



RAPPORT DE PROJET DE FIN D'ÉTUDE

SUJET :

**MISE EN PLACE D'UN SCORE DE COMPORTEMENT POUR LE
CALCUL DES PROVISIONS SOUS LES NORMES IFRS.9**

Réalisé par :

Mohcine ACHNIT

**Option :
Actuariat-Finance**

Juin 2017

Encadrant interne : **M.Abdelaziz CHAOUBI**

Encadrant externe : **M.Adil OUAKIB**

Organisme d'accueil : **WAFASALF-CACF**





Résumé

Avec l'avènement des nouvelles normes IFRS.9 ces normes qui prendront effet début 2019 au Maroc, ont amené « Wafasalaf » à mettre en place un nouveau type de score en respect avec les réglementations dictées par Bâle III en matière de gestion de risque crédit.

Ce score de comportement a pour but de prédire la probabilité de défaillance future des clients de « Wafasalaf » en se basant sur leurs comportements dans le passé ainsi que sur leurs situations socio-économiques.

Cette probabilité de défaillance ainsi calculée va permettre d'estimer les provisions des clients susceptibles de tomber en défaut sur toute la période du crédit contrairement au score d'application qui lui permet de constituer des provisions sur les 12 mois futurs qui suit l'octroi du crédit comme dicté par les normes IFRS.9.

Les modèles de comportement sont des modèles spéciaux qui permettent de prédire le comportement futur des clients en se basant sur leurs historiques, ces modèles inclus entre autres les modèles de survie comme celui de «Cox» et « Kaplan Meier » qui permettent comme leurs noms indiquent 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, Perte attendue, Machine learning, Modèle logistique, Cox, Kaplan Meier, Modèle bayésien, Stress testing, VBA*

Dédicace

À mes très chers parents, pour leurs soutiens et tous leurs sacrifices

À mon frère pour son encouragement et son soutien

À mon cher oncle pour tous ses conseils et ses directives durant mon cursus scolaire

À toute ma famille

À tous mes amis

À ceux qui m'ont soutenu tout au long de ma vie Je dédie ce modeste travail

Mohcine

Remerciements

Il m'est agréable au terme de ce projet de fin d'études, d'exprimer ma reconnaissance et ma profonde gratitude à mon Professeur Abdelaziz CHAOUBI pour son aide, ses encouragements et ses conseils. Je lui suis très reconnaissant pour la qualité de son encadrement. Ainsi, je tiens à remercier Monsieur Abdelatif BELOUALJA, responsable du département Risque, de bien vouloir m'accorder ce stage de fin d'études au sein de son département. Je remercie également mon encadrant : Mr Adil OUAJIB, pour ses directives précieuses et ses conseils pertinents qui m'ont été d'un appui considérable dans mon projet, et un remerciement tout particulier à Mme Youssra Ourti pour ses précieux conseils et son suivi tout au long de ce stage, ainsi qu'à Mme Sara El Naciri et Mme Myriem AITBELLOUALI. Je tiens aussi à remercier l'ensemble du personnel du département pour leur coopération et l'accueil chaleureux qu'ils m'ont offert durant mon stage au sein de WAFASALAF.

Je ne saurais oublier dans mon remerciement tout le cadre professoral de l'INSEA, pour la formation qu'il m'a prodigué durant ces trois années de formation. Enfin, que tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce travail trouvent ici l'expression de ma reconnaissance et de mes remerciements.



Table des matières

Résumé	iii
Dédicace	iv
Remerciements	v
Table des matières	vii
Liste des figures	ix
Liste des tableaux	xiii
1 Présentation de l'organisme d'accueil	1
1.1 Présentation de Wafasalf	2
1.2 Historique	4
1.3 Organigramme	5
1.4 Généralités	6
1.5 Vision	7
2 Les accords de Bâle & Normes IFRS.9	9
2.1 Bâle I & Bâle II	10
2.2 Bâle III	13
2.3 Norme IFRS.9 (International Financial Reporting Standards)	15
3 Risque crédit	17
3.1 Les composants du risque crédit	18
3.2 Credit Scoring	18
4 Présentation du portefeuille et statistique descriptive	23
4.1 Présentaion de la base de données	24
4.2 Statistique descriptive	25
5 Modélisation de la probabilité de défaut de comportement (Score)	29
5.1 Méthodes utilisées en Machine Learning	30
5.2 Régression logistique	30
5.3 Application à notre base de données	32
5.4 Les arbres de décision	37

5.5	Réseau Bayésien	39
5.6	Application à notre base de données	42
5.7	Choix du meilleur modèle	45
6	Les modèles de durée de vie et modélisation du PD de comportement	47
6.1	Les modèles à temps discret	48
6.2	Les modèles à temps continu	53
6.3	La censure	53
6.4	Application à notre portefeuille	54
6.5	Conclusion	64
7	Le modèle Bayésien pour le calcul de la probabilité de défaut	65
7.1	Le modèle bayésien	66
7.2	Application à notre base de données	67
8	Stress Testing	73
8.1	Objectif du Stress Tests	74
8.2	Les types du Stress Testing	74
8.3	Le modèle risque sous la régulation Bâle	76
8.4	Stress Test de la probabilité de défaut basé sur les scénarios	79
9	Développement d'une application VBA	83
9.1	Motivation	84
9.2	Composantes de l'application VBA	84
9.3	Automatisation du calcul de la PD et du score de comportement	85
9.4	Calcul de la perte attendue	85
A	Annexes	I
A.1	Figures annexes	I
A.2	Tableaux annexes	V
A.3	Références	VII
	Bibliographie	VII

Liste des figures

1.1	Légende part de marché et part de capital	2
1.2	Part de marché Prêt personnel à fin 2015	3
1.3	Part de marché automobile à fin 2015	3
1.4	Historique de Wafasalaf	5
1.5	Organigramme de Wafasalaf	6
1.6	Les chiffres clés de Wafasalaf à fin 2015	7
2.1	Les principales caractéristiques de Bâle 1	11
2.2	Les trois pilier de Bâle II	12
2.3	Evolution des trois ratios durant la phase de transition	14
2.4	Evolution des trois ratios durant la phase de transition	14
2.5	Les 3 niveaux de comptabilisation de la provision risque crédit sous les nomes IFRS.9	15
3.1	Score de comportement : De Clip vidéo à un Aperçu	20
3.2	Score dynamique :Clip vidéo à clip vidéo	21
4.1	Boîte à moustache pour les variables continues	25
4.2	Test KHI2 : défaut*variables catégorielles	26
4.3	Tableau de corrélation : Défaut*variables continues	26
4.4	Matrice de corrélation des variables quantitatives	27
4.5	Analyse en composantes principales	28
4.6	Analyse des correspondances multiples	28
5.1	Allure de la fonction logit	31
5.2	Modèle logit : Récapitulatif sur la sélection séquentielle	32
5.3	Modèle logit : Analyse des effets Type 3	33
5.4	Modèle logit : Estimation par l'analyse du maximum de vraisemblance	33
5.5	Modèle logit : Estimation des rapports de cotes	34
5.6	Modèle logit : Association des probabilités prédites et des réponses observées	34
5.7	Modèle logit :courbe ROC	35
5.8	Partie des scorecards crédit du modèle logistique de notre portefeuille	36
5.9	Graphe CART	38

5.10	Courbe Roc pour le graphe CART	39
5.11	Évaluation de l'arbre de décision CART	39
5.12	Graphe du modèle TAN	42
5.13	Graphe du modèle TAN	43
5.14	Extrait du tableau des probabilités de défaut par segment	44
5.15	Comparaison du TAN et CM via leurs courbes ROC	44
5.16	Tableau indice Gini : comparaison TAN et CM	45
5.17	Courbes Roc : comparaison des modèles	45
5.18	Indice Gini : Évaluation des modèles	45
6.1	Fonction de lien non linéaire : Probit et Cloglog	49
6.2	modèle probit	51
6.3	Modèle probit	52
6.4	Modèle probit : TTC	52
6.5	Comparaison des probabilités de défaut moyenne pour les modèles : Logit, PPT et TTC	52
6.6	Les différentes possibilités d'aboutissement d'un crédit : Défaut ou censure	54
6.7	Table des observations après réorganisation des données panels	55
6.8	Tableau récapitulatif des données	56
6.9	Codage de la variable type produit	56
6.10	Méthode ascendante pas à pas	57
6.11	Estimation des paramètres du modèle de Cox	58
6.12	Courbe de survie	59
6.13	Nombre événement par produit	60
6.14	Estimation du modèle de Cox avec stratification par type de produit	61
6.15	Courbes de survie stratifiées selon le type de produit	61
6.16	Comparaison des courbes log(-logS)	62
6.17	Résidus partiels : Score d'octroi et taux de chômage	63
6.18	Courbe Roc du modèle de Cox	64
7.1	Résumé des paramètres MCMC pour un modèle probit	67
7.2	Procédure MCMC pour un modèle probit	68
7.3	Procédure MCMC pour le modèle probit	68
7.4	Procédure MCMC pour le modèle probit	68
7.5	Procédure MCMC pour le modèle probit	69
7.6	Procédure MCMC pour le modèle probit	69
7.7	Procédure MCMC pour le modèle probit	70
7.8	Diagnostic des paramètres β_0 et β_1	70
8.1	Les différents type du Stress testing	74
8.2	Les différents scénarios pour différents variables macroéconomiques	75

8.3	Perte prédite, non prédite et stressée et leurs capitaux correspondants	76
8.4	Le paramètre de corrélation en fonction de la probabilité de défaut pour les différents types de crédit	78
8.5	Le pire scénario du taux de défaut en fonction de la probabilité de défaut pour les différents types de crédit	78
8.6	Sortie du modèle Logistique	79
8.7	validation du modèle	79
8.8	Compraison des probabilités : PD base et PD wcdr	80
8.9	Fonction de distribution cumulative pour les probabilités de défaut :base, pire scénario, stressée	81
9.1	Interface de l'application VBA :calcul de la probabilité de défaut future	84
9.2	Interface de l'application VBA : Prise de décision	85
9.3	Calcul de la perte attendue d'un client donné	86
A.1	Les scorecards crédit du modèle logistique de notre portefeuille avec SAS Miner	I
A.2	Réseau bayésien :structure couverture de Markov	II
A.3	Résidus partiels : type produit	III
A.4	Procédure MCMC pour le modèle probit	III
A.5	Diagnostic des paramètres β_0 et β_1	III
A.6	Extrait du code VBA pour l'automatisation du calcul du score	IV
A.7	Tableau des probabilités conditionnelles de défaut :Réseau bayésien Markovien	V
A.8	Tableau des probabilités conditionnelles de défaut :Réseau bayésien Markovien	VI

Liste des tableaux

1.1	Taux pénétration crédit automobile & évolution entre 2008-2015	3
2.1	Bâle III :Les différents niveau de capitaux requis	13
4.1	Variables idiosyncratiques	24
4.2	Variables macroéconomiques	25

Chapitre 1

Présentation de l'organisme d'accueil

Sommaire

1.1 Présentation de Wafasalf	2
1.1.1 Mission	4
1.1.2 Réseau commercial	4
1.2 Historique	4
1.3 Organigramme	5
1.4 Généralités	6
1.4.1 Les chiffres clé	6
1.4.2 Les métiers chez Wafasalaf	7
1.5 Vision	7

1.1 Présentation de Wafasalf

Filiale du groupe Attijariwafa bank, principal acteur bancaire au Maroc, à rayonnement international à travers une présence au Maghreb, dans la sous région de l'Afrique de l'ouest et en Europe. Wafasalaf a pour deuxième actionnaire de référence CACF (Crédit Agricole Consumer Finance), filiale du groupe Crédit Agricole SA. spécialisée dans les activités de crédit à la consommation et leader Européen sur ce marché. (voir figure 1.1) Société spécialisée dans le crédit à la consommation, leader sur l'ensemble des compartiments du marché (prêt personnel, crédit automobile, équipement des ménages) , œuvrant par une politique d'innovation et de proximité s'appuyant fortement sur ses domaines d'expertises métiers, à faciliter l'accès au crédit à la consommation et apporter des solutions de financement adaptées aux besoins de ses clients et de ses partenaires. Dans un marché en reprise toujours caractérisé par une forte pression concurrentielle, Wafasalaf maintient sa position de leader sur l'ensemble des compartiments du marché avec une part de marché globale de 31,3%.(voir figure 1.1)

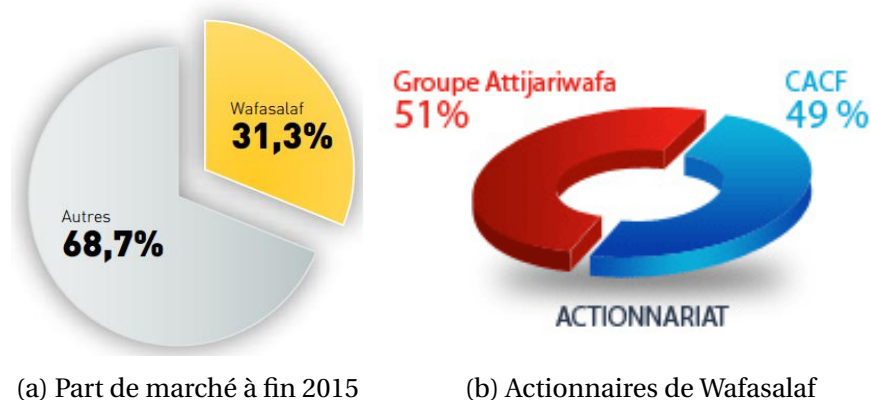


FIGURE 1.1 – Chiffres clés de Wafasalaf concernant le part de marché et le part de capital à fin 2015

