur la probabilité d'apparition de l'évènement observé.

-Test de signification globale :

Afin d'étudier la pertinence du modèle de régression logistique, on préconise la méthode suivante :

la courbe ROC :

La courbe ROC (Receiver Operating Characteristic) est une représentation graphique de la relation qui existe entre la sensibilité (Se) et la spécificité (Sp) d'un modèle de classification d'individus. L'inverse de la spécificité (1-Sp) se place en abscisse tandis que la sensibilité se trouve en ordonnée de ce graphique.

La sensibilité (Se) est la probabilité de classer l'individu dans la catégorie cible étant donné qu'il est observé réellement dans celle-ci, c'est la probabilité de bien détecter un événement.

La spécificité (Sp) est la probabilité de classer l'individu dans l'autre catégorie étant donné qu'il est effectivement observé dans celle-ci, c'est la probabilité de bien détecter un non-événement.

Si la discrimination entre les deux événements est parfaite, la courbe ROC est collée aux extrémités Ouest et Nord du repère. Et dans ce cas, l'aire sous la courbe vaut 1 et la probabilité de distinguer entre les deux événements est de 100%.

ROC	Qualité
=0.5	Aucune discrimination
]0.5, 0.7 [Discrimination faible
[0.7, 0.8 [Discrimination acceptable
[0.8, 0.9 [Discrimination excellente
[0.9, 1]	Discrimination parfaite

TABLE 3.2 – interprétation de l’aire sous la courbe ROC

Si le test ne permet pas la discrimination entre les deux, la courbe ROC se confond avec la première bissectrice. C’est le cas où le modèle n’est pas plus performant qu’un modèle aléatoire. L’aire sous la courbe vaut dans ce cas 0,5 et la probabilité de bien classer un contrat entre les deux événements est de 50%.

Plus la courbe est proche du coin supérieur gauche, meilleur est le modèle. autrement dit, plus l’aire sous la courbe ROC est proche de 1, meilleur est le modèle.

Nous retenons comme règle les interprétations ci-dessus :

Cependant on peut aussi utiliser d’autres fonctions lien tel Probit ou Log-Log comme partie des GLM, or Logit reste la plus utilisée.

a-Donc une première étape consistera à identifier un ensemble effectif de variables explicatives pour modéliser SC_i . Pour cela : la significativité statistique et la solidité économique constitueront les critères clé de ce processus de développement du modèle.

b-Puis on passera à une analyse uni-variée, en procédant à plusieurs régressions uni-variées.

c-Une analyse multi-variée .

d-Et pour clôturer : une régression pas à pas : on examine la liste restante plus en détail par des processus de sélection sur des critères statistiques :

— Soit en Forward Stepwise Regression ;

— Ou en Backward Stepwise Regression

2) Calibration du modèle :

C'est une étape clé, où on passe des SC_i aux PD_i .

Une première étape est la normalisation du score : pour cela on doit définir une nouvelle échelle, c'est-à-dire : définir une fonction de mise à l'échelle puis l'appliquer à toute la base de données et calibrer.

Ici notre objectif consiste à centrer la PD sur son taux le plus récent. Quand on a une base de données complète, on a couramment recours à une régression logistique.

On suppose alors avoir une carte de Scoring basée sur plusieurs années, et notre objectif est d'obtenir une PIT calibration sur une année.

L'IASB-2014 exigera :

$$\varphi_{s,t} = \operatorname{argmax}_{\alpha,\beta} \sum_{i=1}^n \frac{1}{1 + \exp[-(\alpha + \beta * SC_i)]}$$

Avec : $\varphi_{s,t}$: l'indice de solvabilité du crédit du portefeuille 's' en 't'.

Ainsi on importe notre base de données et on lui applique une régression logistique. En cas d'absence de Base de données, on utilisera les informations externes (Rating), si aucune n'est disponible, on passe à la procédure 'Ad-Hoc'.

3) Validation du modèle :

Dans cette étape on analyse la représentativité de notre base de données, ainsi que sa complétude. Puis on passe à la validation de la méthodologie ; en comparant avec d'autre méthode (dans notre cas on comparera avec la méthode du Machine Learning).Finalement, on passe aux tests statistiques :

- Pouvoir discriminatoire : afin de vérifier si le modèle distingue entre les défauts et les non-défauts. Pour cela on utilise les métriques : l'indice de Gini et la courbe de ROC.

- Calibration : on compare les PD modelées et les PD réel.
- La stabilité : Il faut d'abord vérifier que la population sur laquelle le modèle a été fait est similaire à la population en développement. Si nos populations sont différentes, on doit vérifier la stabilité des variables individuellement. En calculant l'indice de stabilité :

$$SI = \sum_{j=1}^J (R_j - O_j) \ln\left(\frac{R_j}{O_j}\right)$$

Avec : R_j est le pourcentage de la population de référence dans la classes j

O_j est le pourcentage de la population observé dans la classe j Ainsi on a si :

- $SI \leq 0.1$: pas de changement significatif n'est survenu
 - $0.1 < SI \leq 0.25$: un petit changement est arrivé
 - $SI > 0.25$: Un important changement a pris place
- Validation de Cross : Ici on divise notre base en g échantillon, puis on développe un modèle sur les g-1 premiers échantillons réunis, et enfin , on essaye de valider le modèle en testant sur le g dernier échantillon.
 - Reproductibilité : Il faut s'assurer que le processus entier est bien défini. Cependant, On devra effectuer une répétition de l'étude sur la même base de données utilisée pour le développement de notre modèle.

Application à notre base de données

La régression logistique sert à modéliser la relation entre des prédicteurs catégoriels ou continus X_i ($i=1, \dots, p$), et une variable réponse catégorielle Y .

Lorsque la variable-réponse ne peut prendre que deux valeurs possibles, un des modèles que l'on peut alors utiliser est le modèle de régression logistique binaire.

En règle générale, le résultat d'une observation binaire est appelé " succès" Ou "échec". Il est représenté mathématiquement par une variable Y telle que $Y = 1$ s'il y a succès et $Y = 0$ s'il y a échec.

Dans notre cas, nous utiliserons une régression logistique binaire car notre variable dépendante <Défaut>, qui prend la valeur 1 si le contrat a connu un défaut et 0 sinon, est binaire.

P.S. : Nous avons utilisé le logiciel **SAS** pour réaliser notre étude .

Tout d’abord, nous avons éliminer les variables qui ont la valeur de l’information (IV) inférieure 0.1 (il reste 8 variables pertinentes), par la suite l’élimination Nous entamons des variables non significatives par la méthode Stepwise qui permet, dans les phases exploratoires de l’élaboration d’un modèle, d’identifier un sous-ensemble utile de prédicteurs. A chaque étape, le procédé ajoute la variable la plus significative ou supprime la variable la moins significative.

Notre base de données comporte 1000 clients.

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	checkingstatus1		2	1	97.5547		<.0001
2	duration		1	2	31.6837		<.0001
3	savings		1	3	19.6947		<.0001
4	purpose		1	4	9.2045		0.0024
5	history		3	5	13.7529		0.0033

FIGURE 3.13 – Présentation des résultats du modèle LOGIT pour les particuliers.

Notre premier tableau de sortie de SAS expose le test Global $\beta = 0$ qui présente un p-value . (Pr>KHI2) inférieur 0.05 qui signifie qu’au moins un des facteurs étudiés impact la probabilité de défaut en l’occurrence *Duration*, *checkingstatus1*, *savings*, *purpose* et *history*.

Le tableau ci-dessous indique que les variables déjà citées entrent en considération dans le modèle (p-value > 0.05).

Tout d’abord, on a fusionner les modalités non significatifs de la variable **Purpose** A41, A42, A43, A44, A45, A46 et A48 en une seule "**Pur**", A40 et A49 en **PPE** , aussi on a regrouper les

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
duration	1	28.6985	<.0001
checkingstatus1	2	53.9909	<.0001
history	3	13.4285	0.0038
savings	1	21.5705	<.0001
purpose	1	9.1060	0.0025

FIGURE 3.14 – Modèle logit : Analyse des effets Type 3.

modalités A62, A63, A64 et A65 de la variable **Saving** . De même pour la variable **checkingstatus1**, on a fusionner les deux modalités A13 et A14. Aussi pour la variable **History**, on a regrouper les modalités A33 et A34 et nous avons la nommer "**His**".

Une fois qu'on a fusionné les modalités, on estime à nouveau notre modèle (Le tableau ci-dessous) :

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.6104	0.3079	137.5021	<.0001
duration		1	0.0386	0.00720	28.6985	<.0001
checkingstatus1	A11	1	1.5088	0.2165	48.5808	<.0001
checkingstatus1	A12	1	1.2359	0.2179	32.1701	<.0001
checkingstatus1	ch	0	0	.	.	.
history	A30	1	1.2443	0.4294	8.3969	0.0038
history	A31	1	0.9315	0.3954	5.5499	0.0185
history	A32	1	0.4506	0.1912	5.5541	0.0184
history	His	0	0	.	.	.
savings	A61	1	0.8904	0.1917	21.5705	<.0001
savings	sav	0	0	.	.	.
purpose	PPE	1	0.5408	0.1792	9.1060	0.0025
purpose	Pur	0	0	.	.	.

FIGURE 3.15 – Modèle logit : Estimation par l'analyse du maximum de vraisemblance.

Nous constatons en conséquence que toutes les modalités sont significatives (autrement dit, celles qui ont une p-value inférieure à 5%).

L'équation de notre modèle est la suivante :

$$P(Y = 0|x_1, \dots, x_1) = \frac{1}{1 + \exp(-3.6104 + 0.0386Duration + \dots + 0.5408(purpose = PPE))}$$

Les estimations des Odds ratio (rapport de cotes) ci - dessous, nous permettent d'avancer qu'un client qui est un Statut du compte courant (A11 : compte < 0 DH) 4.521 fois plus de chances de faire défaut qu'un client qui Statut du compte courant [Ch : (A13 : >= 200 Milles DHs et A14 : pas de compte)]. De même, un client qui pas de crédits pris environ 3.471 fois plus de chance de tomber en défaut comparé à un client qui à [His : A33 : retard de paiement dans le passé et A34 : autres crédits existants (pas dans cette banque)].

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
duration	1.039	1.025	1.054
checkingstatus1 A11 vs ch	4.521	2.958	6.911
checkingstatus1 A12 vs ch	3.441	2.245	5.275
history A30 vs His	3.471	1.496	8.052
history A31 vs His	2.538	1.169	5.510
history A32 vs His	1.569	1.079	2.283
savings A61 vs sav	2.436	1.673	3.547
purpose PPE vs Pur	1.717	1.209	2.440

FIGURE 3.16 – Modèle logit : Estimation des rapports de cotes.

Qualité d'ajustement du modèle : La courbe ROC

La courbe ROC (receiver operating characteristic), est une mesure de la performance d'un classificateur binaire. Graphiquement, on représente souvent la mesure ROC sous la forme d'une courbe qui donne le taux de vrais positifs (TVP, fraction des positifs qui sont effectivement détectés) en

fonction du taux de faux positifs (TVN, fraction des négatifs qui sont incorrectement détectés).