Marché du prêt personnel

Dans un contexte d'accroissement de la concurrence des établissements bancaires et de relative stabilisation des encours en souffrance, le marché du Prêt Personnel enregistre une baisse de $-1,4\%$ par rapport à l'exercice 2014 pour atteindre un encours sain de 19.956 milliards de dirhams. Dans un tel environnement, Wafasalaf affiche une tendance meilleure que le marché avec une croissance de $+1,6\%$ qui lui confère une part de marché de 31,8%, soit +100 points de base par rapport à 2014.(voir figure 1.2)

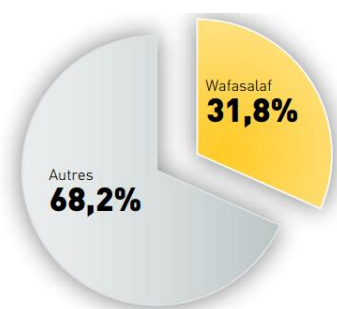


FIGURE 1.2 – Part de marché Prêt personnel à fin 2015

Marché automobile

Le marché Automobile a enregistré une année record en termes de ventes avec 131.935 unités écoulées, soit +8% par rapport à 2014. Cette croissance a été rendue possible, notamment grâce au marché du VP (Véhicule Particulier), en progression de +10.25%. Dans ce contexte, nous avons assisté à une plus forte participation des financements à crédit : le taux de pénétration du crédit dans les ventes Automobiles a atteint 44% en 2015, contre 41% une année auparavant. (voir tableau 1.1 et figure 1.3)

	2009	2010	2011	2012	2013	2014	2015
Ventes Auto	109966	103 436	112 099	130 316	120 766	122 081	131 935
Evolution	-	-55.9%	+8.4%	+16.3%	-7.3%	+ 1.1%	+8%
Taux pénétration crédit	40%	35%	38.9%	42.9%	40%	41%	44%

TABLEAU 1.1 – Taux pénétration crédit auto & Évolution

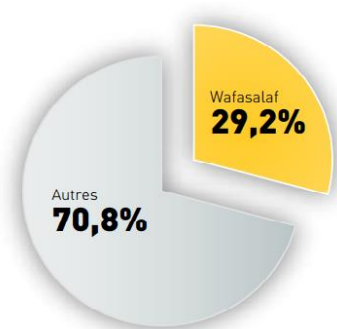


FIGURE 1.3 – Part de marché automobile à fin 2015

Marché équipement des ménages

Le marché des biens d'équipement des ménages a enregistré en 2015 une progression de 18% par rapport à 2014. Selon des cabinets spécialisés, la téléphonie (Smartphones) et l'électroménager blanc (réfrigérateurs, machines à laver, ...) ont été les principaux contributeurs à cette croissance. Grâce à ses partenariats avec les principales enseignes du

Royaume, Wafasalaf maintient une position de choix dans ce marché (part de marché de 99,1%) et confirme année après année, son ambition d'améliorer l'accès aux biens d'équipement aux ménages marocains, pour un mieux vivre général.

1.1.1 Mission

Acteur spécialisé en crédit à la consommation, contribuant au soutien de l'activité économique, en favorisant le développement de la consommation des ménages, par la mise à disposition à la clientèle la plus élargie et de manière responsable des solutions de financement adaptées et innovantes, anticipant l'évolution des habitudes de consommation.

1.1.2 Réseau commercial

Le réseau Wafasalaf s'organise autour de :

42 agences commerciales dont 15 dédiées au financement automobile et d'équipement de la maison. Plus de 2 000 points de ventes agréés.

1.2 Historique

À la fin des années 80; l'octroi des prêts personnels était essentiellement pratiqué par les organismes bancaires qui en faisaient une distribution limitée. En 1987 Wafabank créa un filiale spécialisée en crédit à la consommation : **Wafasalaf**

Depuis son existence Wafasalaf ne fait que innover que ce soit par l'offre de nouveaux produits ou développer ceux déjà existants (voir figure 1.4).

Celle-ci a eu dans un premier temps pour politique d'élargir l'accès aux solutions de financement adaptées aux ménages de différentes catégories socioprofessionnelles.

Wafasalaf s'est dès le départ distinguées comme un opérateur multi-produits, offrant des solutions de crédit pour répondre aux besoins de ses clients, à la fois en prêt personnel, financement automobile et équipement des ménages.



FIGURE 1.4 – Historique de Wafasalaf

1.3 Organigramme

Le rôle de faiseur de marché exige une écoute constante aux clients et aux partenaires afin d'améliorer régulièrement les outils, les offres et les processus de fonctionnement pour s'assurer de répondre au mieux à aux attentes de ces derniers.

Wafasalaf est actuellement organisée en lignes de métier dédiées : une activité commerciale propre, une activité de Gestion Pour Compte d'autrui, une activité Risque ainsi que des fonctions supports.(voir figure 1.5)



FIGURE 1.5

1.4 Généralités

1.4.1 Les chiffres clé

- Bonne performance commerciale (+7,3%) dans un contexte marché très concurrentiel, marquée essentiellement par la dynamisation de l'activité automobile.
- Renforcement des encours et confirmation de la position de leader (+4,4%) grâce au développement commercial, au dynamisme continu des activités de Gestion pour compte et une proximité accrue avec nos clients et partenaires.
- Dans un contexte de marché tendu, un léger retrait du Produit Net Bancaire (-1,5%). Baisse des dotations nettes (-12,8%) se traduisant par une performance au niveau du coût du risque de 31 points de base, réalisée grâce à une politique saine d'octroi de crédit et un processus de recouvrement efficient.

(voir figure 1.6)



FIGURE 1.6 – Les chiffres clés de Wafasalaf à fin 2015

1.4.2 Les métiers chez Wafasalaf

Les métiers de Wafasalaf s'organisent en 4 grands domaines :

Le commercial, le risque, les métiers de la finance d'entreprise et les métiers transversaux, avec une fonction de support et de conseil (systèmes d'information, informatique, audit, contrôle de gestion, Ressources Humaines ...).

Wafasalaf compte plusieurs métiers répartis entre ses différents domaines d'activité. Les rubriques suivantes mettent l'accent sur quelques-uns des métiers les plus représentatifs de l'entreprise.

1.5 Vision

Engagée dans une action économique globale destinée à améliorer le niveau de vie des ménages marocains, Wafasalaf n'en oublie pas pour autant sa responsabilité sociale face aux défis humains que le Maroc doit relever. Son engagement citoyen au profit des personnes en situation précaire ou de jeunes étudiants se développe chaque année.

Chapitre 2

Les accords de Bâle & Normes IFRS.9

Sommaire

2.1 Bâle I & Bâle II	10
2.1.1 Bâle I	10
2.1.2 Bâle II	11
2.1.3 Limites de Bâle I & II	12
2.2 Bâle III	13
2.2.1 Définition	13
2.2.2 La phase de transition	14
2.2.3 Les approches de Bâle et modélisation du risque crédit	14
2.3 Norme IFRS.9 (International Financial Reporting Standards)	15

Introduction

Dans cette partie, nous examinons de plus près les Accords de capitaux Bâle I, Bâle II et Bâle III. Il s'agit de lignes directrices réglementaires qui ont été introduites afin que les institutions financières puissent déterminer de manière appropriée leurs provisions et leurs amortissements de capital pour se protéger contre diverses expositions au risque. Un risque important est le risque de crédit, et dans cette section, nous discutons l'impact de ces accords sur le développement des modèles de risque de crédit en particulier la probabilité de défaut **PD**. Les règlements de Bâle sous-tendent beaucoup d'aspects de l'analyse du risque de crédit, et nous reviendrons sur les différentes questions dans les chapitres suivants.

2.1 Bâle I & Bâle II

2.1.1 Bâle I

Les Accords de Bâle ont été proposés par le Comité de Bâle par la Banque de Supervision. Ce comité a été fondé en 1974 par les banques centrales G10. De nos jours, il compte 27 membres. Ils se rencontrent régulièrement à la Banque des règlements internationaux (BIS) à Bâle, en Suisse. Le premier accord introduit était l'Accord Bâle I Capital, en 1988. Comme déjà mentionné, l'objectif était de mettre en place des exigences réglementaires de capital minimum afin pour s'assurer que les banques sont capables pour retourner les fonds des déposants. L'accord Bâle I a axé principalement sur le risque de crédit et introduit l'idée du ratio Cooke, et le ratio pondéré du risque. Il a placé une limite inférieure à ce ratio de 8%; En d'autres termes, la capitale devrait être supérieure à 8% des actifs pondérés en fonction du risque. Ainsi la formule introduit par *Cooke* qui s'écrit comme suit :

$$RatioCooke = \frac{FondsPropres}{RisqueCrédit + RisqueMarché} \geq (8\%)$$

Cette formule est considérée comme une relation phare en ce qui concerne la réglementation des fonds propres.

Bâle I	
Ratio de fonds propres	Ratio Cooke = $\frac{\text{Fonds propres}}{\text{risque de crédit} + \text{risque de marché}} \geq 8\%$
Méthode de calcul des risques	Méthode de calcul uniforme.
Le risque de crédit	<p>Méthode standard</p> <ul style="list-style-type: none"> - Catégories d'emprunteurs : État OCDE, banque, hypothécaire et « normal » (entreprises, particuliers, États hors OCDE). - Pondérations respectives : 0%, 20%, 50% et 100%.
Le risque de marché	Risque de marché mesuré par une approche standard ou une approche de modèle interne
Le risque opérationnel	Pas de prise en compte du risque opérationnel.

FIGURE 2.1 – Les principales caractéristiques de Bâle I

2.1.2 Bâle II

Pour remédier aux lacunes de l'Accord de Bâle I, l'Accord de Bâle II a été introduit. Il se compose de trois principaux piliers :

Le pilier I : couvre le capital minimal Exigence.

Le pilier II : couvre le processus d'examen de la surveillance.

Le pilier III : discipline de marché et la divulgation.

Sous le pilier I, trois types de risques différents sont inclus. Le risque de crédit est le risque rencontré en prêtant de l'argent aux obligataires. Le risque opérationnel est défini comme le risque de perte indirecte résultant de processus internes, de personnes et de systèmes inadéquats ou échoués ou des événements externes. Des exemples populaires ici sont la fraude, les dégâts matériels et les pannes du système. Le risque de marché est le risque dû aux mouvements de marché défavorables face au positionnement d'une banque sur le marché par l'intermédiaire d'espèces ou de produits dérivés. Exemples populaires sont

le risque de change, le risque et le risque de taux d'intérêt. (voir figure 2.2)

Finance et gestion des risques

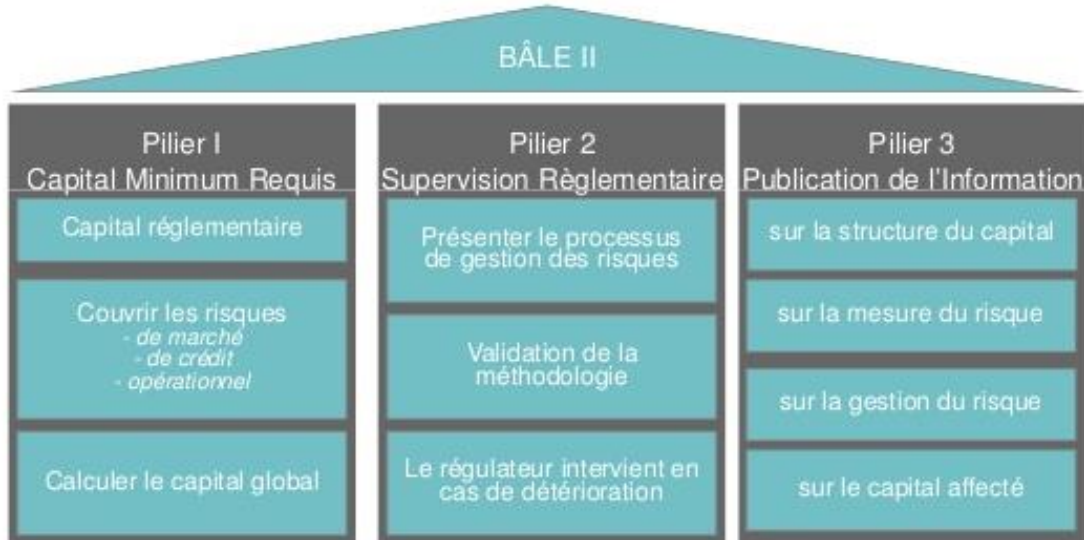


FIGURE 2.2 – Les trois pilier de Bâle II

2.1.3 Limites de Bâle I & II

Concernant Bâle I, seul le risque crédit était pris en compte. Or l'affaire Kerviel nous montre clairement que les risques opérationnels et de marché sont eux aussi très importants. De plus, le fait de limiter la capacité de prêt entraîne les banques à devoir optimiser la rentabilité de leur portefeuille. Elles vont, pour ce faire, sortir de leur portefeuille de créances, les prêts ayant un faible taux de rentabilité ainsi qu'un faible risque, de façon à ne garder en portefeuille que les créances au plus fort taux de rentabilité. Ainsi elles optimisent la rentabilité de leur capacité de prêt. Enfin, la pondération des crédits ne se faisait qu'en fonction du type d'acteur ou du type de garantie couvrant le prêt. Cette approche n'était pas assez personnalisée pour réellement couvrir et surtout estimer les risques liés aux prêts. En effet, la notion de qualité de la signature de l'emprunteur n'était pas prise en compte (ou tout du moins de manière très simpliste) et ce n'était pas les banques qui décidaient de la pondération de leurs lignes de crédit alors qu'elles sont les plus à mêmes de qualifier la qualité de la créance et le risque qui s'y attache.