Mathématiquement :

$$TVP = \frac{TP}{TP + FN}$$

$$TVN = \frac{FP}{TP + FN}$$

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	77.6	Somers' D	0.556
Percent Discordant	22.0	Gamma	0.558
Percent Tied	0.4	Tau-a	0.240
Pairs	138096	c	0.778

FIGURE 3.17 – Modèle logit : AUC et Indice GINI.

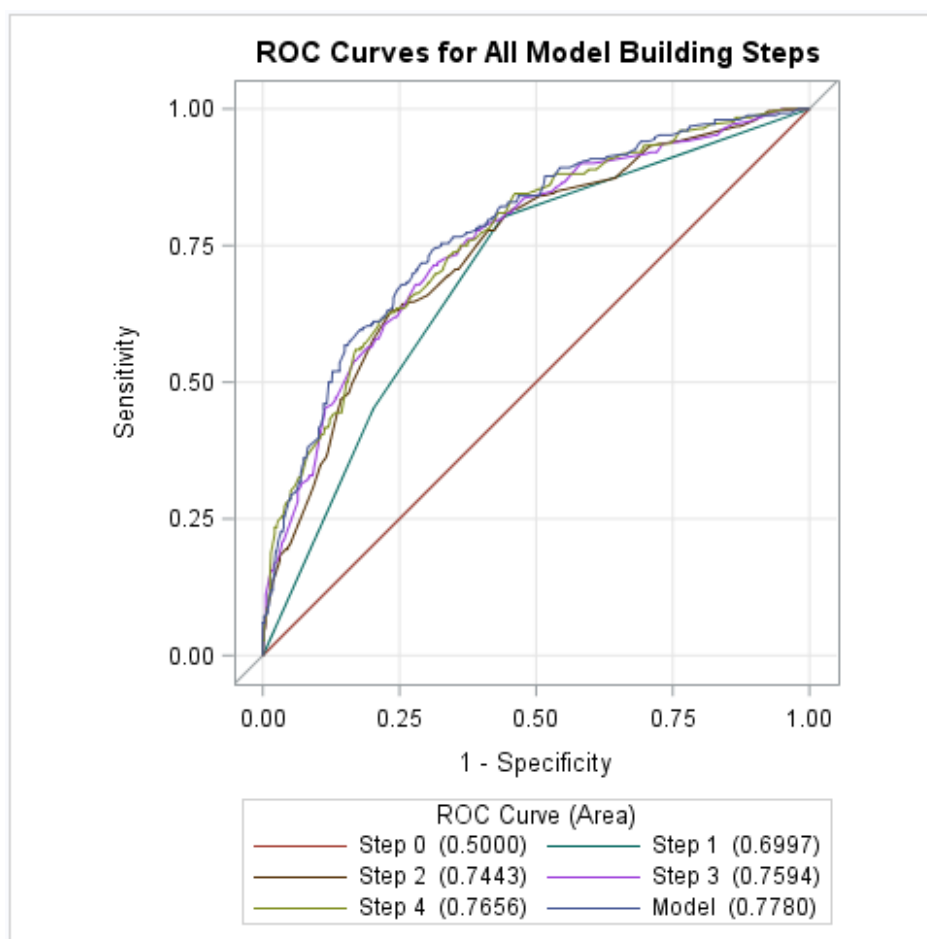


FIGURE 3.18 – Modèle logit : Courbe ROC pour toutes les étapes de création de modèles.

Le tableau ci-dessus montre que notre modèle est de bonne qualité puisque $c = 0.778$ est supérieur à 0.75% et D de Somer's (l'indice de Gini)=0.556 est supérieur 0.5. Ainsi L'image suivante présente les courbes ROC pour les différentes étapes lors de la création du modèle (méthode stepwise), ainsi comme mentionné plus haut le ROC de notre modèle final est de 0.778

Adéquation du modèle :Test de Hosmer-Lemeshow

- La figure ci-dessous montre que $p_value = 0.4613 > 0.05$. Donc notre modèle est compatible avec les données.

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
7.7200	8	0.4613

FIGURE 3.19 – Modèle logit : Adéquation du modèle.

3.2.2 Méthode de Machine Learning : Arbre de décision et Forêts aléatoires

Récemment, et avec l'avancement qu'a connu le monde de la modélisation, de plus en plus de regards se posent sur les procédures du Machine Learning. Ce dernier représente un ensemble d'algorithmes spécifiquement conçu pour les problèmes de reconnaissance de forme basé sur de grandes bases de données. Ceci dit ces techniques s'adaptent parfaitement aux analyses de risques crédits.

Avant d'entamer l'explication de cette méthode et son importance dans la modélisation des PD, il faut garder en compte que les PD qui en résulteront ne vont pas servir directement dans le calcul des ECL. En effet, vu les exigences de l'IFRS 9; un passage par l'approche prospective (Forward Looking) est nécessaire.

1) CART : arbre de régression et de classification :

Les arbres de décisions, souvent utilisés pour les problèmes de classification, servent pour les variables numériques comme pour les catégorielles. Ils visent à diviser la population étudiée en sous-groupes homogènes en se basant sur la variable explicative la plus significative.

En pratique, on se retrouve souvent avec l'un des deux arbres suivants :

- Les arbres de régression : pour le cas de variable quantitative,
- Les arbres de classification : pour ce qui est des variables qualitatives.

Pour mieux comprendre, l'idée générale repose sur la division de l'espace prédictif en sous-ensembles homogènes qui ne se chevauchent pas, afin d'obtenir des modèles à la fois explicatifs et prédictifs.

Commençant par les arbres de régression : On peut reformuler leur procédure en deux étapes simples :

On divise l'ensemble des valeurs possibles X_1, \dots, X_p en J cases exhaustives et non chevauchantes R_1, \dots, R_J .

Pour les observations d'une région R_i : on fait la même prévision qui est la moyenne des valeurs de réponse à R_i .

On cherche les divisions R_1, \dots, R_J qui minimisent la fonction de perte :

$$\sum_{j=1}^J \sum_{i \in R_j} (y_i - \tilde{y}_{R_j})^2$$

Avec :

y_i : la variable réponse pour l'observation i

\tilde{y}_{R_j} : moyenne de la variable réponse dans R_j .

Un fractionnement binaire récursif est utilisé à chaque étape du processus de construction de l'arbre pour obtenir deux nouvelles branches. Compte tenu de la disponibilité d'une série de variables prédictives, on choisit X_j de telles façons à ce qu'on divise en deux sous-régions R_1 et R_2 sans oublier de vérifier qu'on minimise la somme des carrés des résidus comme suit :

$$\boxed{\operatorname{argmin}_{(j,s)} \sum_{i: X_i \in R_1(j,s)} (y_i - \tilde{y}_{R_1})^2 + \sum_{i: X_i \in R_2(j,s)} (y_i - \tilde{y}_{R_2})^2}$$

Pour ce qui est des arbres de classification ; c'est la même procédure sauf qu'on ne peut pas utiliser la somme des carrés des résidus. Une alternative est de la remplacer par le taux d'erreur de classification. En pratique, on utilise souvent l'indice d'impureté de Gini G :

$$\boxed{G = \sum_{k=1}^K pr_{m,k}^{\sim} * (1 - pr_{m,k}^{\sim})}$$

Avec :

$pr_{m,k}^{\sim}$: la proportion des observations estimées dans la région m qui appartiennent à la classe k (exemple de classe : défaut ou non-défaut).

Quand G tend vers 0, ceci implique qu'un nœud contient des observations prédominées par l'une des m classes de la variable cible.

Une alternative à cet indice est l'entropie croisée qui reprend la même interprétation que l'indice de Gini en cas de valeurs tendant vers 0. L'entropie croisée se calcule comme suit :

$$\boxed{D = - \sum_{k=1}^K pr_{m,k}^{\sim} * \log(1 - pr_{m,k}^{\sim})}$$

2) Bagging, Random Forest et Boosting :

Les arbres de décision restent des méthodes pas très performantes au niveau de la prédiction comparés à d'autres tel la régression. Cependant on retrouve le Boosting, Bagging et Random Forest qui viennent améliorer cette performance prédictive.

- **Bagging** : Les arbres de décision se caractérisent par une grande variance, ce qui fait qu'ils sont très sensibles aux fluctuations de l'échantillon. C'est dans ce sens qu'on a recours au Bagging, qui a pour but de réduire la variance de l'estimateur afin de corriger l'instabilité des arbres de

décision. Ce dernier, aussi connu sous le nom de Bootstrap Aggregating, a pour principe de créer de nouveaux échantillons en tirant au hasard et avec remise dans l'ancien échantillon. Ainsi on applique l'algorithme de décision sur ces échantillons, et on prend la moyenne des estimateurs obtenus. La variance est donc réduite grâce à la combinaison de ces estimateurs supposés indépendants. C'est donc une estimation plus précise que celle d'un seul arbre. Or le problème reste que l'interprétation du modèle peut devenir plus difficile. Pour ce qui est des prédicteurs, on peut tester l'importance de chacun en utilisant la somme des carrés résiduels RSS dans le cas des arbres de régression ou l'indice de Gini dans le cas des arbres de classification. Ainsi pour le premier cas, une forte baisse dans la RSS dû à l'exclusion d'une variable renvoie à l'importance de l'indicateur. De même pour le 2ème cas.

- **Random Forest** : Les forêts aléatoires est un algorithme de classification qui combine de nombreux arbres de décisions développés sous l'approche bagging. Donc, par rapport à la méthode précédente, la forêt aléatoire introduit des contraintes afin de réduire la corrélation, améliorant ainsi le processus. En effet, lors du développement de la forêt, à chaque nœud : l'algorithme est contraint de ne pas considérer la majorité des prédicteurs disponibles. Ainsi, pour chaque arbre construit sur l'échantillon aléatoire d'apprentissage, chaque division ne prend en compte qu'un sous-ensemble des prédicteurs. Ce qui fait qu'on n'utilise pas tous les prédicteurs puissants. Ainsi, la principale différence entre le bagging et la forêt aléatoire est la chaîne du sous-ensemble des prédicteurs. Donc si une forêt aléatoire est construite sur tous les prédicteurs ; on aura des résultats identiques à ceux du bagging.

3) Calibration du modèle :

Le processus de calibration pour le modèle développé grâce au Machine Learning reste le même que celui vu pour le modèle GLM.

Le redimensionnement (scaling) a été initialement introduit sur cette base comme méthode de

calibrage des machines à vecteur de support. Le processus consiste à trouver les paramètres d'une fonction sigmoïde, maximisant la vraisemblance du jeu de calibration.

La régression isotonique a été suggérée comme méthode de calibration qui peut être considérée comme une forme générale de répartition en classes qui n'exige aucun nombre ou taille spécifique d'intervalle à prédéterminer.

4) Validation du modèle :

On se retrouve avec la même procédure que pour la validation du modèle GLM.

- Validation de la base de données : exactement le même processus que pour les GLM
- Validation de la méthodologie : Le Machine Learning est souvent l'objet de plusieurs challenges. Il faut alors être sûr de bien comprendre la procédure et le but de chacun des algorithmes que ça soit pour les forêts aléatoires ou le boosting, et les partager avec les parties prenantes. Des comparaisons méthodologiques sont souvent faites avant d'adopter les procédures du Machine Learning.
- Validation par tests statistiques : Le plus important est de faire attention au pouvoir discriminatoire. On devra alors procéder à plusieurs analyses tel : L'analyse de la courbe de ROC, l'analyse de Kolmogorov-Smirnov et l'étude de l'indice de Gini. Sans oublier de valider les PD calibrées, le processus se fait de la même façon que pour celles obtenues par GLM ; on compare les PD réelles et les PD ajustées.