Quant à Bâle II et malgré son jeune âge, le ratio Bâle II présente déjà des signes de faiblesse. L'actualité nous le montre, la gestion des risques bancaire est encore loin d'être résolue. Appliqué dans son intégralité depuis le 1 janvier 2008, le ratio Bâle II présente de

nombreuses faiblesses telles que :

- L'application uniquement à l'échelle européenne
- Le principe de procyclicité
- L'impact des normes I.F.R.S. sur ce ratio
- Les différents problèmes d'évaluation des risques
- L'insuffisance du ratio Mc Donough pour la couverture des risques bancaires.

2.2 Bâle III

2.2.1 Définition

L'Accord sur les capitaux de Bâle III a été introduit en conséquence directe de la crise financière de 2007. Il s'appuie sur l'Accord de Bâle II, vise à renforcer les normes mondiales en matière de capital. Son principal point d'attention est l'accent mis davantage sur le capital corporel, car il s'agit de la composante ayant la plus grande capacité de perte d'absorption. Cela réduit la dépendance vis-à-vis des modèles développés en interne par la banque et les notes obtenues auprès des agences de notation externes. Cela met également davantage l'accent sur les tests de stress (Stress testing). Pour les grandes banques, il souligne la nécessité d'avoir une capacité d'amortissements des pertes au-delà des normes communes. Il met davantage l'accent sur le capital de niveau I pondéré (l'encours pondéré = RWA¹). Une nouveauté clé est qu'elle introduit un ratio de levier à risque et sensitif en tant que backstop pour aborder le risque de modèle. (voir tableau 4.2)

	Bâle II	Bâle III
CET1	2%* RWA	4.5%* RWA
Ratio Tier I	4%* RWA	6%* RWA
Ratio Tier II	4%* RWA	2%* RWA
volant de conservation	-	2.5%* RWA
fonds propres contracycliques	-	0%-2.5%* RWA

TABLEAU 2.1 – Bâle III : Les différents niveaux de capitaux requis

1. Note : RWA=Risk-Weighted-Asset

2.2.2 La phase de transition

Deux mesures du Comité de Bâle ont quelque peu atténué, durant l'été 2010, les impacts Bâle III : l'une est pérenne, le dispositif de la Franchise ; l'autre est destinée à ménager une (longue) période transitoire. La mise en application à l'échelle nationale, par les pays membres, débutera le 1er janvier 2013. Ils devront avoir transposé ces règles dans leurs législations et réglementations respectives avant cette date. Le Comité de Bâle a introduit une phase de transition importante dans cette mise en œuvre des déductions. Les ajustements réglementaires (déductions et filtres prudentiels), et notamment les montants au-delà de la Franchise pour les participations dans d'autres établissements financiers, les MSR et les DTA, ne seront totalement déduits qu'à compter du 1er janvier 2018 du CET1.

Plus concrètement, ces ajustements réglementaires seront déduits à hauteur de 20% le 1er janvier 2014, puis passeront à 40% le 1er janvier 2015, à 60% le 1er janvier 2016, à 80% le 1er janvier 2017, pour atteindre ces 100% le 1er janvier 2018 (voir figure 2.4).

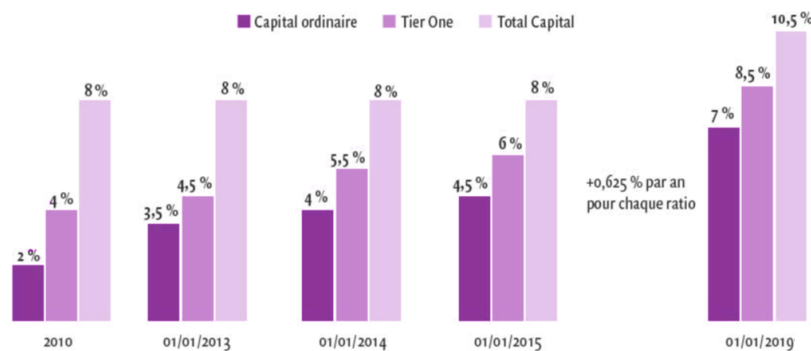


FIGURE 2.3

FIGURE 2.4 – Evolution des trois ratios durant la phase de transition

Durant cette période de transition, la partie non déduite de cette composante continuera à être soumise aux traitements nationaux en vigueur. Les déductions avec le système de Franchise seront ainsi progressives entre 2013 et 2018 (pourcentage en hausse de 20% tous les ans). Une année donnée, si 40% de la déduction et de la Franchise Bâle III sont appliqués, alors, sur ces mêmes opérations, les banques devront prendre 60% de la charge Bâle II.

2.2.3 Les approches de Bâle et modélisation du risque crédit

L'approche standard

L'approche standard actuelle prévoit deux options pour l'application des pondérations aux banques. L'option 1 lie la pondération appliquée à une banque à la note souve-

raîne de l'État où elle est établie, tandis que l'option 2 applique la pondération qui correspond à la note de crédit la banque. En révisant le traitement des expositions sur les banques, le Comité entend supprimer les références aux notations externes et la relation avec le risque souverain. Il cherche à mettre au point une approche unique fondée sur deux déterminants du risque pertinents pour les banques.

L'approche Internal Ratings Based (IRB)

L'approche IRB est une approche plus sophistiquée pour quantifier le risque crédit. Elle se base sur 3 paramètres clés, probabilité de défaut **PD**, loss given default **LGD** et exposure at default **EAD** introduite dans le chapitre suivant d'une manière approfondie.

2.3 Norme IFRS.9 (International Financial Reporting Standards)

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. (voir figure 2.5)

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

FIGURE 2.5 – 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.

Chapitre 3

Risque crédit

Sommaire

3.1 Les composants du risque crédit	18
3.2 Credit Scoring	18
3.2.1 Définition	18
3.2.2 Les types de score crédit	19
3.2.3 Scores de comportement/dynamique et Normes IFRS.9	21

Introduction

3.1 Les composants du risque crédit

Probabilité de défaut "*PD*"

La probabilité de défaut est, dans le cadre de Bâle II, la probabilité qu'un débiteur ne puisse faire face à ses obligations de remboursement.

La probabilité de défaut est un paramètre utilisé dans la détermination du capital économique, ECap et du capital réglementaire, RegCap sous Bâle II pour les institutions bancaires.

Loss Given Default "*LGD*"

Loss Given Default (LGD) est un des trois indicateurs de risque de crédit de la réglementation Bâle II correspondant à l'évaluation la perte encourue en cas de défaut de la part d'une contrepartie. Le Loss Given défaut est exprimé en pourcentage.

Exemple : $LGD = 100\%$: perte du total du montant en cas de défaut d'une contrepartie.

Exposure At Default "*EAD*"

L'« exposition en cas de défaut » est l'exposition encourue par un créancier en cas de défaut de son débiteur. En d'autres termes, c'est le montant du prêteur exposé au risque de défaut de l'emprunteur. qui est le risque que le débiteur ne remplisse pas son obligation de remboursement au créancier.

Exposure At Default (EAD) - est un paramètre utilisé dans la détermination du capital économique, ECap et du capital réglementaire, RegCap sous Bâle II pour les institutions bancaires.

3.2 Credit Scoring

3.2.1 Définition

Le **credit scoring** est un terme anglo-saxon qui correspond à l'évaluation du risque de défaillance-crédit d'un particulier ou d'une entreprise à partir d'un score statistique.

Le principe de base est d'anticiper la solvabilité de l'emprunteur grâce à une série de ratio recoupés dans une « **fonction score** ».

Pour noter un individu, plusieurs paramètres sont pris en compte. Les plus importants

sont :

1. L'historique de crédit : paradoxalement, il est préférable d'avoir contractés plusieurs emprunts (auto, etc.) pour améliorer son credit score en démontrant sa capacité de remboursement.
2. La régularité des paiements effectués (loyer, factures, remboursements de cartes de crédit...).
3. Le type de crédit demandé.
4. Le nombre de recherches effectuées pour obtenir un prêt (credit inquiry) NINQ.

3.2.2 Les types de score crédit

Les trois approches statistiques clés dans le score crédit pour des portefeuilles de détail sont :

1. **Le score d'application ou d'octroi** : c'est la 1ère approche statistique la plus importante dans le score crédit pour les portefeuilles de détail, son objectif est d'attribuer un score au client lors de l'octroi du crédit, ce score doit refléter le risque de défaut de ce client. L'importance de ce score réside donc dans le fait qu'il permet à l'organisme prêteur la possibilité d'accepter ou de refuser les clients lors de l'octroi de crédit en se basant sur ce score calculé.
2. **Le score de comportement (cf. Chapitre 5 p. 29)** est une autre approche statistique du score de crédit de détail, dans ce cas on étudie le comportement des clients auxquels on a déjà accordés un crédit. *Pour bien comprendre le score de comportement, considérons l'exemple suivant :*
 - Imaginons un client X qui demande un crédit de sa banque ou d'un organisme de crédit de détail. Premièrement, à travers le calcul du score d'octroi la banque va décider si elle doit accorder le crédit à X, supposons que l'octroi a eu lieu (Client accepté, Bon client¹?). À un moment donné dans le temps la banque veut réévaluer le risque crédit de son client, en prenant en compte cette fois-ci et en plus des variables considérées lors de l'octroi du crédit, son comportement récent (12 mois typiquement si la durée de crédit dépasse 2 ans sinon le nombre de mois dès l'octroi du crédit jusqu'au moment d'observation). Ces informations récentes peuvent être le suivi (comportement) du compte bancaire du client résumé par la moyenne de ou bien sa tendance durant les 12 derniers mois. Autres informations intéressantes c'est la défaillance du client si le client n'a pas honoré ses engagements lors des crédits passés ou existants, ou encore le retard des paiements des traites. (voir figure 3.1)

1. La banque peut accorder à un client un crédit mais qui va par la suite s'avérer que c'est un mauvais client (Défaillant).

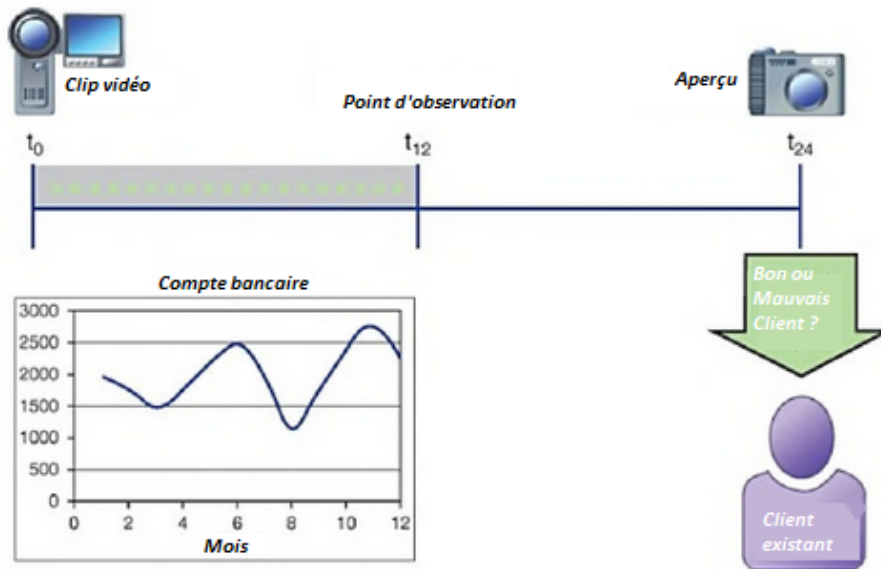


FIGURE 3.1 – Score de comportement : De Clip vidéo à un Aperçu

Comparé au score d’octroi, le score de comportement est un score dynamique dans le sens où il est changeant dans le temps en fonction des variables de comportement et ou toute donnée qui peut enrichir ce score. Tous ces variables de comportements combinées donne lieu au *score de comportement*.

3. **Score dynamique**(cf.Chapitre 6 p. 47) On a déjà abordé le score d’octroi comme problème d’un aperçu à un autre aperçu dans le temps, le score de comportement quant à lui est un problème d’un vidéo clip à un aperçu, dans ce qui suit on va discuter le score dynamique comme approche statistique du score crédit cette fois on parle d’un vidéo clip à un autre vidéo clip.

Pour le score dynamique, le risque crédit est évalué à n’importe quel moment futur dans le temps, contrairement au score d’octroi et au celui de comportement où le risque est évalué à un horizon de temps fixé (18 et 12 mois resp), dans le cas du score dynamique le risque est évalué sur 3, 4, 5, 6, 12, ou 18 mois (ou plus) dans le futur (voir figure 3.2). En tant que tel, il donne plus d’information que les scores d’octroi et de comportement. Cependant, il est plus difficile de construire ce genre de score, il nécessite des techniques statistiques plus poussées comme *l’analyse de survie*. Principalement, ces techniques nous permettent non seulement de savoir si un client va faire défaut, mais aussi quand. C’est cette dernière propriété qui donne à cette technique sa particularité. D’une manière simpliste on peut considérer que :

$$\text{Score dynamique} = \text{Score de comportement}(T)^2$$

2. le score dynamique n’est que le score de comportement dans le temps en faisant varier T (Score dynamique(12)=Score de comportement classique)

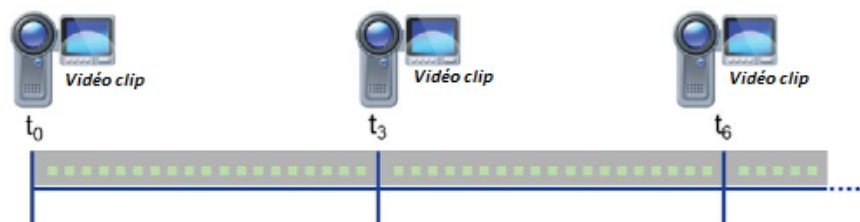


FIGURE 3.2 – Score dynamique :Clip vidéo à clip vidéo

3.2.3 Scores de comportement/dynamique et Normes IFRS.9

Les score de comportement/dynamique sont d'une importance primordiale dans l'application des nouvelles normes IFRS.9 d'une part car ils vont permettre de prédire le comportement des clients existants quant au paiements de leurs traites futures sur la base des paiements passés, d'autre part d'anticiper toute défaillance future en constituant des provisions.

Comme mentionné auparavant, la spécificité de ces scores c'est qu'ils permettent de prédire la probabilité de défaillance sur toute la durée du crédit. Dans ce cas on peut distinguer deux niveaux :

1. Si la probabilité de défaillance ne dépasse pas un certain seuil, généralement 10% du portefeuille en question, l'organisme de crédit n'est pas obligé de constituer des provisions sur toute la durée de crédit.
2. Si par contre la probabilité de défaillance est grande i.e. dépasse un certain seuil 10%, l'organisme de crédit est dans l'obligation de constituer des provisions sur toute la durée de crédit.

Chapitre 4

Présentation du portefeuille et statistique descriptive

Sommaire

4.1	Présentaion de la base de données	24
4.1.1	Les variables propres au client	24
4.1.2	Les variables de comportement	24
4.1.3	Les variables macroéconomiques	24
4.2	Statistique descriptive	25
4.2.1	Introduction	25
4.2.2	Traitement des valeurs extrêmes	25
4.2.3	Test d'indépendance : KHI2	25
4.2.4	Mesure de corrélation	26
4.2.5	Mesure des corrélations entre les différentes variables	27
4.2.6	Analyse des données	27

Introduction

Dans cette partie de notre projet on va commencer à manipuler notre base de données, premièrement on va exposer les types de variables que notre base comporte, ensuite on procédera à une analyse du jeu de données à savoir une analyse univarié, bivariée et enfin multivariée .

4.1 Présentaion de la base de données

4.1.1 Les variables propres au client

Le premier type de variables que notre base comporte sont les variables à trait personnel c-à-d qu'elles sont propres au client, qu'on appelle aussi des variables idiosyncratiques, le tableau suivant présente quelques-unes :

Intitulé variable	Libellé	Type
ID	Identifiant du client	Entier
Age_cli	âge du client	Entier
Compt_temps	Compte bancaire du client au moment d'observation	Réel
NINQ	Le nombre de recherches effectuées pour obtenir un prêt	Entier

TABLEAU 4.1 – Variables idiosyncratiques

4.1.2 Les variables de comportement

Les variables de comportement sont la nouveauté du score de comportement puisqu'elles intègrent la dimension comportementale du client et son assiduité vis à vis le paiement des traites dans le passé, on trouve entres autre les variables suivantes :

1. `nombr_impay` : Le nombre impayé de traites par le client sur la période d'observation
2. `MT_Cred_ex` : le montant des crédits existants
3. `sold_compt` : le solde du compte bancaire du client au moment d'observation

4.1.3 Les variables macroéconomiques

Ce sont des variables qui définissent l'état de l'économie, elles vont être utilisées lors de notre étude du *stresstesting* ainsi que dans notre modélisation de la probabilité de défaut **PD** par les modèles de durée de vie.

Intitulé variable	Libellé	Type
PIB	Taux de croissance du produit intérieur brut	pourcentage
Taux_Chom	Taux de chômage	pourcentage
Taux_Infla	Taux d'inflation	pourcentage
Taux_Interet	Taux d'intérêt du marché	pourcentage

TABLEAU 4.2 – Variables macroéconomiques

4.2 Statistique descriptive

4.2.1 Introduction

Dans partie de ce chapitre on va faire une analyse descriptive de notre portefeuille pour appréhender la relation entre les variables et voir s'il existe a priori une relation entre notre variable à expliquer à savoir la probabilité de défaut **PD** des clients existants et les différentes variables explicatives.

4.2.2 Traitement des valeurs extrêmes

L'étude des valeurs extrêmes est extrêmement importante, cela nous permet de comprendre si des observations correspondent à un profil particulier ou une catégorie particulière des individus et qui nécessite un traitement à part notamment dans notre cas à savoir la prédiction du risque d'impayé.

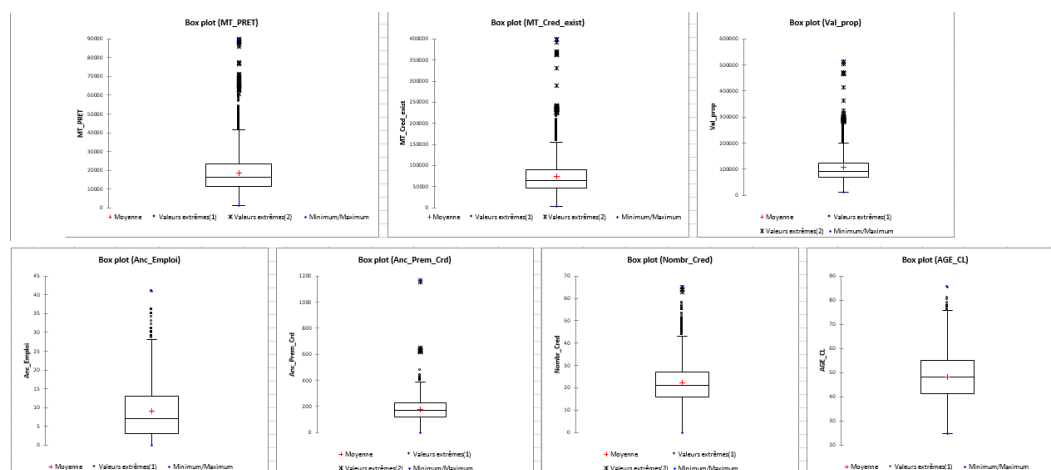


FIGURE 4.1 – Boîte à moustache pour les variables continues

4.2.3 Test d'indépendance : KHI2

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

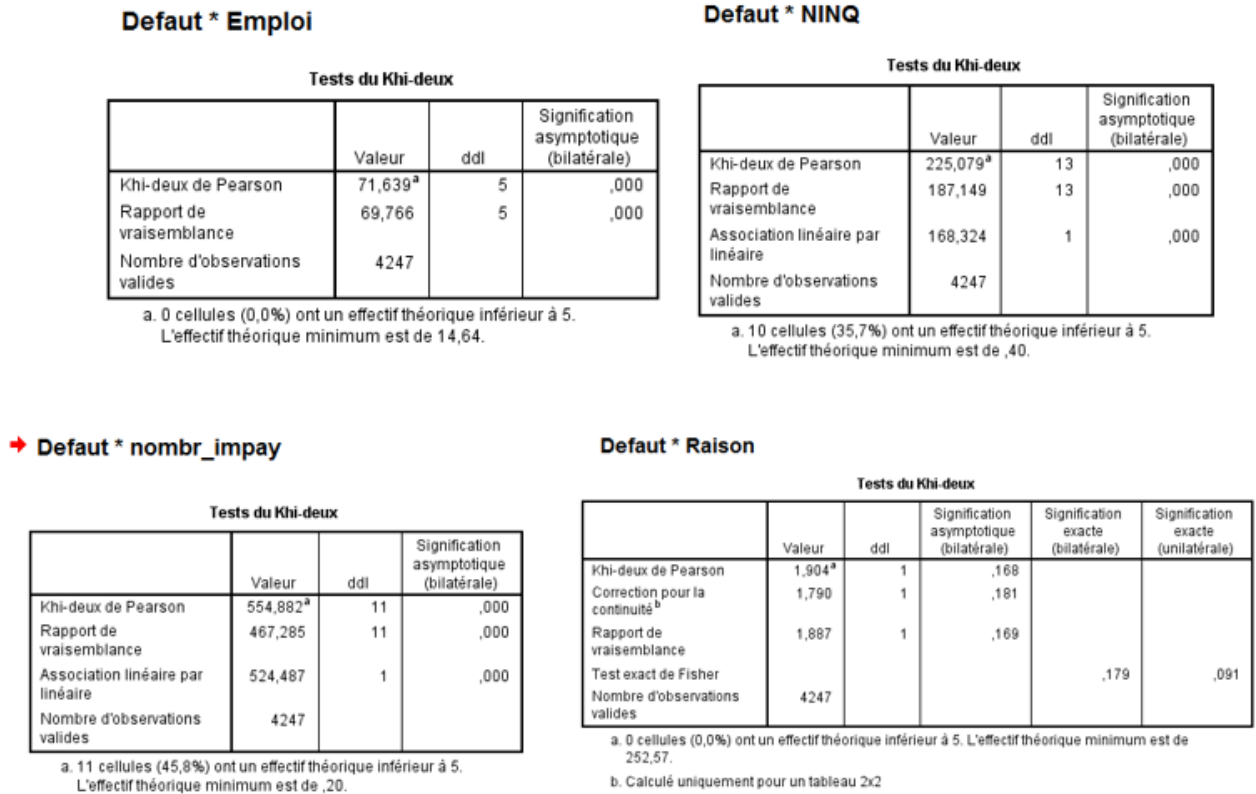


FIGURE 4.2 – Test KHI2 : défaut*variables catégorielles

variable **défaut** dépend des variables catégorielles **Emploi, NINQ et nombre d'impayé**, par contre il ne dépend pas de la variable **Raison**.

4.2.4 Mesure de corrélation

De même que dans la partie précédente, on va étudier la corrélation qui existe entre la variable **défaut** et les variables continues c'est un peu l'équivalent du test khi2 dans le cas des variables continues. On remarque que la variable **défaut** est significativement cor-

Corrélations

	Defaut	MT_PRET	Val_prop	Anc_Emploi	nombr_impay	AGE_CL	Nombr_Cred	Anc_Prem_Cr d	MT_Cred_exi st	NINQ
Defaut	1									
Corrélation de Pearson		-.096**	-.069*	-.077*	,351**	,002	,000	-.164**	-.062**	,199**
Sig. (unilatérale)		,000	,000	,000	,000	,441	,499	,000	,000	,000
N	4247	4247	4247	4247	4247	4247	4247	4247	4247	4247

** La corrélation est significative au niveau 0.01 (unilatéral).

* La corrélation est significative au niveau 0.05 (unilatéral).

FIGURE 4.3 – Tableau de corrélation : Défaut*variables continues

rélee avec **MT_pret, val_prop, ANC_emploi, nombr_impay, Anc_Prem_Cr, MT_Cred_ex et NINQ**, par contre le défaut n'est pas significativement corrélé à l'âge du client et au nombre de crédit.

4.2.5 Mesure des corrélations entre les différentes variables

Parmi les hypothèses qu'on doit vérifier c'est la non corrélation des variables de notre base de données, sinon cela va baisser notre modèle.

La figure suivante liste les différentes variables continues et leurs corrélations.

Comme première remarque on voit que l'ancienneté d'emploi est corrélée avec l'âge donc on va se contenter d'une seule variable, ainsi que les variables montant des crédits existants avec la valeur de la propriété du client ce qui est aussi explicable, en effet les clients qui résident dans des propriétés luxueuses tendent à contracter plusieurs crédits donc un montant plus conséquent et vice versa de même on va choisir une seule des deux variables.