Application à notre base de données

CART désigne une méthodostatistique, introduite par Breiman et al. (1984) qui construit des prédicteurs par arbre aussi bien en régression qu'en classification.

Nous retenons 6 variables explicatives : **checkingstatus1**, **Duration**, **Employ**, **Savings**, **History** et **Amount**

Le premier nœuds segmente la variable **checkingstatus1** en deux segments, un comprend les clients qui ont eux un compte inférieur 200 Milles Dhs d'impayé et une deuxième branche qui comprend

les clients qui ont eu un compte supérieur 200 Mille Dhs et les clients qui n'ont pas de compte. ces derniers sont à leur tour divisés à deux segments, un segment ont eux une Durée supérieure 31.5 mois et l'autre ont eux une Durée inférieure 31.5.

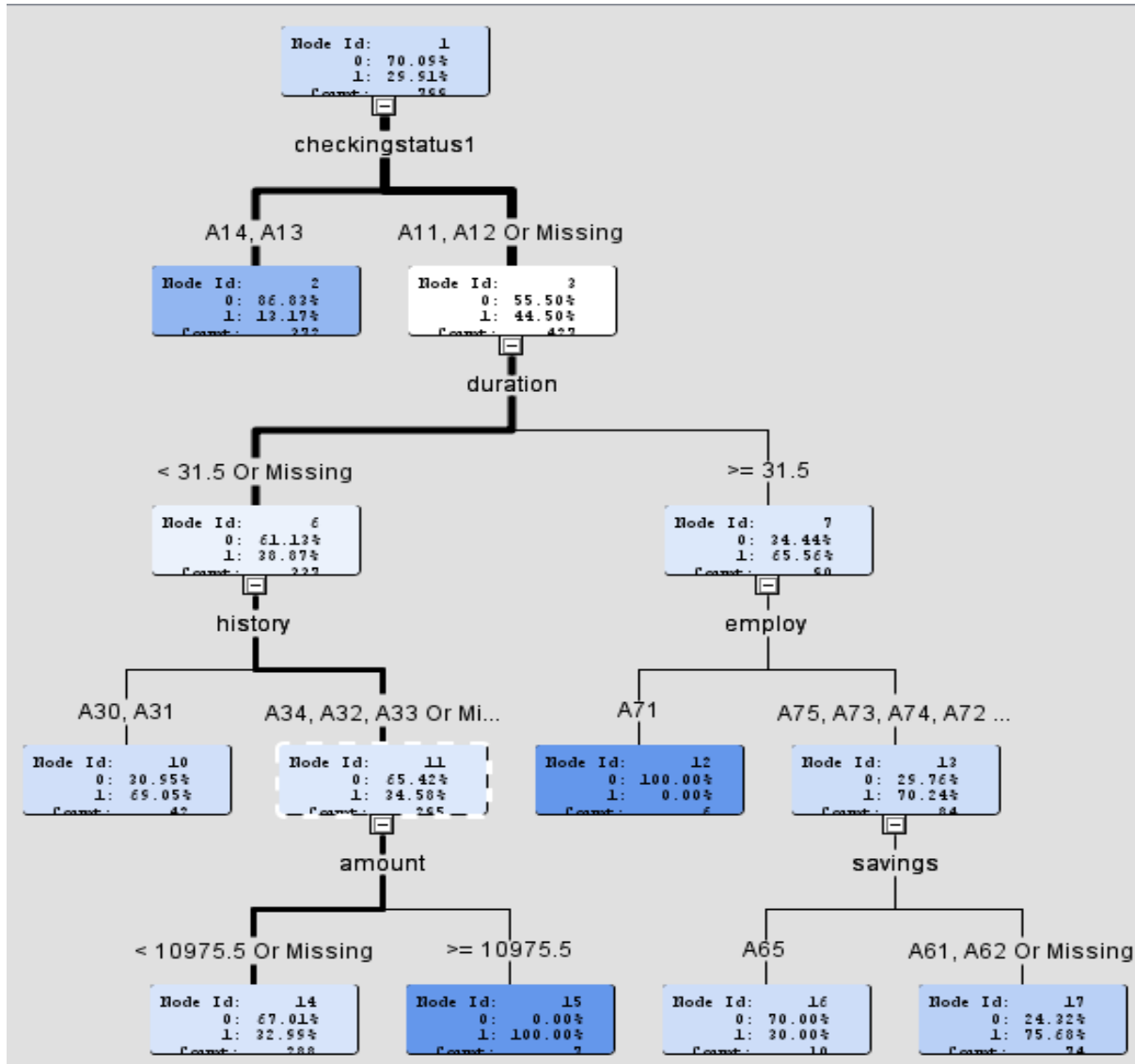


FIGURE 3.20 – Arbres de décision

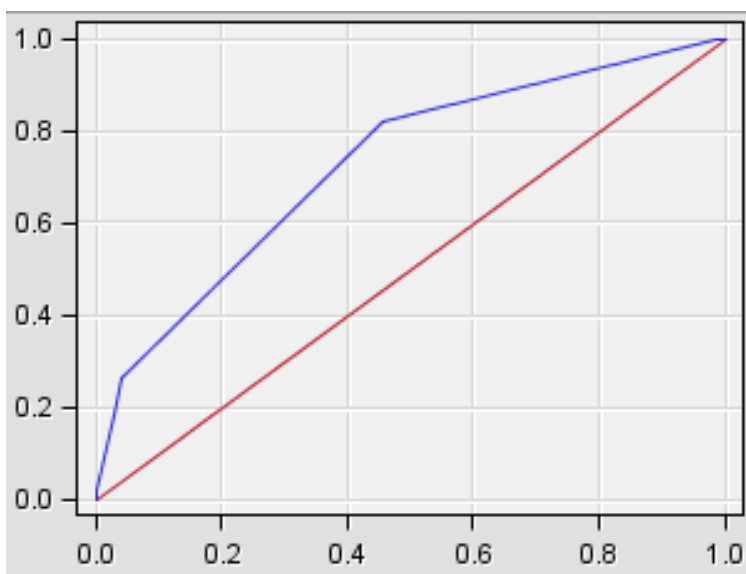


FIGURE 3.21 – Arbres de décision : Courbe ROC

Modèle	AUC	GINI
CART	0.726	0.452

FIGURE 3.22 – Arbres de décision : index ROC et coefficient de Gini

-Nous remarquons que l'indice de GINI = 0.452 est inférieur 0.5. Donc notre modèle est mauvais. Le modèle logistique étudié plus haut est plus performant en terme de prédire les clients qui vont tomber en défaut et ceux qui ne le vont pas que le modèle CART .

Forêts aléatoires

L'objectif est l'utilisation de forêts aléatoires pour la sélection des variables importantes à inclure dans le modèle. les variables à prendre en compte pour l'inclusion dans un modèle peuvent être classées par ordre de importance. L'indice d'importance variable (également appelé indice de Gini) basé sur les forêts prennent en compte l'interaction entre les variables. Cela en fait une méthode robuste pour trouver variables importantes qui peuvent être utilisées dans un modèle de prédiction. Cela implique le classement de variables explicatives (indépendantes) utilisant le score aléatoire des forêts d'importance.

Loss Reduction Variable Importance					
Variable	Number of Rules	Gini	OOB Gini	Margin	OOB Margin
checkingstatus1	6775	0.059087	0.04420	0.118173	0.103424
duration	12950	0.041978	0.00004	0.083957	0.040309
savings	4624	0.014451	-0.00203	0.028902	0.014507
purpose	6690	0.020713	-0.00438	0.041426	0.018175
history	5394	0.012648	-0.00621	0.025296	0.009768
property	5840	0.013584	-0.00994	0.027167	0.006198
age	12089	0.024216	-0.01907	0.048431	0.002500
amount	15242	0.036967	-0.02043	0.073933	0.012059

FIGURE 3.23 – Forêts aléatoires : Selection des variables

Le tableau ci-dessus montre les variables importantes, classées par ordre d'importance (du plus important au moins important) sont : **checkingstatus1**, **duration**, **savings**, **purpose**....

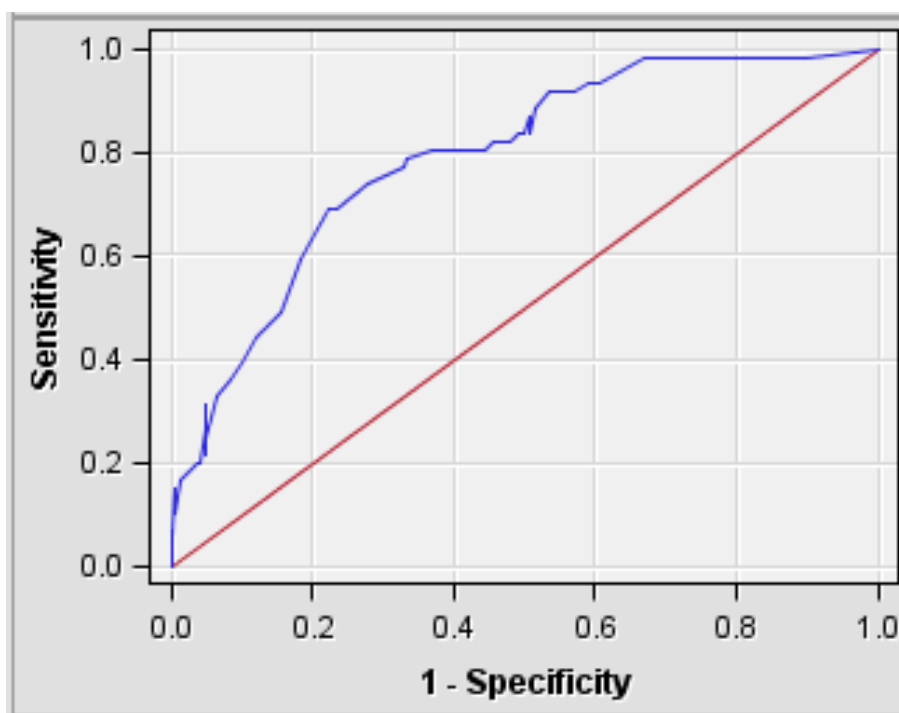


FIGURE 3.24 – Forêts aléatoires : Courbe ROC

Modèle	AUC	GINI
CART	0.793	0.585

FIGURE 3.25 – Forêts aléatoires : index ROC et coefficient de Gini

Nous remarquons que l'indice de GINI = 0.593 est supérieur 0.5. Donc notre modèle est bon. Le modèle logistique étudié plus haut est plus performant en terme de prédire les clients qui vont tomber en défaut et ceux qui ne le vont pas que le modèle CART

3.2.3 Choix du meilleur modèle

Dans cette partie finale de notre chapitre, on va comparer les trois méthodes exposées précédemment quant à leur performance vis à vis la prédiction des clients qui vont faire défaut dans le futur.