À part ces deux cas, on peut affirmer que nos variables ne sont pas significativement corrélées.

	MT_PRET	MT_Cred_exist	Val_prop	Anc_Emploi	Anc_Prem_Crd	Nombr_Cred	AGE_CL	nombr_impay	NINQ
MT_PRET	1	0,3030	0,4088	0,0780	0,0889	0,1005	-0,0045	-0,0537	0,0398
MT_Cred_exist	0,3030	1	0,8744	-0,0870	0,1442	0,3142	-0,0256	-0,0453	0,0389
Val_prop	0,4088	0,8744	1	-0,0109	0,2034	0,2255	-0,0082	-0,0540	0,0007
Anc_Emploi	0,0780	-0,0870	-0,0109	1	0,2206	0,0147	0,8680	0,0350	-0,0504
Anc_Prem_Crd	0,0889	0,1442	0,2034	0,2206	1	0,2116	-0,0030	0,0052	-0,1075
Nombr_Cred	0,1005	0,3142	0,2255	0,0147	0,2116	1	0,0198	0,1307	0,0798
AGE_CL	-0,0045	-0,0256	-0,0082	0,8680	-0,0030	0,0198	1	0,0028	0,0165
nombr_impay	-0,0537	-0,0453	-0,0540	0,0350	0,0052	0,1307	0,0028	1	0,0509
NINQ	0,0398	0,0389	0,0007	-0,0504	-0,1075	0,0798	0,0165	0,0509	1

FIGURE 4.4 – Matrice de corrélation des variables quantitatives

4.2.6 Analyse des données

Pour comprendre d'avantage le comportement de notre base de données, on a effectué deux analyses une sur les variables quantitatives et l'autre sur les variables nominales à savoir respectivement l'ACP et l'ACM.

Analyse en composantes principales

La sortie de l'ACP figure 4.5 montre que nos variables peuvent être regrouper en deux facteur, F1 et F2 dont la contribution dans la variabilité est de 26% et 14% respectivement, on peut voir la première figure à gauche que le nombre de crédit augmente avec l'âge et l'ancienneté d'emploi, aussi que le montant du prêt est positivement corrélé avec montant crédit et valeur de la propriété du client.

La figure de gauche indique la répartition des observations en fonction de défaut des clients selon les deux axes.

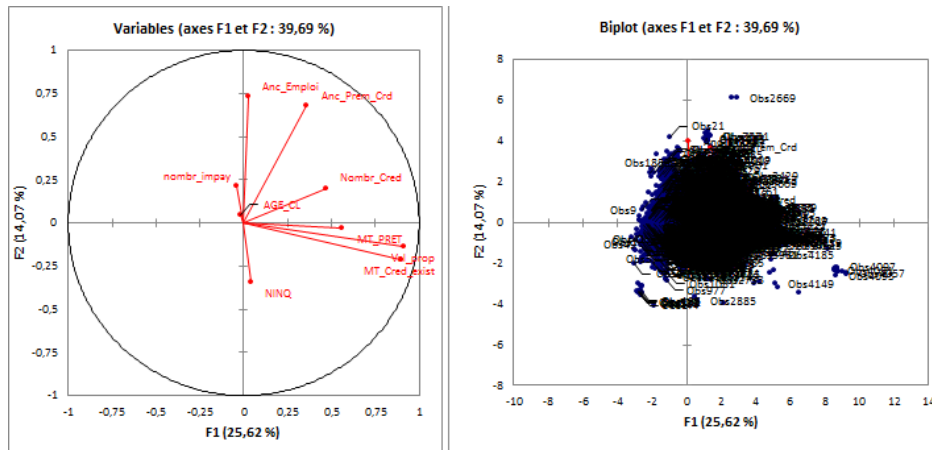


FIGURE 4.5 – Analyse en composantes principales

Analyse des correspondances multiples

Quant à l'analyse effectuée par l'ACM sur les variables qualitatives figure 4.6 de dessous a réduit les variables à deux facteurs aussi F1 et F2 avec des contributions respectives de 73% et 2%, on remarque que les hommes de notre base de données ont tendance à avoir la modalité **autre** emploi par contre les femmes ont pour emploi **manager**.

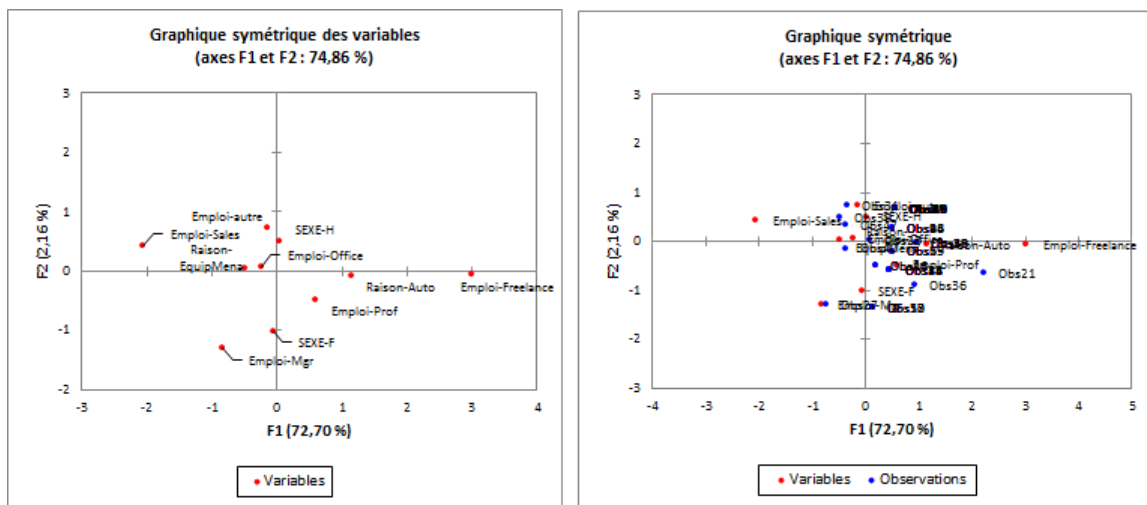


FIGURE 4.6 – Analyse des correspondances multiples

Chapitre 5

Modélisation de la probabilité de défaut de comportement (Score)

Sommaire

5.1 Méthodes utilisées en Machine Learning	30
5.1.1 Définition	30
5.2 Régression logistique	30
5.2.1 Généralités	30
5.2.2 Le modèle logistique dans la modélisation du score crédit	32
5.3 Application à notre base de données	32
5.3.1 Calibration des probabilités de défaut	35
5.4 Les arbres de décision	37
5.4.1 Généralités	37
5.4.2 Application à notre base de données	38
5.5 Réseau Bayésien	39
5.5.1 Généralités	40
5.5.2 Définition formelle	41
5.5.3 Tree Augmented Naïve Bayes TAN	42
5.6 Application à notre base de données	42
5.6.1 Tree Augmented Naïve Bayes TAN	42
5.6.2 Couverture de Markov	43
5.6.3 Comparaison des modèles TAN et CM	44
5.7 Choix du meilleur modèle	45
5.7.1 Conclusion	46

Introduction

Dans ce chapitre on va détailler deux des méthodes les plus populaires et les plus utilisées dans la modélisation des scores ou plus spécifiquement la probabilité de défaut qu'il soit un score d'octroi ou de comportement, on va également exposer les points forts et les points faibles de chaque méthode, à savoir *Régression Logistique*, *Arbre de décision* et *Réseau bayésien*

5.1 Méthodes utilisées en Machine Learning

5.1.1 Définition

L'apprentissage automatique ou apprentissage statistique (machine learning en anglais), champ d'étude de l'intelligence artificielle, concerne la conception, l'analyse, le développement et l'implémentation de méthodes permettant à une machine (au sens large) d'évoluer par un processus systématique, et ainsi de remplir des tâches difficiles ou problématiques à remplir par des moyens algorithmiques plus classiques.

L'analyse peut concerner des graphes, arbres, ou courbes (par exemple, la courbe d'évolution temporelle d'une mesure; on parle alors de données continues, par opposition aux données discrètes associées à des attributs-valeurs classiques) au même titre que de simples nombres. Les algorithmes utilisés :

- La régression logistique
- Les arbres de décision
- Le Réseau bayésien

5.2 Régression logistique

5.2.1 Généralités

Lors de la modélisation de la variable à expliquer (PD) en utilisant la régression linéaire, on obtient une équation de la forme :

$$D = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Avec : les $x_i, i = 1..n$: les variables explicatives

Lorsqu'on estime les β_i par la méthode du maximum du vraisemblance, deux problèmes majeurs surviennent :

1. La variable à expliquer PD (probabilité de défaut) ne suit pas une loi normale \mathcal{N} mais elle suit une Bernoulli \mathcal{B}
2. Rien ne s'assure que le résultat obtenu va être compris entre 0 et 1

Considérons maintenant la fonction :

$$f(z) = \frac{1}{1+e^{-z}}$$

Pour toute valeur possible de z , le résultat est toujours compris entre 0 et 1. La probabilité peut ainsi être écrite sous la forme suivante :

$$P(D = 1 | x_1, x_2, \dots, x_n) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

on obtient ainsi le modèle logistique est qui a pour allure (voir figure ??)

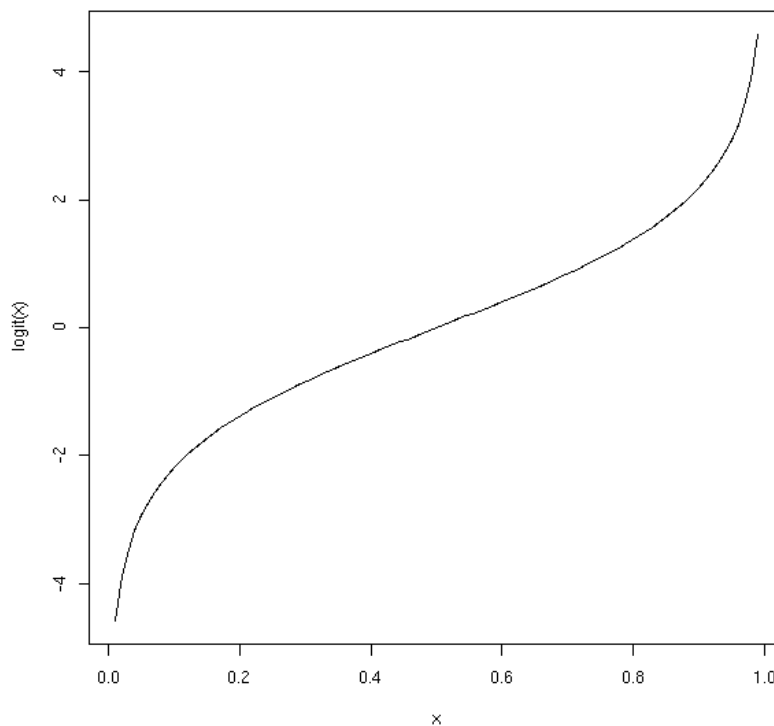


FIGURE 5.1 – Allure de la fonction logit

la probabilité complémentaire de l'événement de défaut s'écrit alors :

$$P(D = 0 | x_1, x_2, \dots, x_n) = 1 - \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} = \frac{1}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

En reformulant en terme de côtes (odds), le modèle devient :

$$\frac{P(D = 1 | x_1, x_2, \dots, x_n)}{P(D = 0 | x_1, x_2, \dots, x_n)} = e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}$$

ou en terme de $\log(odds)$:

$$\ln\left(\frac{P(D = 1 | x_1, x_2, \dots, x_n)}{P(D = 0 | x_1, x_2, \dots, x_n)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

5.2.2 Le modèle logistique dans la modélisation du score crédit

La régression logistique est une méthode populaire dans la modélisation du score crédit, en particulier pour sa simplicité et sa performance. Comme la régression linéaire, une fois les paramètres sont estimés, la régression peut être évaluée d'une manière simple, ainsi que son efficacité opérationnelle. D'un point de vue interprétation, la régression logistique est facile à interpréter.

5.3 Application à notre base de données

Notre base de données comporte 50000 clients, Comme pour toutes les régressions, la sortie présente un bref tableau récapitulatif de l'étude menée, où apparaît notamment le type de la régression demandée en l'occurrence logit (binaire), dans notre cas la probabilité modélisée est **Defaut=0**.

Notre premier tableau de sortie de SAS expose le test Global $\beta = 0$ qui présente un p-value

Récapitulatif sur la sélection séquentielle							
Etape	Effet		DDL	Nombre dans	Khi-2 du score	Khi-2 de Wald	Pr > Khi-2
	Saisi	Supprimé					
1	nombr_impay		1	1	524.6104		< .0001
2	NINQ		1	2	158.3861		< .0001
3	Anc_Prem_Crd		1	3	109.0616		< .0001
4	Emploi		5	4	74.0451		< .0001
5	MT_PRET		1	5	24.9716		< .0001
6	Anc_Emploi		1	6	9.5390		0.0020
7	Raison		1	7	4.5114		0.0337
8		Raison	1	6		4.5039	0.0338

FIGURE 5.2 – Modèle logit :Récapitulatif sur la sélection séquentielle

(Pr>KHI2) inférieur à 0.05 qui signifie qu'au moins un des facteurs étudiés impact la probabilité de défaut en l'occurrence *nombr_impay*, *NINQ*, *Anc_Prem_Crd*, *Emploi*, *MT_pret*, *Anc_Emploi* et la variable *Raison* qui va sortir de notre modèle.