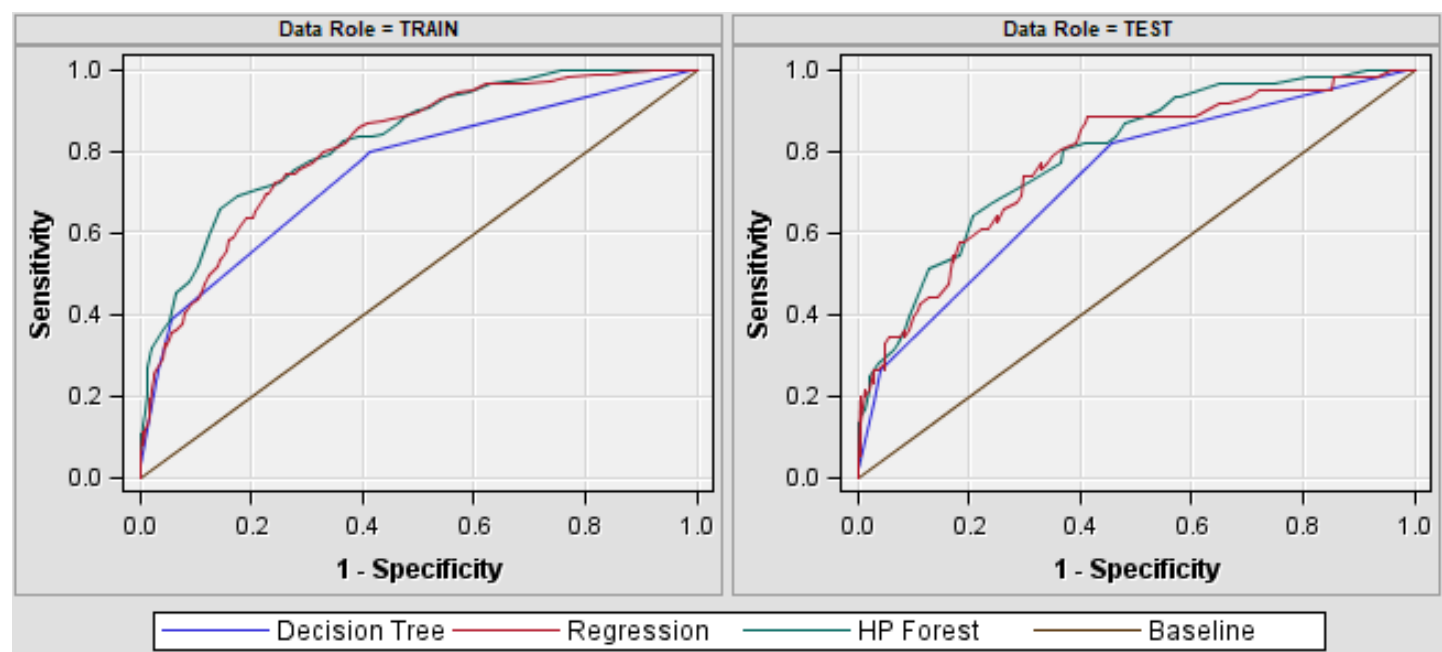


FIGURE 3.26 – Courbes ROC : comparaison des modèles

Modèle	AUC	GINI
CART	0.726	0.452
Logistique	0.778	0.556
Forêts aléatoires	0.793	0.586

FIGURE 3.27 – Indice Gini :Évaluation des modèles

D'après les deux figures ci-dessus, on constate que le type Forêts aléatoires est le plus performant puisque sa courbe est en dessus des autres, la figure vient confirmer cette constatation avec un $AUC(\text{Forêts aléatoires}) = 79.3\%$ et $Gini(\text{Forêts aléatoires}) = 58.6\%$ suivi par le modèle logistique tel que $AUC(\text{Logistique}) = 77.8\%$ et $Gini(\text{Logistique}) = 55.6\%$ et finalement le modèle CART tel que $AUC(\text{CART}) = 72.6\%$ et $Gini(\text{CART}) = 45.2\%$.

Chapitre 4

Modèle multipériodiques de probabilité de défaut PD

Sommaire

4.1	Modèles de risque en temps discret	65
4.1.1	L'événement par défaut	65
4.1.2	Les modèles non lineaire : Probit	65
4.1.3	Application notre base de donnée	66
4.2	Modèles de risque en temps continu	69
4.2.1	Modèle de survie : COX	69
4.2.2	Application à notre portefeuille	72
	Conclusion générale	79

4.1 Modèles de risque en temps discret

Les probabilités de défaut peuvent être modélisées avec des modèles d'aléa en temps discret. Les données doivent être organisées sous forme de panel.

Chaque ligne d'un ensemble de données de panel représente l'observation d'un prêt à chaque date t . Pour chaque individu, il y a autant d'observations que de dates $t = 1, 2, \dots, m$ où il est exposé au risque. On est donc amené à travailler avec un fichier de données "personne-période", il comporte 622489 lignes qui correspondent chacune à 1 mois d'observation d'un individu. A chaque observation personne-période on associe une variable de statut (default_tmtps dans notre base) qui prend la valeur 1 pour la dernière observation de chaque individu (celle où l'événement survient) et 0 pour les autres observations du même individu.

4.1.1 L'événement par défaut

Un PD décrit la probabilité d'un événement par défaut. Les banques observent si les emprunteurs défaut, et l'indiquent généralement avec un indicateur par défaut :

$$D_{it} = \begin{cases} 1 & \text{l'emprunteur } i \text{ fait défaut au moment } t \\ 0 & \text{sinon.} \end{cases}$$

Avec : $i = 1, \dots, I$ et $t = 1, \dots, T$

D variable aléatoire et d sa réalisation.

Un événement par défaut peut être défini par l'un des événements suivants :

- Faillite de l'emprunteur.
- Retard de paiement de plusieurs jours ou plus ; les seuils populaires sont 30, 60 et 90 jours.

Dans notre étude, on travaillera avec la définition par Bâle qui déclare qu'un client est en défaut si le nombre de jours impayés est 90 jours ou plus.

4.1.2 Les modèles non linéaire : Probit

L'équation du modèle probit est :
$$P(D_{it} = 1|X_{it-1}) = \Phi(\beta'x_{it-1})$$

Avec $\Phi(\cdot)$ la fonction de densité cumulée de la distribution normale standard, β vecteur des paramètres et x_{it-1} vecteur des covariables.

Estimation de vraisemblance maximale

La vraisemblance est définie comme suit :

$$L(\beta, x_{it-1}) = \prod_{i=1}^I \prod_{t=1}^T ((P(D_{it} = 1))^{d_{it}} \cdot (1 - P(D_{it} = 1))^{1-d_{it}})$$

La vraisemblance est alors transformée par le logarithme naturel monotone (c.-à-d. log-vraisemblance) :

$$\log L(\beta, x_{it-1}) = \sum_{i=1}^I \sum_{t=1}^T ((d_{it} \log(P(D_{it} = 1))) \cdot (1 - d_{it}) \log(1 - P(D_{it} = 1)))$$

Tout au long du cycle (TTC) versus Point-in-Time (PIT)

- Une mesure du risque de crédit **PIT** utilise toutes les informations disponibles et pertinentes à une date donnée pour estimer la PD d'un client sur un horizon temporel donné. L'ensemble d'informations comprend non seulement les attentes sur l'évolution du risque de crédit d'un client à long terme, mais également les tendances macroéconomiques. Donc la probabilité de défaut du modèle PIT va expliquer d'une manière plus précise et plus adéquate le taux de défaut réel.
- Une mesure du risque de crédit **TTC** ne prend généralement pas en compte l'état de l'économie en excluant les variables macroéconomiques .

4.1.3 Application notre base de donnée

Notre base de données sur les prêts se présente sous la forme d'un panel et présente des observations sur la création et le rendement de 50000 emprunteurs sur 60 périodes. Comme dans le monde réel, les prêts peuvent naître avant le début de la période d'observation . Les observations

de prêt peuvent ainsi être censurées à mesure que les prêts arrivent à échéance .

Notre rapport montre comment travailler avec les données du panel de crédit pour créer un modèles à travers la méthodologie (PIT) en utilisant le modèle probit.

La figure suivante est la sortie du modèle probit sous SAS en modélisant la probabilité que le client tombe en défaut par les variables **Score** du client, **rapport valeur prêt** et les variables macroéconomiques **PIB**, et **taux d'intérêt** .

Nous constatons que toutes les p-values sont inférieures à 5%. Les quatre variables exogènes sont donc statistiquement significatives et contribuent donc l'explication de la probabilité qu'un contrat connaisse un défaut.

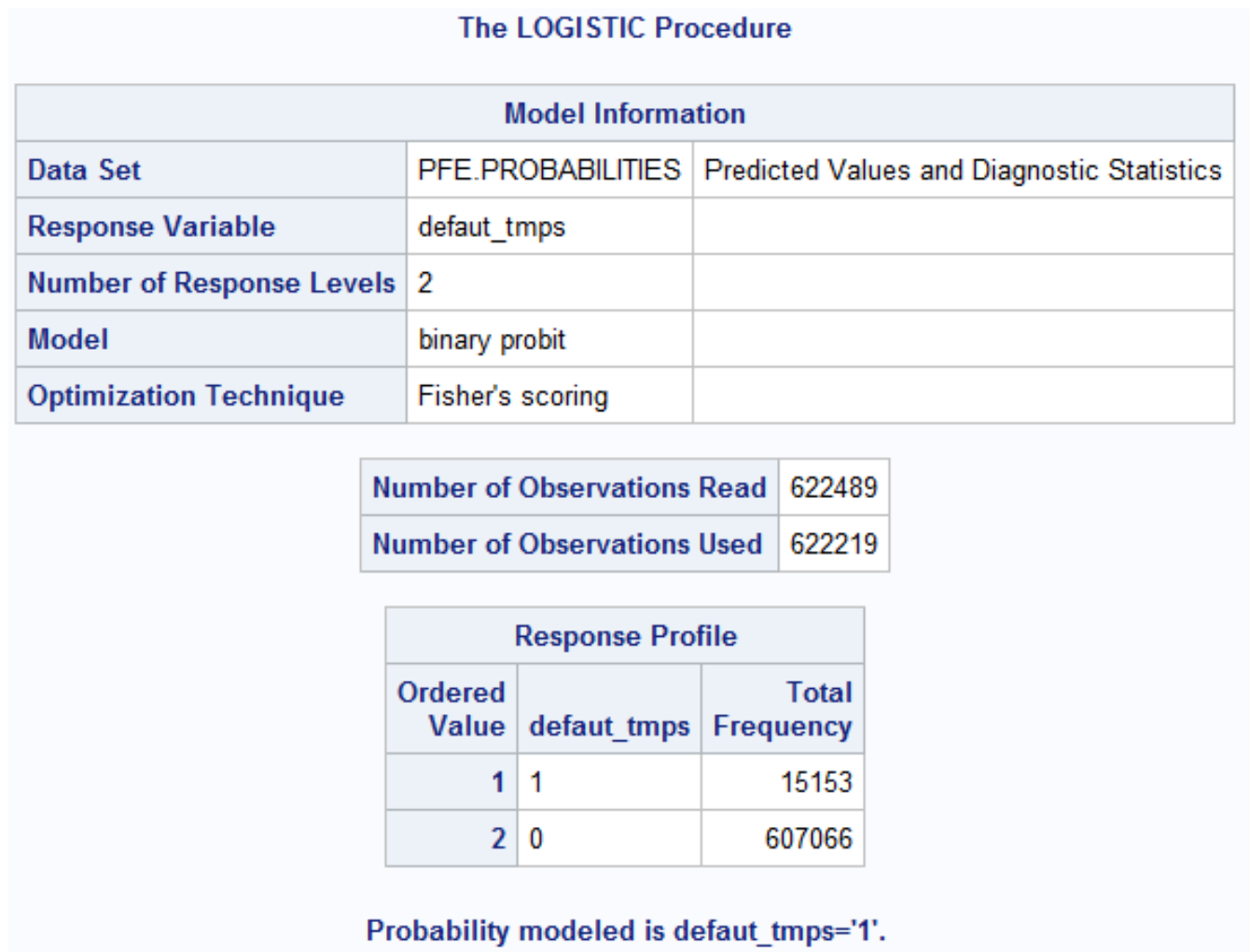


FIGURE 4.1 – Modèle probit

Notre tableau d'estimation des paramètres indique que notre modèle contient la constante qui égale

-2.0312, on constate que le paramètre estimé du score est négatif, ceci peut être expliqué par le fait que les clients qui ont un bon score lors de l'octroi du crédit tendent à avoir un comportement assidu quant au paiement de leur crédit.

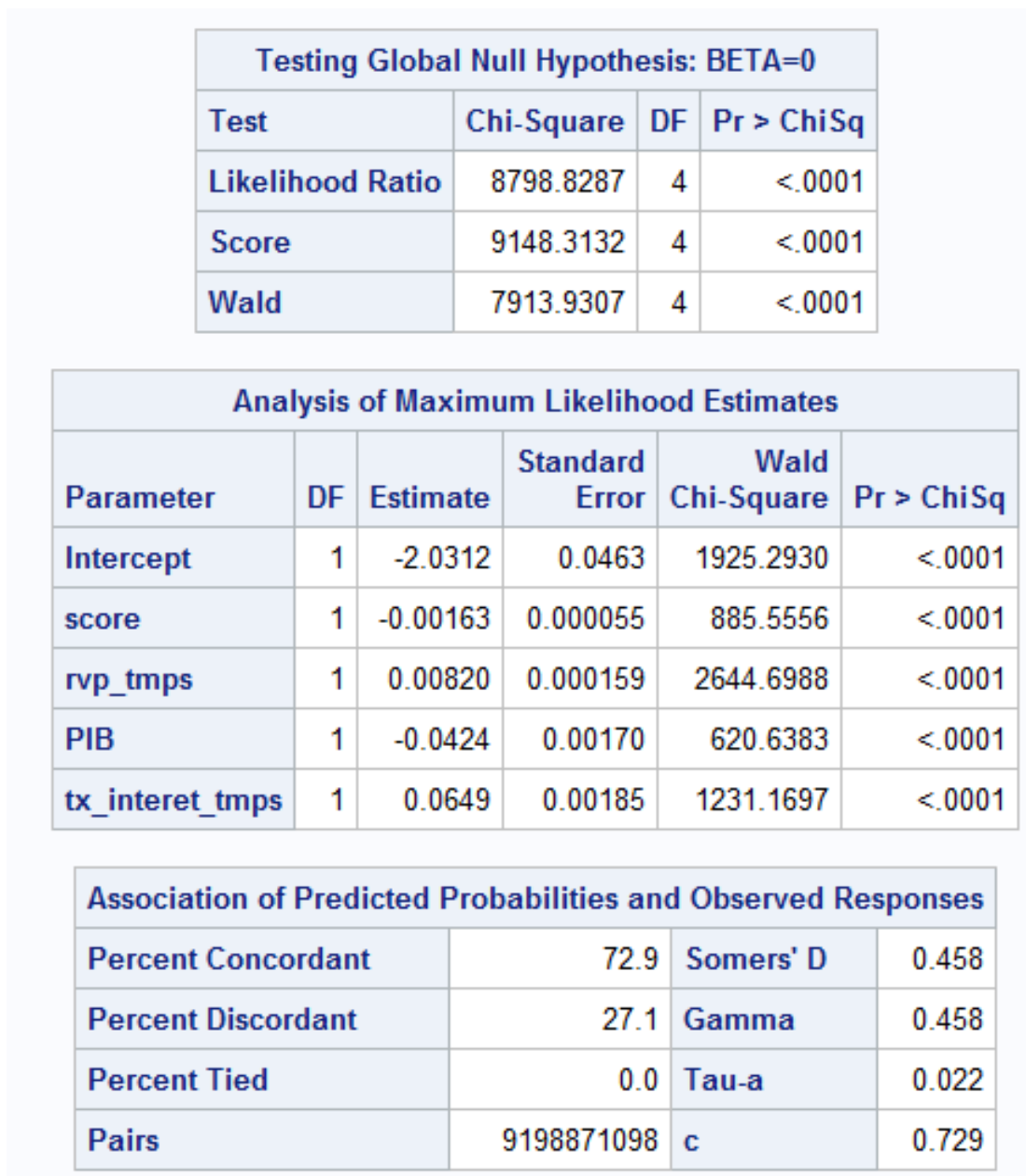


FIGURE 4.2 – Modèle probit : PIT

Nous remarquons aussi que le coefficient associé à la variable "PIB" est négative et l'OR de la variable est inférieur à 1, $\exp(-0.0424) = 0,9584$. Nous pouvons affirmer donc qu'à mesure que le PIB augmente, les chances d'effectuer un défaut diminuent.

Nous constatons aussi que le coefficient associé à la variable "rvp_tmps" est positive. Dans le

même sens, on a l'OR associé à cette variables calculé par qui est supérieur à 1. Cela signifie que plus le rapport prêt valeur est important, plus la probabilité de faire un défaut augmente.

Nous remarquons aussi que le coefficient associé à la variable "taux d'intérêt" est positive et l'OR de la variable est supérieur à 1, $\exp(0.0649) = 1,067$. Nous pouvons affirmer donc qu' mesure que le montant bloqué augmente, plus la probabilité de faire un défaut augmente.

Conclusion : Le pourcentage concordant et le coefficient c (AUC) nous informe sur l'adaptation du modèle aux données. Ici on a un pourcentage de 72,9% de concordance entre le modèle et la réalité. Ainsi, le modèle peut être jugé bien adapté avec la réalité observée.

4.2 Modèles de risque en temps continu

4.2.1 Modèle de survie : COX

Vu qu'on étudie le défaut temps et que celui-ci représente la durée jusqu'à la survenance de l'événement terminale « Défaut », de plus du fait que cette étude est limitée dans le temps. C'est donc naturellement qu'on se retourne vers les modèles d'analyse de survie, plus précisément le modèle des risques proportionnels de Cox « CPH ». On note aussi plusieurs autres approches telles celle de l'estimateur de Kaplan Meier ou encore le modèle AFT.

Mais ce qui caractérise le plus ces modèles, c'est le fait que dans le temps prévu par l'étude on n'observe pas forcément l'évènement terminale chez tous les clients. On parle alors de données censurées où la durée de survie au défaut n'est pas connue pour tous. Ainsi n dirait que la variable dépendante est constituée de deux paramètres importants :

- Le temps de survie qui est une variable positive continue,
- La survenance ou non du défaut qui est une variable discrète : variable binaire prenant un 0 en cas de non-défaut et 1 en cas de défaut.

Représentation du modèle « CPH » et ses fonctions

Pour commencer, on va d'abord définir la notion de censure, un phénomène lié à la disponibilité des informations :

- Censure à droite : évènement non survenu avant la fin de la période d'observation,
- Censure à gauche : évènement survenu avant la date du point et donc sans que l'on en connaisse la date.

Donc pour notre étude, on considère l'évènement qu'est le « défaut » de l'entreprise. Notre approche reposera alors sur l'estimation de la fonction de survie S_t qui représente la probabilité qu'un individu ne vivra pas le défaut avant t .

$$S_t = P(T \geq t) \quad , t \geq 0 \quad \text{Avec } T : \text{variable aléatoire positive et continue.}$$

Pour ce qui est du modèle de CPH on s'intéresse à la fonction du risque instantané de survenance du défaut :

$$h(t|X) = h_0(t) * \exp(\beta_1 * X_1 + \dots + \beta_n * X_n)$$

Avec : $X = (X_1, \dots, X_n)$: qui représente le vecteur des variables explicatives,

Et les $\beta = (\beta_1, \dots, \beta_n)$: sont les coefficients de régression de CPH,

$h_0(t)$: est la fonction de risque instantané de base.

Ainsi, on liera la fonction de survie à la fonction de risque instantané par l'équation suivante :

$$S_t = \exp\left(-\int_0^t h(x)dx\right)$$

L'estimation des coefficients se fera par la méthode du maximum de vraisemblance.

Hypothèses du modèle à vérifier

Afin d'effectuer un modèle de Cox, il faudra d'abord vérifier les deux principales hypothèses suivantes :

- • La loglinéarité des variables : vérifier que le logarithme de $h(t)$ est une fonction linéaire des variables explicatives $X = (X_1, \dots, X_n)$. Ceci peut être fait grâce à une évaluation

graphique.

- • La proportionnalité des risques : les coefficients de régression doivent être indépendants du temps.

Tests de significativité des coefficients

L'objectif des tests de significativité est d'éprouver le rôle d'une, de plusieurs, de l'ensemble des variables explicatives.

Dans le cas où l'on veut tester la significativité d'un seul coefficient c'est-à-dire :

$$H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0$$

Tandis que si l'on veut tester la significativité de plusieurs coefficients :

$$H_0 : \beta_1 = \dots = \beta_k = 0 \quad vs \quad H_1 : \exists i \text{ tel que } \beta_i \neq 0$$

On va utiliser les tests suivants pour tester la significativité des variables explicatives et l'adéquation du modèle :

- • Test du rapport de vraisemblance
- • Test du Score
- • Test du Wald

Validation du modèle de survie

On refait les mêmes étapes faites dans les modélisations passées (validation des données, de la méthodologie et des tests statistiques) : on prédit par l'estimation basé sur le modèle CPH, puis on calcule la courbe ROC pour différents intervalles h. Puis on passe à l'étalonnage.

À commencer par la précision du modèle : il faut savoir que tous les modèles de survie fournissent une estimation sur un horizon donné. Cependant, on peut être intéressé par des sous-périodes. Dans

ce cas on la déduit grâce au rapport suivant :

$$\widehat{PD}_{i,s,t_j} = \widehat{PD}_{i,s,(t_{t_j,init},t_{t_j,end})} = \frac{\widehat{S}_{t_{t_j,end}} - \widehat{S}_{t_{t_j,init}}}{\widehat{S}_{t_{t_j,init}}}$$

Avec $t_{t_j,init}$ et $t_{t_j,end}$ le début et la fin de l'intervalle t_j .