La partie Analyse des effets Type 3 indique que les variables déjà citées entrent en considération dans le modèle (*p-value* > 0.05).

Les paramètres estimés des variables significatifs (*p-value* < 0.05) apparaissent ensuite, on voit bien que la probabilité de défaut diminue avec l'âge du client et le montant du prêt octroyé ce qui est souvent le cas puisque les clients plus âgés ont tendance à payer leurs crédits comparés aux jeunes et augmente par contre avec le nombre d'impayé des traites ce qui est logique puisque l'événement de défaut est le fait d'avoir un nombre d'impayé supérieur à trois traites ou plus , ainsi l'équation de notre modèle est la suivante :

Analyse des effets Type 3			
Effet	DDL	Khi-2 de Wald	Pr > Khi-2
MT_PRET	1	22.3768	<.0001
Anc_Emploi	1	9.5009	0.0021
Anc_Prem_Crd	1	78.3664	<.0001
Emploi	5	67.7464	<.0001
nombr_impay	1	331.7303	<.0001
NINQ	1	97.5231	<.0001

FIGURE 5.3 – Modèle logit : Analyse des effets Type 3

$$P(D = 0|x_1, x_2, \dots, x_n) =$$

1

$$1 + e^{(0.53+0.0002MT_{PRET}+0.02Anc_{Emploi}+0.0057Anc_{PremCrd}-0.780Emploi1+0.631Emploi3+\dots-0.23NINQ)}$$

Estimations par l'analyse du maximum de vraisemblance						
Paramètre		DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > Khi-2
Intercept		1	0.5342	0.1377	15.0449	0.0001
MT_PRET		1	0.000023	4.764E-6	22.3768	<.0001
Anc_Emploi		1	0.0202	0.00655	9.5009	0.0021
Anc_Prem_Crd		1	0.00573	0.000647	78.3664	<.0001
Emploi	Freelance	1	-0.7805	0.2403	10.5501	0.0012
Emploi	Mgr	1	0.0147	0.1302	0.0128	0.9098
Emploi	Office	1	0.6318	0.1415	19.9401	<.0001
Emploi	Prof	1	0.2330	0.1161	4.0269	0.0448
Emploi	Sales	1	-1.3313	0.2700	24.3024	<.0001
Emploi	autre	0	0	.	.	.
nombr_impay		1	-0.8105	0.0445	331.7303	<.0001
NINQ		1	-0.2317	0.0235	97.5231	<.0001

FIGURE 5.4 – Modèle logit : Estimation par l'analyse du maximum de vraisemblance

Les estimations des Odds ratio (rapport de cotes) ci - dessous, nous permettent d'avancer qu'un client qui est Freelance a 0.458 fois plus de chances de faire défaut qu'un client qui professe (autre) De même, un client travaillant au bureau a environ 2 fois plus de chance (1.881) de tomber en défaut comparé à un client qui à pour métier (autre) . Finalement, le tableau suivant expose les paramètres qui permettent de mesurer la performance et la qualité de notre modèle, en particulier **le paramètre c** qui n'est en fait que la courbe

Estimations des rapports de cotes			
Effet	Valeur estimée du point	95% Intervalle de confiance de Wald	
MT_PRET	1.000	1.000	1.000
Anc_Emploi	1.020	1.007	1.034
Anc_Prem_Crd	1.006	1.004	1.007
Emploi Freelance vs autre	0.458	0.286	0.734
Emploi Mgr vs autre	1.015	0.786	1.310
Emploi Office vs autre	1.881	1.425	2.482
Emploi Prof vs autre	1.262	1.005	1.585
Emploi Sales vs autre	0.264	0.156	0.448
nombr_impay	0.445	0.408	0.485
NINQ	0.793	0.758	0.831

FIGURE 5.5 – Modèle logit : Estimation des rapports de cotes

ROC et D de Somer's qui n'est que l'indice de Gini, en l'occurrence on peut considérer que notre modèle est de bonne qualité puisque $c = 79\%$ est supérieur à 75% et **D de Somer's**= 57% est supérieur à 55% .

La figure suivante expose les courbes ROC pour les différentes étapes lors de la création

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	78.3	D de Somers	0.567
Pourcentage discordant	21.7	Gamma	0.567
Pourcentage lié	0.0	Tau-a	0.180
Paires	2861880	c	0.783

FIGURE 5.6 – Modèle logit : Association des probabilités prédites et des réponses observées

du modèle (méthode stepwise), ainsi comme mentionné plus haut le ROC de notre modèle final est de 0.783

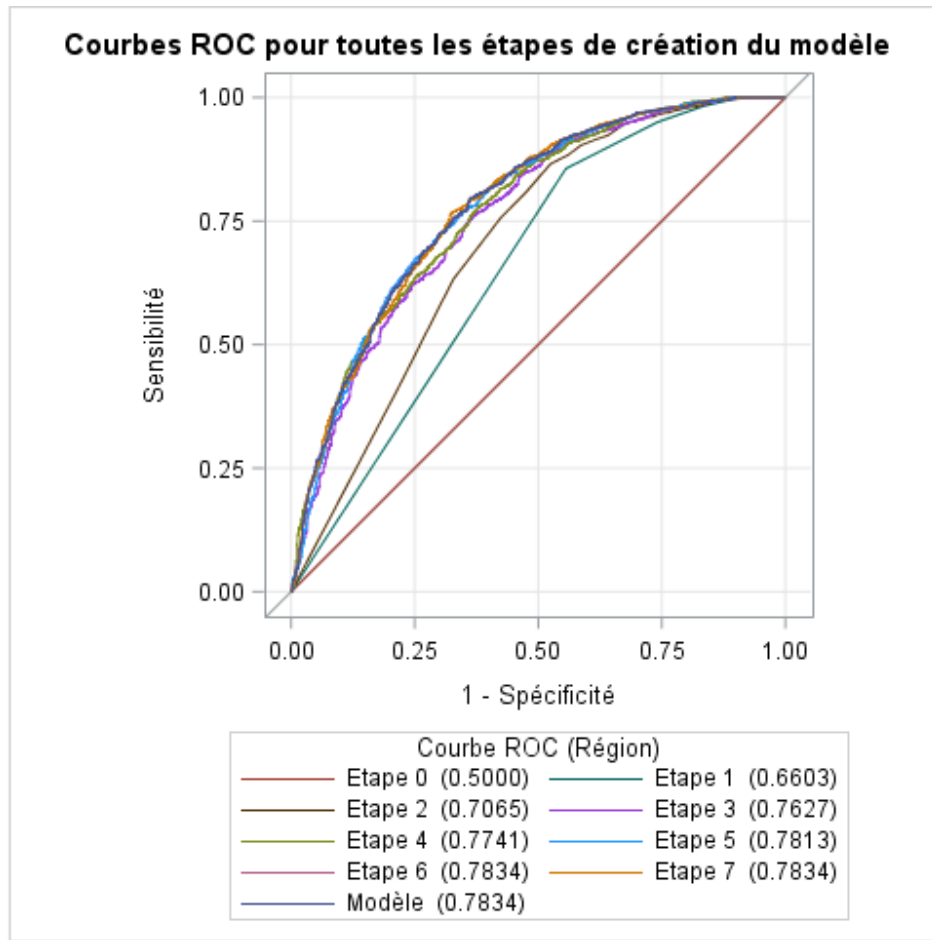


FIGURE 5.7 – Modèle logit :courbe ROC

5.3.1 Calibration des probabilités de défaut

La calibration est l'étape où la probabilité de défaut obtenue pour chaque client est transformée en **scorecard**, pour cela on introduit deux paramètres à savoir :

- Factor : Un des deux paramètres de calibration durant le processus du calcul du score. l'expression du paramètre Factor est : $Factor = \frac{pdo}{\ln(2)}$ avec *pdo* : Points to double the odds, paramètre donné par l'expert. Pour comprendre la signification du *pdo*, combien le score varie si on augmente les côtes de 100 : 1 à 200 : 1.

Une valeur commune du *pdo* est 20, car il permet de créer des intervalles de scores crédit que les clients apprécient.

Donc l'interprétation de $Factor = \frac{20}{\ln(2)}$, pour 20 points de plus d'un score, la côte double.

- Offset : le deuxième paramètre de calibration
- WOE : Weight of evidence ou poids d'évidence en français en pourcentage permet de mesurer le pouvoir prédictif pour chaque variable, il est calculé comme suit :

$$WOE = \ln\left(\frac{Distribution_Bon}{Distribution_Mauvais}\right) \times 100$$

Le calcul des deux premiers paramètres *Factor et Offset* ce fait en fixant un Score (par l'expert). Supposons par exemple qu'on cherche un score de crédit de 100 pour côte 50 : 1, et un score crédit de 120 pour côte 100 : 1, on obtient le système suivant :

$$\begin{aligned} 100 &= offset + factor * \log 50 \\ 120 &= offset + factor * \log 100 \end{aligned}$$

la solution de cette équation est :

$$\begin{aligned} factor &= \frac{20}{\ln(2)} = 28,85 \\ offset &= 100 - factor \times \ln(50) = -12,87 \end{aligned}$$

Après le calcul de ces deux paramètres on peut maintenant obtenir le score crédit pour chaque segment de client de notre portefeuille.

Finalement on calcule le score crédit donné par la formule suivante :

$$\text{Score crédit} = \sum_{i=1}^N (WOE_i * \beta_i + \frac{\beta_0}{N}) * \frac{offset}{N}$$

Avec :

β_0 : la constante dans le modèle logistique

β_i : l'estimation des paramètres des variables significatives

WOE : Weight of evidence (poids d'évidence)

N : le nombre de paramètres significatifs

Sous SAS Miner on a calculé le scorecard pour chaque segment de client, ainsi que leurs WOE : poids d'évidence, figure 5.8 pour voir le tableau dans sa totalité voir Annexe A.1 page I.

Scorecard							
		Groupe	Points du scorecard	Poids d'évidence	Taux d'événement BIN_DEFAULT = 1	Pourcentage de population	Coefficient
Anc_Prem_Crd	Anc_Prem_Crd < 69.13	1.00	58	-1.08	41.98	4.99	-0.97
	69.13 <= Anc_Prem_Crd < 172.5, _MISSING_	2.00	79	-0.33	25.59	45.00	-0.97
	172.5 <= Anc_Prem_Crd < 240.69	3.00	101	0.47	13.34	30.00	-0.97
	240.69 <= Anc_Prem_Crd < 312.74	4.00	113	0.88	9.26	15.00	-0.97
	312.74 <= Anc_Prem_Crd	5.00	96	0.30	15.49	5.02	-0.97
MT_PRET	MT_PRET < 8000	1.00	63	-1.02	40.53	9.70	-0.86
	8000 <= MT_PRET < 15200, _MISSING_	2.00	87	-0.05	20.51	34.90	-0.86
	15200 <= MT_PRET < 21600	3.00	94	0.23	16.34	25.36	-0.86
	21600 <= MT_PRET < 37700	4.00	99	0.44	13.75	25.01	-0.86
	37700 <= MT_PRET	5.00	85	-0.13	21.96	5.04	-0.86

FIGURE 5.8 – Partie des scorecards crédit du modèle logistique de notre portefeuille

Grâce à ce tableau on peut donc calculer le score crédit (de comportement) des clients de notre portefeuille, ce qui permet de donner une idée sur leur comportement futur vis à vis le paiement de leurs traites, par exemple un client qui a les caractéristiques suivantes :

1. Nombre d'impayé :2
2. Ancienneté premier crédit :60*mois*
3. Emploi :*Autre*
4. Montant prêté :40000DHs
5. Valeur propriété :120000DHs
6. NINQ : 1

Ce client dont les caractéristiques sont listées en dessus a pour score :

$$Score_Crédit = 42 + 58 + 85 + 97 + 93 = 375$$

5.4 Les arbres de décision

Dans cette partie, on va évaluer la performance d'un type d'arbre de décision à savoir **CART** (Classification and Regression Tree).

5.4.1 Généralités

Complémentaires des méthodes statistiques plus classiques : analyse discriminante, régression linéaire, les solutions obtenues à travers les arbres de décision sont présentées sous une forme graphique simple à interpréter et constituent une aide efficace pour l'aide à la décision.