4.2.2 Application à notre portefeuille

Avant de procéder à la modélisation et aux différents tests, on commence par traiter les données en se concentrant sur la censure. On note ainsi que l'indicateur « default_tmps » est égale à 0 si l'observation est censurée et 1 si le défaut arrive. On néglige la censure à gauche, en considérant que tous les crédits commencent par la première période d'observation.

Le traitement des données consistera alors à classer les observations par client en utilisant la variable « id » afin de faciliter le constat de chaque client sur si ou non il a fait défaut ou a remboursait son crédit. Puis on passe au calcul des durées de vie du crédit pour chacun sous le nom « duree ».

id	temps	orig_time	first_time	default_tmps	payoff_time	rvp_tmps	PIB	tx_chmge_tmps	tx_interet_tmps	Interest_Rate_orig_time	hpi_time
1	25	-7	25	0	0	24.498336255	2.8991367167	4.7	9.2	9.2	226.29
1	26	-7	25	0	0	24.483867189	2.1513648831	4.7	9.2	9.2	225.1
1	27	-7	25	0	0	24.626794958	2.3617217155	4.4	9.2	9.2	222.39
1	28	-7	25	0	0	24.735883472	1.2291722496	4.6	9.2	9.2	219.67
1	29	-7	25	0	0	24.92547567	1.6929686703	4.5	9.2	9.2	217.37
1	30	-7	25	0	0	25.318291178	2.2742178006	4.7	9.2	9.2	212.73
1	31	-7	25	0	0	26.566116797	1.8506891663	4.7	9.2	9.2	200.67
1	32	-7	25	0	0	25.872558649	1.1041627602	5	9.2	9.2	186.12
1	33	-7	25	0	0	25.584425341	0.8368586627	5	9.2	9.2	180.52
1	34	-7	25	0	0	26.008072958	-0.314447747	5.8	9.2	9.2	173.35
1	35	-7	25	0	0	27.286499381	-2.805844087	6.5	9.2	9.2	162.09
1	36	-7	25	0	0	28.963629196	-3.516567987	7.8	9.2	9.2	151.48
1	37	-7	25	0	0	28.347862215	-4.146710947	9	9.2	9.2	153.35
1	38	-7	25	0	0	27.38013901	-3.339549297	9.5	9.2	9.2	158.77
1	39	-7	25	0	0	27.485740204	-0.241080652	10	9.2	9.2	158.16
1	40	-7	25	0	0	27.828850078	1.5859655122	9.8	9.2	9.2	156.21
1	41	-7	25	0	0	26.687874923	2.6824968062	9.9	9.2	9.2	161.06
1	42	-7	25	0	0	26.387715083	3.0294986398	9.4	9.2	9.2	161
1	43	-7	25	0	0	26.899116563	2.6941105489	9.4	9.2	9.2	156.04
1	44	-7	25	0	0	27.2027313	1.8755149375	9.2	9.2	9.2	150.91
1	45	-7	25	0	0	26.100975904	1.6391549258	9.1	9.2	9.2	154.87
1	46	-7	25	0	0	25.58158643	1.1759090317	9	9.2	9.2	155.61
1	47	-7	25	0	0	26.229863984	1.6681823188	8.8	9.2	9.2	149.58
1	48	-7	25	1	0	26.658065492	2.7159028846	8.3	9.2	9.2	146.45
2	18	18	25	0	0	65.231342577	2.8991367167	4.7	7.68	7.68	226.29
2	26	18	25	0	1	65.46985108	2.1513648831	4.7	7.68	7.68	225.1
3	25	-6	25	0	0	30.551659012	2.8991367167	4.7	10.5	8.75	226.29
3	26	-6	25	0	0	30.626801806	2.1513648831	4.7	10.5	8.75	225.1
3	27	-6	25	0	0	30.913949712	2.3617217155	4.4	11.375	8.75	222.39
3	28	-6	25	0	0	31.214709602	1.2291722496	4.6	11.375	8.75	219.67

FIGURE 4.3 – Aperçu de la base de données initiale

id	time	orig_time	first_time	mat_time	balance_time	LTV_time	Interest_rate_time	hpi_time	gdp_time	uer_time	RBype_CO_orig_time	RBype_PU_orig_time
1	48	-7	25	113	29087.21	26.658065492	9.2	146.45	2.7159028846	8.3	0	0
2	26	18	25	138	105654.77	65.46985108	7.68	225.1	2.1513648831	4.7	0	0
3	29	-6	25	114	44378.6	31.459734885	11.375	217.37	1.6929686703	4.5	0	0
4	60	-2	25	119	52686.35	34.898841933	10.5	189.82	2.8363581757	5.7	0	0
5	27	18	25	138	52100.71	66.346342871	9.155	222.39	2.3617217155	4.4	0	0
6	56	19	25	139	190474.11	75.83479541	6.58	181.43	1.7170529058	6.6	0	0
7	26	18	25	138	107916.38	77.919574241	9	225.1	2.1513648831	4.7	0	0
8	25	18	25	138	152393.21	58.819187192	7.25	226.29	2.8991367167	4.7	0	0
9	37	18	25	138	130140.31	99.13810459	8	153.35	-4.146710947	9	0	0
10	29	18	25	139	88046.35	67.296389785	10.23	217.37	1.6929686703	4.5	0	0
11	27	18	25	138	154664.04	78.781217589	9.13	222.39	2.3617217155	4.4	0	0
12	27	19	25	139	324361.02	67.274831362	5.75	222.39	2.3617217155	4.4	0	0

FIGURE 4.4 – Aperçu de la base de données après traitement des données

type_CO_orig_time	RBype_PU_orig_time	RBype_SF_orig_time	investor_orig_time	balance_orig_time	FICO_orig_time	LTV_orig_time	Interest_Rate_orig_time	hpi_orig_time	defaut_time	payoff_time	status_time	duree	indcator
0	0	1	0	45000	715	69.4	9.2	87.03	1	0	1	24	1
0	0	1	0	107200	558	80	7.68	106.91	0	1	2	2	1
0	0	1	0	48600	680	83.6	8.75	89.58	0	1	2	5	1
0	0	1	0	63750	587	81.8	10.5	97.99	0	0	0	36	1
0	0	1	0	52800	527	80	9.155	106.91	0	1	2	3	1
0	0	1	0	212000	670	80	6.58	191.42	1	0	1	32	1
0	0	1	0	108250	601	95	9	106.91	0	1	2	2	1
0	0	1	0	192600	663	90	7.25	106.91	0	1	2	1	1
0	0	1	0	128000	501	80	8	106.91	1	0	1	13	1

FIGURE 4.5 – Aperçu de la nouvelle variable « duree »

Maintenant qu'on obtient de façon plus claire la durée et l'état du crédit de chaque client, on procède à la construction du modèle continu. Ce dernier sera construit à l'aide des 2 variables suivantes :

- ● « fico_orig_time » : c'est le score crédit. En pratique, il peut être déduit des probabilités de défaut sur un an grâce à la calibration.
- ● « Ltv_orig_time » : c'est le ratio Loan To Value.

La modélisation se fera grâce à la fonction « PROC PHREG » vu qu'on veut un modèle à risque proportionnel, puis on générera la fonction de survie. On prend comme variable dépendante « duree » et la variable censurée « default_time ».

On obtient alors un total de 34846 de valeurs censurées, soit un pourcentage de 69.69% du total d'observation.

Donc toutes les variables insérées sont significatives. On constate du signe positif de la variable « LTV_orig_time » que le risque augmente avec le ratio Loan To Value, alors que le contraire arrive en cas d'augmentation du score crédit.

On a aussi le test global BETA=0 qui présente des p-value inférieur à 5%, ce qui signifie qu'au moins un des facteurs étudiés joue un rôle dans le modèle.

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi-2	DDL	Pr > Khi-2
Likelihood Ratio	2164.1072	2	<.0001
Score	2198.7891	2	<.0001
Wald	2183.5710	2	<.0001

Estimations par l'analyse du maximum de vraisemblance						
Paramètre	DDL	Valeur estimée des paramètres	Erreur type	Khi-2	Pr > Khi-2	Rapport de risque
FICO_orig_time	1	-0.00442	0.0001114	1571.3114	<.0001	0.996
LTV_orig_time	1	0.01554	0.0008019	375.6040	<.0001	1.016

FIGURE 4.6 – Sortie de la fonction « PROC PHREG » montrant les coefficients du modèle

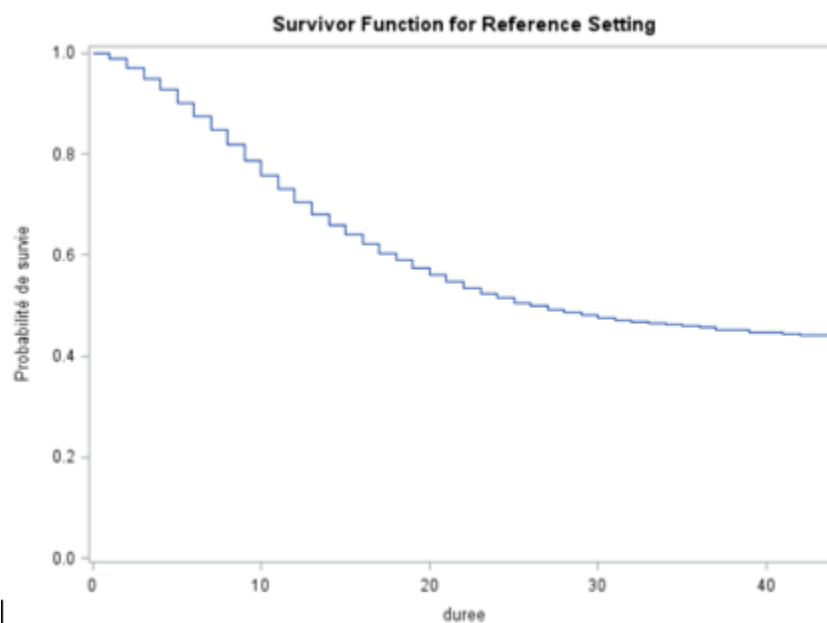


FIGURE 4.7 – Graphe montrant la probabilité de survie en fonction de la durée du crédit

Ici, vu qu'on a élaboré un modèle de Cox de base, on a donc supposé que les caractéristiques individuelles sont constantes dans le temps pendant la période observée. Cependant, beaucoup de caractéristiques peuvent évoluer avec le temps. On note comme exemple le taux d'intérêt ainsi que le PIB et plein d'autre.

Pour intégrer cette évolution dans notre modèle, on considère que les variables indépendantes, supposées fixes dans ce qui précède comme le ratio Loan to Value, varient au fil du temps et on ajoute aussi le taux d'intérêt ainsi que le PIB. Le principe repose alors sur des sous épisodes de durée définie, et dont on connaît la date de début. On parle de covariables variant dans le temps, tel qu'on commence un nouvel épisode chaque fois que l'une des covariables change de valeur.

On définit les durées d'épisode en 1 mois pour chacun ainsi que la date de début de chaque épisode.

```

/* covariable variant dans le temps*/
data pfe.phreg;
set pfe.mort;
time1=time-first_time;
time2=time-first_time+1;
run;

```

FIGURE 4.8 – Code utilisé pour générer les deux variables définissant le début et la fin de chaque épisode

On construit alors notre nouveau modèle grâce aux variables suivantes :

- ● « Fico_orig_time » : le score crédit calculé à l'origine.
- ● « interest_rate_time » : taux d'intérêt évoluant avec le temps
- ● « gdp_time » : PIB évoluant avec le temps
- ● « Ltv_time » : Loan to value évoluant avec le temps

La figure ci-dessus montre que notre modèle est globalement significatif. De même les p-value des différentes variables nous montrent qu'elles ont toutes significatives.

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi-2	DDL	Pr > Khi-2
Likelihood Ratio	5612.9774	4	<.0001
Score	6949.7801	4	<.0001
Wald	8180.6688	4	<.0001

Estimations par l'analyse du maximum de vraisemblance						
Paramètre	DDL	Valeur estimée des paramètres	Erreur type	Khi-2	Pr > Khi-2	Rapport de risque
FICO_orig_time	1	-0.00382	0.0001232	958.6669	<.0001	0.996
LTV_time	1	0.00868	0.0001234	4952.2684	<.0001	1.009
gdp_time	1	-0.09781	0.00395	614.1728	<.0001	0.907
interest_rate_time	1	0.11526	0.00407	803.9748	<.0001	1.122

FIGURE 4.9 – Modèle de survie CPH covariables variant dans le temps

Et pour finir, on génère la probabilité de survie pour chaque client, en se référant au client "i" par la variable « id » égale à « i », pour un intervalle de temps allant de 1 mois à 44 mois.

_orig_time	default_time	payoff_time	status_time	time1	time2	Survival Function Estimate
87.03	0	0	0	0	0	1
87.03	0	0	0	0	1	0.993039928
87.03	0	0	0	0	2	0.9822170059
87.03	0	0	0	0	3	0.9692751189
87.03	0	0	0	0	4	0.9562240473
87.03	0	0	0	0	5	0.9423951907
87.03	0	0	0	0	6	0.929063628
87.03	0	0	0	0	7	0.9161577979
87.03	0	0	0	0	8	0.9024897019
87.03	0	0	0	0	9	0.8894503535
87.03	0	0	0	0	10	0.8769112233
87.03	0	0	0	0	11	0.8646179853
87.03	0	0	0	0	12	0.8527664608
87.03	0	0	0	0	13	0.8412215894
87.03	0	0	0	0	14	0.8296260768
87.03	0	0	0	0	15	0.8189557213
87.03	0	0	0	0	16	0.8085177061
87.03	0	0	0	0	17	0.7989589467
87.03	0	0	0	0	18	0.7894867558
87.03	0	0	0	0	19	0.7799336801
87.03	0	0	0	0	20	0.7701180017
87.03	0	0	0	0	21	0.7604098137
87.03	0	0	0	0	22	0.7515999389
87.03	0	0	0	0	23	0.7431275584
87.03	0	0	0	0	24	0.7355359106
87.03	0	0	0	0	25	0.7279533559
87.03	0	0	0	0	26	0.7217905587
87.03	0	0	0	0	27	0.7158559438
87.03	0	0	0	0	28	0.7111153513

FIGURE 4.10 – aperçu des probabilités de survie du client 1

On déduit les probabilités de défaut en fonction du temps, du ratio Loan to Value, du PIB, du taux d'intérêt et du score à l'origine.

	LTV_time	interest_rate_time	gdp_time	FICO_orig_time	time2	PD_time
79920	82.60988107	6.875	2.2742178006	714	29	0.0095086375
79921	82.60988107	6.875	2.2742178006	714	30	0.0100810767
79922	82.60988107	6.875	2.2742178006	714	31	0.0093555626
79923	82.60988107	6.875	2.2742178006	714	32	0.0082044347
79924	82.60988107	6.875	2.2742178006	714	33	0.0071403541
79925	82.60988107	6.875	2.2742178006	714	34	0.0048484801
79926	82.60988107	6.875	2.2742178006	714	35	0.0050555697
79927	82.60988107	6.875	2.2742178006	714	36	0.0081563435
79928	82.60988107	6.875	2.2742178006	714	37	0.007003054
79929	82.60988107	6.875	2.2742178006	714	38	0.0053567895
79930	82.60988107	6.875	2.2742178006	714	39	0.0093606307
79931	82.60988107	6.875	2.2742178006	714	40	0.0036451738
79932	82.60988107	6.875	2.2742178006	714	41	0.0048705305
79933	82.60988107	6.875	2.2742178006	714	42	0.0055922127
79934	82.60988107	6.875	2.2742178006	714	43	0.0027476429
79935	82.60988107	6.875	2.2742178006	714	44	0.0204692022
79936	87.574625006	6.875	1.8506891663	714	0	0
79937	87.574625006	6.875	1.8506891663	714	1	0.0102229468
79938	87.574625006	6.875	1.8506891663	714	2	0.0159931879
79939	87.574625006	6.875	1.8506891663	714	3	0.0193246964
79940	87.574625006	6.875	1.8506891663	714	4	0.0197465792
79941	87.574625006	6.875	1.8506891663	714	5	0.0212039366
79942	87.574625006	6.875	1.8506891663	714	6	0.020742946
79943	87.574625006	6.875	1.8506891663	714	7	0.020369917
79944	87.574625006	6.875	1.8506891663	714	8	0.0218715949
79945	87.574625006	6.875	1.8506891663	714	9	0.0211838522
79946	87.574625006	6.875	1.8506891663	714	10	0.0206715566
79947	87.574625006	6.875	1.8506891663	714	11	0.0205563627
79948	87.574625006	6.875	1.8506891663	714	12	0.020101002
79949	87.574625006	6.875	1.8506891663	714	13	0.0198538249
79950	87.574625006	6.875	1.8506891663	714	14	0.0202133997

FIGURE 4.11 – Aperçu des Probabilité de défaut pour des durées allant de 1 mois à 44 mois

Pour la validation du modèle, on sait que le modèle d'analyse de survie CPH reste très flexible vu qu'il ne suppose aucune distribution des durées des épisodes. Cependant on passe par deux hypothèses qu'il faut tout de même vérifier : la log-linéarité pour les variables continues et les risques proportionnels.

Pour ce qui est des risques proportionnels, on va se baser sur les résidus de Schoenfeld :

Pour calculer le résidu de Schoenfeld d'une variable X, on compare la valeur de X à un temps t pour un client « i » à la valeur moyenne parmi les clients encore à risque de défaut à cette instant t. Il faut donc noter que ces résidus ne sont calculés que pour les individus non censurés à chaque instant t. Ainsi pour tester notre hypothèse, on test l'existence, ou pas, d'une corrélation entre les résidus de Schoenfeld et le temps.

Ainsi si on constate que le risque est constant dans le temps, on admet l'hypothèse de proportionnalité.

Statistiques simples							
Variable	N	Moyenne	Ecart-type	Somme	Minimum	Maximum	Libellé
time2	622489	12.19344	9.99615	7590285	1.00000	60.00000	
rfico	15153	-0.00191	66.04823	-28.99095	-245.59590	167.73693	Schoenfeld Residual for FICO_orig_tir
rltv	15153	-0.00921	19.97865	-139.48528	-94.48471	101.62255	Schoenfeld Residual for LTV_time
rgdp	15153	-0.0000260	1.99072	-0.39420	-6.36558	5.18669	Schoenfeld Residual for gdp_time
rinterest	15153	-0.0004772	1.85198	-7.23025	-6.79701	8.16159	Schoenfeld Residual for interest_rate_!

Coefficients de corrélation de Pearson					
Proba > r sous H0: Rho=0					
Nombre d'observations					
	time2	rfico	rltv	rgdp	rinterest
time2	1.00000	0.08750	-0.01330	0.10243	-0.17483
		<.0001	0.1016	<.0001	<.0001
	622489	15153	15153	15153	15153
rfico	0.08750	1.00000	0.19445	-0.09319	-0.37464
Schoenfeld Residual for FICO_orig_time	<.0001		<.0001	<.0001	<.0001
	15153	15153	15153	15153	15153
rltv	-0.01330	0.19445	1.00000	-0.37326	-0.17310
Schoenfeld Residual for LTV_time	0.1016	<.0001		<.0001	<.0001
	15153	15153	15153	15153	15153
rgdp	0.10243	-0.09319	-0.37326	1.00000	0.04179
Schoenfeld Residual for gdp_time	<.0001	<.0001	<.0001		<.0001
	15153	15153	15153	15153	15153
rinterest	-0.17483	-0.37464	-0.17310	0.04179	1.00000
Schoenfeld Residual for interest_rate_time	<.0001	<.0001	<.0001	<.0001	
	15153	15153	15153	15153	15153

FIGURE 4.12 – Sortie des résidus de Schoenfeld de notre modèle

En se basant sur les coefficients de corrélation dans la sortie, on constate qu'ils sont tous presque proche de la valeur 0, donc la relation linéaire entre les résidus de Schoenfeld de chaque variable explicative du modèle et le temps est faible. Et donc on admet avec une probabilité de faire une erreur de type 1 égale à 5% que l'hypothèse est vérifiée.

Conclusion générale

Notre stage PFE au sein du cabinet Grant Thornton nous a rapproché du monde de l'ingénierie financière. Une expérience qui nous a introduit à la nouvelle norme comptable IFRS9. Cette dernière a permis aux banques et aux institutions financières de faire un grand pas dans l'analyse risque crédit. Ainsi en remplaçant le modèle de dépréciation des pertes encourues de l'IAS 39 par le modèle des pertes attendues, l'IFRS 9 permet aux financiers d'anticiper les pertes liées au défaut avant l'apparition d'éléments déclencheurs de défaut et ainsi éviter d'importantes pertes. Un modèle très intéressant dont le principe repose sur le passage du calcul des ECL sur 12 mois au calcul de celle-ci sur la durée du crédit lorsque ce dernier connaît une augmentation significative du risque de défaut. Un sujet d'actualité très entichissant, qui nous a beaucoup appris en modélisant les probabilités de défaut par plusieurs méthodes : à savoir « les modèles linéaires généralisés » par le modèle logistique ainsi que le « machine learning » grâce à l'arbre « CART » et aux « forêts aléatoires ». On a aussi eu l'opportunité de renforcer nos compétences en analyse de survie grâce à la modélisation des probabilités de défaut sur le long terme par le « modèle à risques proportionnels de Cox » ainsi que l'utilisation du modèle de PIT . Ce qu'on a retenu dans notre étude c'est l'introduction des variables macroéconomiques et des variables sensibles aux fluctuations temporelles pour les modèles à risques proportionnels et PIT, donnant ainsi le moyen de prédire l'évolution des probabilités de défaut dans le temps. Ainsi une fois que les probabilités de défaut sur un an sont calculées, on peut les transformer en score crédit par calibration. Ce dernier est utilisé, en cas d'augmentation significative du risque, à l'élaboration des probabilités de défaut sur toute la durée du crédit en y ajoutant les variables macro-économiques tel le PIB et le taux d'intérêt. Grâce à ces probabilités, on peut prédire le comportement de nos clients et ainsi juger si ils sont solvables ou pas. Cette expérience a aussi été riche sur le plan personnel, tel que confronté à la situation sanitaire actuelle, on s'est retrouvé devant un vrai défi à gérer le stage à distance et à chercher une deuxième base de données pour la modélisation comportant les variables macroéconomiques.