Construction de l'arbre CART

L'arbre est de type binaire, les données sont constituées de l'observation de p variables quantitatives ou qualitatives explicatives X_j et d'une variable à expliquer Y qualitative à m modalités $\Gamma \{1, \dots, m$ ou qualitative réelle, observées sur un échantillon de n individus. La construction d'un arbre de discrimination binaire consiste à déterminer une séquence de nœuds selon l'algorithme suivant :

- Un nœud est défini par le choix conjoint d'une variable parmi les explicatives et d'une division qui induit une partition en deux classes. Implicitement à chaque nœud correspond donc un sous-ensemble de l'échantillon auquel est appliquée une dichotomie.
- Une division est elle-même définie par une valeur seuil de la variable quantitative sélectionnée ou un partage en deux groupes des modalités si la variable est qualitative.
- À la racine ou nœud initial correspond l'ensemble de l'échantillon ; la procédure est ensuite itérée sur chacun des sous-ensembles

5.4.2 Application à notre base de données

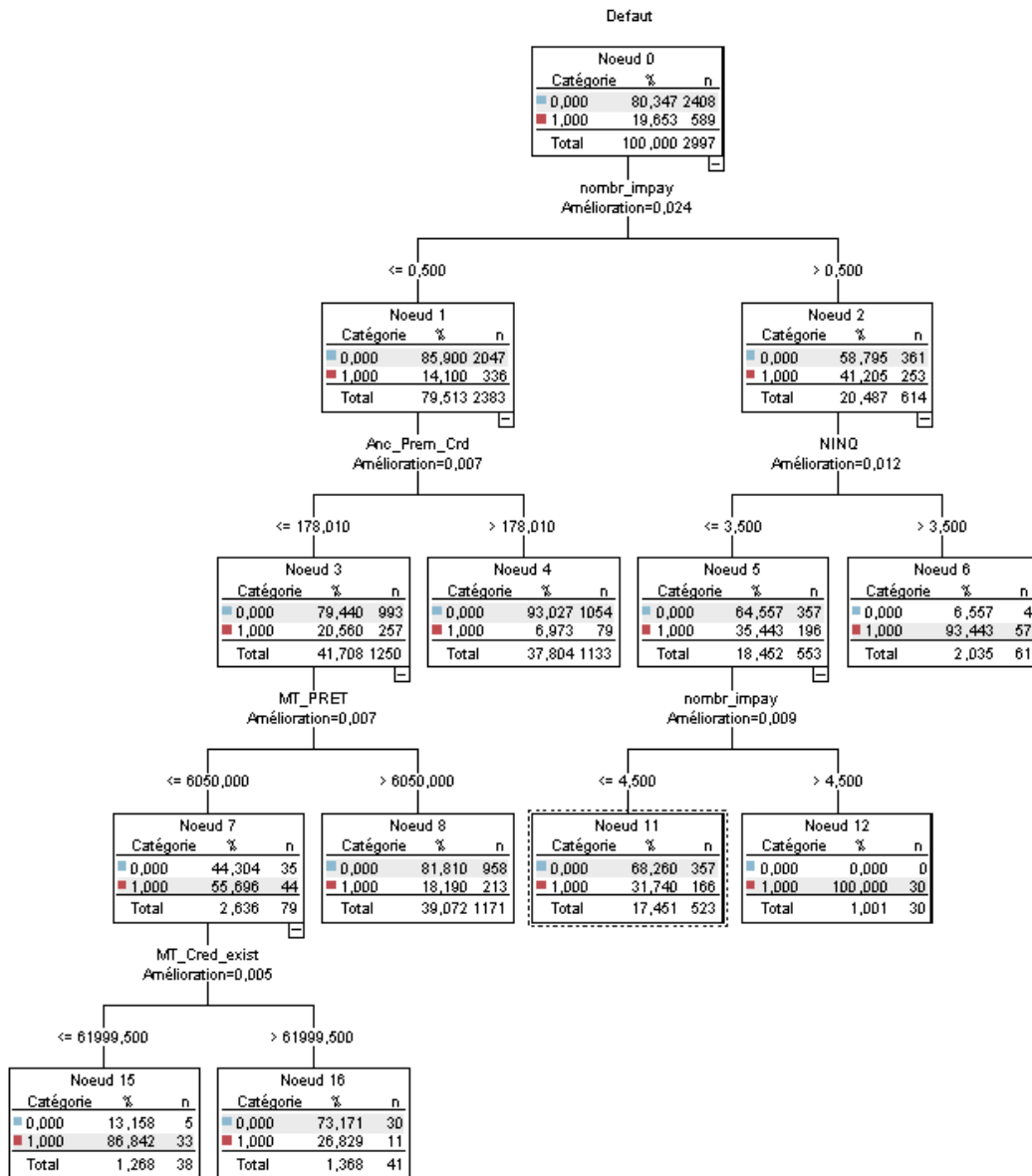


FIGURE 5.9 – Graphe CART

Le graphe CART de la sortie retiens 5 variables explicatives à savoir **Nombr_impaye**, **Anc_Prem_Cred**, **NINQ**, **MT_Pret** et **MT_Cred_exist**, il contient 16 nœuds, le premier nœuds segmente la variable nombre d'impayé en deux segments un comprend les clients qui n'ont jamais eu d'impayé et une deuxième branche qui comprend les clients qui ont eu un ou plus d'impayé, les clients qui ont plus d'un impayé sont à leur tour divisés à trois segments, de même que pour les clients qui n'ont jamais eu d'impayé sont segmentés en 6 segments.

Même si la technique de l'arbre de décision CART permet de donner un premier aperçu sur la distribution du défaut pour chaque segment, elle présente un inconvénient princi-

pal, ce qu'elle ne fait pas de différence entre les variables qualitatives continues et celles discrètes, en effet dans notre exemple la variable **Nombr_impay** qui est une variable discrète est segmentée en deux segments $\text{Nombr_impay} \leq 0.5$ et l'autre segment $\text{Nombr_impay} > 0.5$, ce qu'il n'a pas de sens puisque la variable **Nombr_impay** est un nombre entier positif.

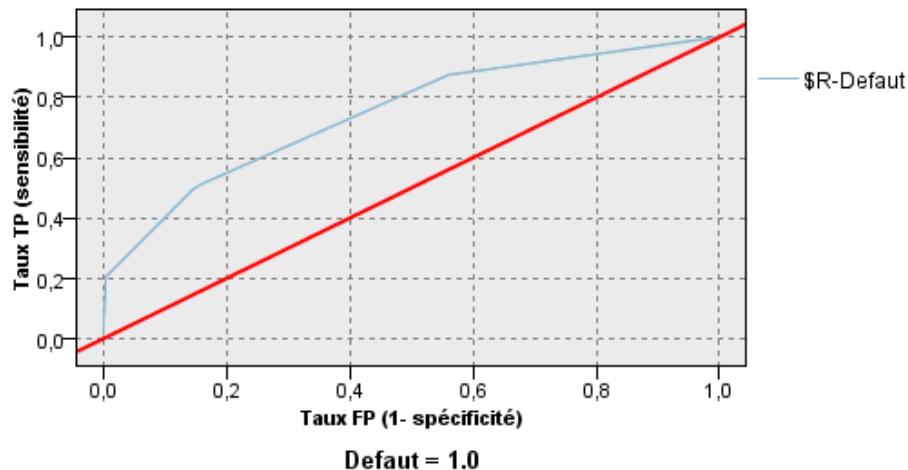


FIGURE 5.10 – Courbe Roc pour le graphe CART

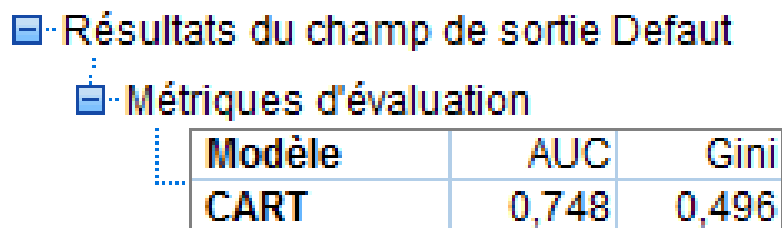


FIGURE 5.11 – Évaluation de l'arbre de décision CART

Évaluation du modèle

Les figures 5.10 et 5.11 nous informe sur la performance du graphe CART, d'après l'indice de Gini qui égale environ à 50% on peut considérer que c'est un mauvais modèle, comparé au modèle logistique étudié plus haut le modèle CART est moins performant en terme de prédire les clients qui vont tomber en défaut et ceux qui ne le vont pas.

5.5 Réseau Bayésien

Introduction

Les réseaux Bayésiens permettent de créer un modèle de probabilité en combinant les preuves observées et enregistrées avec les connaissances réelles "de bon sens" pour

établir la probabilité des occurrences.

Par exemple, imaginons qu'une banque s'inquiète des prêts susceptibles de ne pas être remboursés. Si les données par défaut des prêts précédents peuvent être utilisées pour prédire quels clients potentiels risquent d'avoir des difficultés à rembourser leur prêt, il est alors possible de refuser un prêt à ces clients à "fort risque" ou de leur proposer d'autres produits.

Dans notre étude, on se concentre sur l'utilisation de données de défaut sur des prêts existants permettant de prédire quels futurs clients sont susceptibles de ne pas pouvoir rembourser leur prêt, on examinera pour cela deux différents types de modèle de réseau (TAN et **Couverture de Markov**) Bayésien afin d'établir celui le plus adapté à cette situation.

5.5.1 Généralités

Un réseau bayésien est en informatique et en statistique un modèle graphique probabiliste représentant des variables aléatoires sous la forme d'un graphe orienté acyclique. Intuitivement, ils sont à la fois :

1. des modèles de représentation des connaissances
2. des modèles de représentation des connaissances
3. une base pour des Systèmes d'aide à la décision

Pour un domaine donné (par exemple médical), on décrit les relations causales entre variables d'intérêt par un graphe. Dans ce graphe, les relations de cause à effet entre les variables ne sont pas déterministes, mais probabilisées. Ainsi, l'observation d'une cause ou de plusieurs causes n'entraîne pas systématiquement l'effet ou les effets qui en dépendent, mais modifie seulement la probabilité de les observées.

L'intérêt particulier des réseaux bayésiens est de tenir compte simultanément de connaissances a priori d'experts (dans le graphe) et de l'expérience contenue dans les données.

Les réseaux bayésiens sont surtout utilisés pour le diagnostic (médical et industriel), l'analyse de risques (ex : crédit), la détection des spams et le data mining.

Construction de réseaux bayésiens

Construire un réseau bayésien, se décompose en :

1. définir le graphe du modèle
2. définir les tables de probabilité de chaque variable, conditionnellement à ses causes.

Le graphe est aussi appelé la « structure » du modèle, et les tables de probabilités ses « paramètres ». Structure et paramètres peuvent être fournis par des experts, ou calculés à partir de données, même si en général, la structure est définie par des experts et les paramètres calculés à partir de données expérimentales.

5.5.2 Définition formelle

Loi de probabilité jointe

Il existe plusieurs façons de définir un réseau bayésien. La plus simple exprime la loi de probabilité jointe sur l'ensemble des variables aléatoires modélisées dans le réseau. Un réseau bayésien est un graphe orienté acyclique $G = (V, E)$ avec V l'ensemble des nœuds du réseaux et E l'ensemble des arcs. À chaque nœud x appartenant à V du graphe est associé la distribution de probabilité conditionnelle suivante :

$$P(x|pa(x)) \text{ où } pa(x) \text{ représente les parents immédiats de } x \text{ dans } V$$

L'ensemble V est donc un ensemble de variables aléatoires discrètes. Chaque nœud de V est conditionnellement indépendant de ses non-descendants, étant donné ses parents immédiats. Il est ainsi possible de factoriser les distributions de probabilité conditionnelles sur l'ensemble des variables en en faisant le produit :

$$P(V) = \prod_{x \in V} P(x|pa(x)) \text{ Ce résultat est parfois noté JPD, pour distribution de probabilité jointe}$$

Propriété de Markov globale

Une autre façon de définir un réseau bayésien est d'évaluer si un graphe orienté acyclique $G = (V, E)$ vérifie la propriété de Markov globale (dite orientée ici, en raison de la nature de G) étant donné une loi de probabilité P sur l'ensemble des variables V . Soit X, Y, S trois sous-ensemble de V , tels que X et Y sont d-séparés par S , noté $(X|S|Y)$: la propriété de Markov globale exprime l'indépendance conditionnelle entre X et Y , c'est-à-dire que X est indépendant de Y conditionnellement à S . Formellement, la propriété dans un graphe orienté s'énonce ainsi :

$$\forall X, Y, S \subset V \text{ disjoints, } (X|S|Y) \Rightarrow X \perp Y|S$$

Il convient de définir la d-séparation, pour « séparation dirigée » ou « séparation orientée », dans un graphe orienté. Les sous-ensembles X et Y sont d-séparés par S si et seulement si toute chaîne de nœuds de X à Y est « bloquée » par S . Pour trois nœuds x appartenant à X , y appartenant à Y et s appartenant à S , une chaîne est bloquée dans les deux cas suivants :

- soit la chaîne contient une séquence $x \rightarrow s \rightarrow y$ ou $x \leftarrow s \rightarrow y$
- soit la chaîne contient une séquence $x \rightarrow z \leftarrow y$ où z n'appartient pas à S et aucun descendant de z appartient à S .

L'intuition est que S « bloque » l'information entre X et Y .