Annexe A

Annexe

age_IV						
Obs	nondefault	default	somme	somme1	woe	iv
1	63	42	700	300	-0.44183	0.02209
2	47	38	700	300	-0.63474	0.03778
3	74	27	700	300	0.16093	0.00253
4	79	41	700	300	-0.19142	0.00456
5	72	33	700	300	-0.06714	0.00048
6	55	17	700	300	0.32682	0.00716
7	89	24	700	300	0.46328	0.02184
8	70	28	700	300	0.06899	0.00046
9	84	21	700	300	0.53900	0.02695
10	67	29	700	300	-0.00990	0.00001
	700	300				0.12386

Duration_IV						
Obs	nondefault	default	somme	somme1	woe	iv
1	84	10	700	300	1.28093	0.11101
2	69	17	700	300	0.55360	0.02320
3	130	49	700	300	0.12842	0.00287
4	59	13	700	300	0.66529	0.02725
5	72	43	700	300	-0.33183	0.01343
6	30	10	700	300	0.25131	0.00239
7	128	56	700	300	-0.02062	0.00008
8	11	6	700	300	-0.24116	0.00103
9	75	51	700	300	-0.46164	0.02902
10	42	45	700	300	-0.91629	0.08247
	700	300				0.29275

Amount_IV						
Obs	nondefault	default	somme	somme1	woe	iv
1	68	31	700	300	-0.06178	0.00038
2	72	30	700	300	0.02817	0.00008
3	73	26	700	300	0.18507	0.00326
4	78	22	700	300	0.41837	0.01594
5	70	30	700	300	0.00000	0.00000
6	76	24	700	300	0.30538	0.00873
7	78	23	700	300	0.37392	0.01300
8	70	29	700	300	0.03390	0.00011
9	62	38	700	300	-0.35775	0.01363
10	53	47	700	300	-0.72715	0.05886
	700	300				0.11399

FIGURE A.1 – Information value : Age, Duration et Amount

checkingstatus1_IV					
Obs	checkingstatus1	nondefault	default	woe	IV
1	A11	139	135	-0.81810	0.20569
2	A12	164	105	-0.40139	0.04645
3	A13	49	14	0.40547	0.00946
4	A14	348	46	1.17626	0.40441
		700	300		0.66601

history_IV					
Obs	history	nondefault	default	woe	IV
1	A30	15	25	-1.35812	0.08407
2	A31	21	28	-1.13498	0.07188
3	A32	361	169	-0.08832	0.00421
4	A33	60	28	-0.08516	0.00065
5	A34	243	50	0.73374	0.13242
		700	300		0.29323

Purpose_IV					
Obs	purpose	nondefault	default	woe	IV
1	A40	145	89	-0.35920	0.03216
2	A41	93	22	0.59426	0.03537
3	A42	123	58	-0.09556	0.00168
4	A43	218	62	0.41006	0.04296
5	A44	8	4	-0.15415	0.00029
6	A45	14	8	-0.28768	0.00192
7	A46	28	22	-0.60614	0.02020
8	A48	8	1	1.23214	0.00997
9	A49	63	34	-0.23052	0.00538
		700	300		0.14994

property_IV					
Obs	property	nondefault	default	woe	IV
1	A121	222	60	0.46103	0.05401
2	A122	161	71	-0.02857	0.00019
3	A123	230	102	-0.03419	0.00039
4	A124	87	67	-0.58608	0.05805
		700	300		0.11264

Savings_IV					
Obs	savings	nondefault	default	woe	IV
1	A61	386	217	-0.27136	0.04665
2	A62	69	34	-0.13955	0.00206
3	A63	52	11	0.70605	0.02656
4	A64	42	6	1.09861	0.04394
5	A65	151	32	0.70425	0.07680
		700	300		0.19601

FIGURE A.2 – Information value : Checkingstatus1, History, Property, Purpose et Savings

Default	property				Total
	A121	A122	A123	A124	
0	222	161	230	87	700
1	60	71	102	67	300
Total	282	232	332	154	1000

Default	purpose								Total	
	A40	A41	A42	A43	A44	A45	A46	A48		A49
0	145	93	123	218	8	14	28	8	63	700
1	89	22	58	62	4	8	22	1	34	300
Total	234	115	181	280	12	22	50	9	97	1000

Default	housing			Total
	A151	A152	A153	
0	109	527	64	700
1	70	186	44	300
Total	179	713	108	1000

FIGURE A.3 – Tables de contingence entre défaut et Property, Purpose et Housings

Default	checkingstatus1				Total
	A11	A12	A13	A14	
0	139	164	49	348	700
1	135	105	14	46	300
Total	274	269	63	394	1000

Default	history					Total
	A30	A31	A32	A33	A34	
0	15	21	361	60	243	700
1	25	28	169	28	50	300
Total	40	49	530	88	293	1000

FIGURE A.4 – Tables de contingence entre défaut et Checkingstatus1 et History

```

proc logistic data=pfe.bd_train descending;
class property checkingstatus1 history Savings Purpose / param = glm ;
Model default = duration amount age property checkingstatus1 history
Savings Purpose / selection = stepwise lackfit ;
score data=pfe.bd_train out = pfe.Logit_Training fitstat outroc=troc;
score data=pfe.bd_test out = pfe.Logit_Validation fitstat outroc=vroc;
run;

```

FIGURE A.5 – Code utilisé pour la modélisation LOGIT

```

/*PIT*/
PROC LOGISTIC DATA=pfe.probabilities DESCENDING;
MODEL default_tmps = score rvp_tmps PIB tx_interet_tmps / LINK=PROBIT;
OUTPUT OUT=pfe.probabilities2 PREDICTED=PD_PIT_tmps;
RUN;

```

FIGURE A.6 – Code utilisé pour la modélisation PROBIT

```

/* covariable variant dans le temps*/
❑ data pfe.phreg;
  set pfe.mort;
  time1=time-first_time;
  time2=time-first_time+1;
  run;
ods graphics on;
❑ proc phreg data=pfe.phreg plots(overlay)=survival;
  model (time1,time2)*default_time(0)=fico_orig_time ltv_time gdp_time interest_rate_time/
  ties=efron;
  baseline;
  run;
ods graphics off;
/*probabilité de survie*/
❑ proc phreg data=pfe.phreg ;
  model (time1,time2)*default_time(0)=fico_orig_time ltv_time gdp_time interest_rate_time/
  ties=efron;
  baseline covariates=pfe.phreg out=pfe.survival
  survival=survival;
  run;
/*on passe au probabilité de défaut entre t1 et t2*/
/*PD(t1-t2)=(St1-St2)/St1*/
❑ Data pfe.survival1;
  Set pfe.survival;
  if time2>=1 then PD_time=(lag(survival)-survival)/lag(survival);
  if time2=1 then PD_time=1-survival;
  keep fico_orig_time LTV_time gdp_time interest_rate_time time2 PD_time;
  run;
/*on passe à la validation du modèle*/
/*on fusionne d'abord les deux tables de données obtenues*/
❑ proc sort data=pfe.phreg;
  by time2 fico_orig_time ltv_time gdp_time interest_rate_time;

```

FIGURE A.7 – Code 1 utilisé pour la modélisation par durée de vie

```

/*Modèle continue : analyse de survie "CPH"*/
/*on note que l'indicateur default_time est égale à 0 si l'observation est censurée et 1 si le défaut arrive*/
/*on ne se concentre que sur la censure à droite, en considérant que tous les crédits commencent par la première période d'observation*/
/*on classe par id afin de faciliter le constat de chaque clients sur si ou non il a fait défaut ou rembourser son crédit*/
❑ proc sort data=pfe.mort;
  By id;
  run;
/*on calcule les durées de vie selon id de chaque client*/
❑ Data pfe.client;
  set pfe.mort;
  duree = time-first_time+1;
  by id;
  retain id;
  if last.id then indicator=1;
  run;
/*on ne laisse que la dernière ligne pour chaque id, afin d'afficher la durée de chaque crédit*/
❑ data pfe.client1;
  set pfe.client;
  if indicator=1 or default_time=1;
  run;
❑ data pfe.clientfinal;
  set pfe.client1;
  by id;
  retain id;
  if first.id then output;
  run;
/* on modélise le modèle continue de survie grace à la fonction proc lifetest*/
ods graphics on;
/*le modèle sera construit grace au score crédit (nomé ici fico score)
et au ratio loan to value (LTV)et au taux d'intérêt*/
/*on modélise grace à proc phreg vu qu'on veut un modèle à risque proportionnel, puis on génère la fonction de survie*/
❑ proc phreg data=pfe.clientfinal plots(overlay)=survival;
  model duree*default_time(0)= fico_orig_time ltv_orig_time / ties=efron;
  run;
ods graphics off;
/* covariable variant dans le temps*/

```

FIGURE A.8 – Code 2 utilisé pour la modélisation par durée de vie

```

/*on passe à la validation du modèle*/
/*on fusionne d'abord les deux tables de données obtenues*/
proc sort data=pfe.phreg;
  by time2 fico_orig_time ltv_time gdp_time interest_rate_time;
run;

proc sort data=pfe.survival1 nodupkey;
  by time2 fico_orig_time ltv_time gdp_time interest_rate_time;
run;

data pfe.pdlt;
  merge pfe.phreg(IN=a) pfe.survival1;
  by time2 fico_orig_time ltv_time gdp_time interest_rate_time;
  if a;
run;

/*avant on analyse les données obtenue en calculant la moyenne*/
proc means data=pfe.pdlt;
  var pd_time default_time;
run;

/* on remarque que les deux variable ont presque la meme moyenne à savoir 0.024*/
/*pour la validation on test les hypothèses en commençant par la loglinéarité des variables*/

/*Pour la proportionnalité des risques*/
proc phreg data=pfe.phreg;
  model (time1,time2)*default_time(0)=fico_orig_time ltv_time gdp_time interest_rate_time/
  ties=efron;
  output out=pfe.resultat ressch=rfico rltv rgdp rinterest ;
run;

proc corr data=pfe.resultat;
  var time2 rfico rltv rgdp rinterest;
run;

```

FIGURE A.9 – Code 3 utilisé pour la modélisation par durée de vie

Bibliographie

- [1] *CREDIT RISK ANALYTICS (MEASUREMENT TECHNIQUES APPLICATIONS, and EXAMPLES in SAS)*
- [2] *ANALYSE DU RISQUE DE CREDIT (Banque & Marchés) de Cécile Kharoubi et Philippe Thomas*
- [3] *IFRS 9 and CECL Credit Risk Modelling and Validation de Tiziano Bellini*
- [4] *Cours de GLM et d'analyse de durée de vie de Monsieur CHAOUBI Abdelaziz, professeur à l'Institut National de Statistique et d'Economie Appliquée*
- [5] <https://www.iasplus.com/fr-ca/standards/part-i-ifs/broad-topics/ifrs9>
- [6] https://www.leblogdesfinanciers.fr/2018/01/12/classification_instruments_financiers_selon_ifrs9/
- [7] https://andreberchtold.com/UNIGE/survie/4_Modele_Cox.pdf
- [8] https://andreberchtold.com/UNIGE/survie/5_Reg_temps_discret.pdf