L'indépendance conditionnelle enfin, notée ci-dessus $X \perp Y|S$, exprime que X est conditionnellement indépendant de Y étant donné S . Formellement :

$$X \perp Y|S \text{ si et seulement si } P(X|Y, S) = P(X, S) \text{ et } P(Y|X, S) = P(Y, S)$$

5.5.3 Tree Augmented Naïve Bayes TAN

Modèle bayésien naïf est un modèle probabiliste pour un classifieur est le modèle conditionnel $P(C|F_1...F_n)$ où C est une variable de classe dépendante dont les instances ou classes sont peu nombreuses, conditionnée par plusieurs variables caractéristiques $F_1...F_n$.

Lorsque le nombre de caractéristiques n est grand, ou lorsque ces caractéristiques peuvent prendre un grand nombre de valeurs, baser ce modèle sur des tableaux de probabilités devient impossible. Par conséquent, nous le dérivons pour qu'il soit plus facilement soluble.

5.6 Application à notre base de données

5.6.1 Tree Augmented Naïve Bayes TAN

On va en premier lieu utiliser la méthode TAN pour créer le réseau bayésien de notre base de donnée et faire ressortir les variables qui ont un effet discriminant sur le défaut du client.

La figure de dessous illustre le graphe du modèle TAN : On remarque d'après le graphe

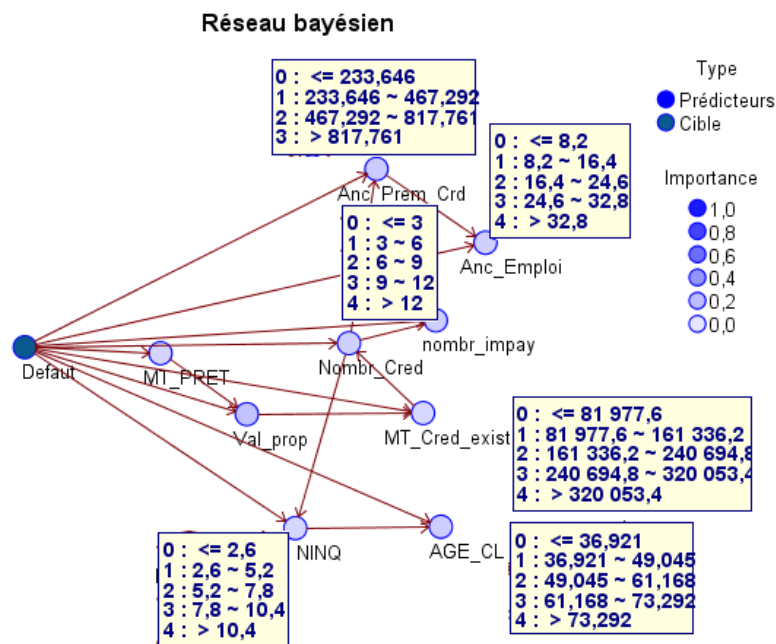


FIGURE 5.12 – Graphe du modèle TAN

que la variable cible à savoir **le défaut** peut être expliqué par 9 variables explicatives, le graphe nous indique aussi l'importance l'importance de ces prédicteurs dans notre cas tous les variables explicatives ont presque la même importance qui est d'environ 0.2. Le graphe nous permet aussi de savoir s'il y a une dépendance entre les variables expli-

catives en l'occurrence les variables MT_pret et Val_prop sont reliées par une flèche c-à-d qu'il existe une dépendance entre ces deux variables, cette constatation est déjà faite lors d'étude descriptive de notre base de données.

On remarque aussi que le réseau bayésien permet de segmenter les variables dans notre cas que la variable ancienneté emploi est segmentée en 4 segments. La figure suivante est

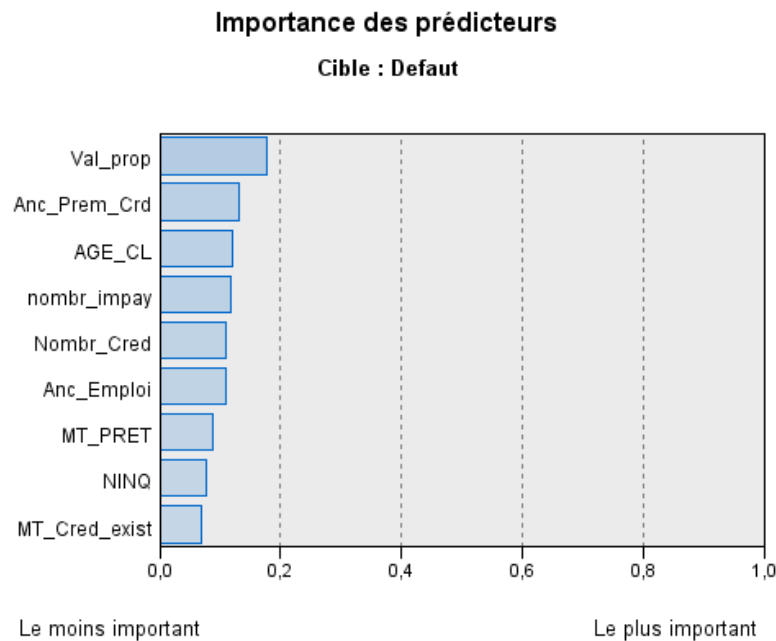


FIGURE 5.13 – Importance des variables du modèle TAN

complémentaire de celle qui l'a précédé, il indique l'importance des variables retenues par le réseau bayésien, en effet la variable val_prop se distingue des autres variables par une importance d'environ 0.2 suivi par la variable Anc_Prem_Crd par une importance de 0.13 et la variable retenue par le réseau et qui explique le moins c'est MT_Cred_exist avec une importance 0.07.

5.6.2 Couverture de Markov

Dans cette deuxième partie, on va appliquer un autre type de structure appelé **couverture de Markov**, contrairement au TAN qui ne permet pas d'explicitement les probabilités de défaut par client en fonction de ces caractéristiques, la structure de Markov calcule la probabilité de défaut pour chaque segment des variables explicatives comme indiqué par l'extrait de tableau des probabilités conditionnelles ci-dessous :

Ce tableau 5.14 expose les variables explicatives retenues par la couverture de Markov, comparé au type TAN, la couverture de Markov retient 2 variables de moins, on en plus remarque que ces variables sont segmenter en fonction du graphe de représentation (voir **Annexe**), et pour chaque segment une probabilité de défaillance est calculée.

Probabilités conditionnelles de Défaut								
Parents							Probabilité	
MT_PRET	MT_Cred_exist	Val_prop	Anc_Emploi	Anc_Prem_Crd	nombr_impay	NINQ	1	0
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	<= 2,6	0,23	0,77
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	2,6 ~ 5,2	0,36	0,64
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	5,2 ~ 7,8	0,80	0,20
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	7,8 ~ 10,4	0,38	0,62
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	> 10,4	1,00	0,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	3 ~ 6	<= 2,6	0,75	0,25
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	3 ~ 6	2,6 ~ 5,2	1,00	0,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	3 ~ 6	5,2 ~ 7,8	1,00	0,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	6 ~ 9	<= 2,6	1,00	0,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	233,646 ~ 467,292	<= 3	<= 2,6	0,09	0,91
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	0,33	0,67

FIGURE 5.14 – Extrait du tableau des probabilités de défaut par segment

Prenons par exemple la première ligne de notre tableau, pour un client qui a les caractéristiques suivantes :

- un montant prêté inférieur ou égale à 18800
- le montant des crédits existants inférieur ou égale à 81977
- la valeur de son logement ou égale 112530
- Une ancienneté d’emploi inférieur ou égale à 8.2 ans
- Un nombre d’impayé inférieur ou égale à 3 impayés

à une probabilité de défaillance futures de 0.23.

5.6.3 Comparaison des modèles TAN et CM

Dans cette section on va comparer les deux modèles du réseau bayésien TAN et CM exposés plus haut afin d’établir celui le plus adapté à notre base de données, pour cela on va comparer leur ROC ainsi que leur indice Gini. D’après les deux figures 5.15 et 5.16 on

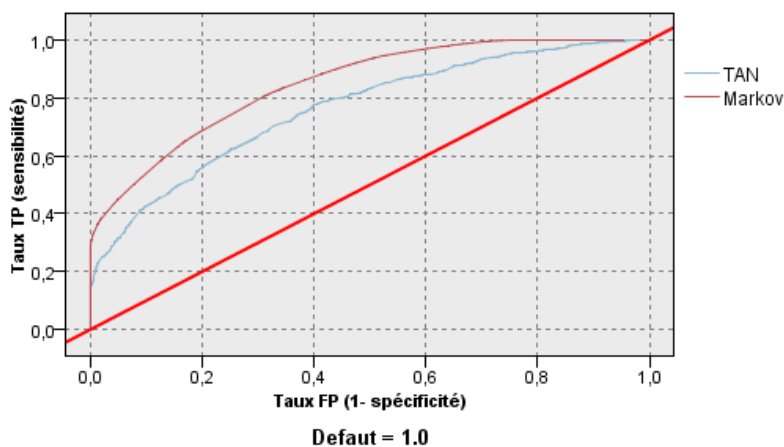


FIGURE 5.15 – Comparaison du TAN et CM via leurs courbes ROC

peut affirmer que le modèle CM est plus performant que le modèle TAN puisque la courbe

☐ Résultats du champ de sortie Defaut

☐ Métriques d'évaluation

Modèle	AUC	Gini
TAN	0,763	0,526
Markov	0,848	0,696

FIGURE 5.16 – Tableau indice Gini : comparaison TAN et CM

ROC de CM (en rouge) est en dessus de celle du TAN (en bleu), d'autre part l'indice de Gini qui correspond à celui de modèle CM (0,696) est supérieur à celui du TAN (0,526), l'indice de Gini vient à son tour confirmer la performance du modèle Markovien, avec un indice d'environ 0.7.

5.7 Choix du meilleur modèle

Dans cette section finale de notre chapitre on va comparer les quatre méthodes exposées précédemment quant à leur performance vis à vis la prédiction des clients qui vont faire défaut dans le futur. Il ressort de la figure 5.17 que le modèle du réseau bayésien de

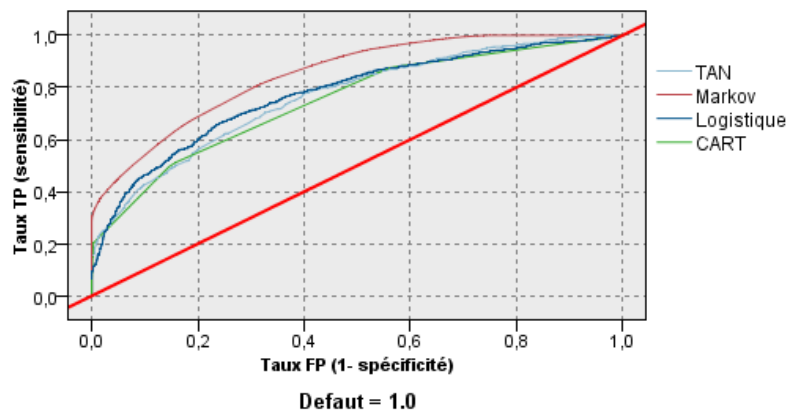


FIGURE 5.17 – Courbes Roc : comparaison des modèles

☐ Métriques d'évaluation

Modèle	AUC	Gini
TAN	0,763	0,526
Markov	0,848	0,696
Logistique	0,773	0,547
CART	0,748	0,496

FIGURE 5.18 – Indice Gini : Évaluation des modèles

type Markovien est le plus performant puisque sa courbe est en dessus des autres, la figure 5.18 vient confirmer cette constatation avec un $AUC(Markov) = 84\%$ et $Gini(Markov) = 70\%$ suivi par le modèle logistique tel que $AUC(Logistique) = 77\%$ et $Gini(Logistique) = 55\%$.

5.7.1 Conclusion

On a modélisé la probabilité de défaut par trois méthodes, il s'avérait que la plus performante est le réseau bayésien markovien, Cependant le réseau bayésien manque de transparence dans le sens où on ne peut pas l'interpréter à cause du manque des coefficients des paramètres des variables explicatives on parle dans ce cas d'une boîte noire puisque le traitement et le choix des variables explicatives se fait sans indication de même pour le graphe CART, contrairement au modèle logistique qui lui est considéré comme une boîte transparente du moment où les coefficients des paramètres des variables explicatives sont explicités de plus il est facilement interprétable grâce aux côtes (odds).

Chapitre 6

Les modèles de durée de vie et modélisation du PD de comportement

Sommaire

6.1 Les modèles à temps discret	48
6.1.1 L'événement de défaut	48
6.1.2 Le défaut conditionnel	48
6.1.3 Les modèles non-linéaire :Probit,Cloglog	49
6.2 Les modèles à temps continu	53
6.2.1 Introduction	53
6.3 La censure	53
6.4 Application à notre portefeuille	54
6.4.1 Traitement des données	54
6.4.2 Modèle semi-paramétrique : Modèle de Cox	55
6.4.3 Estimation des paramètres du modèle	57
6.4.4 Estimation de la fonction de survie	59
6.4.5 Modèle stratifié	60
6.4.6 Test de proportionnalité	62
6.4.7 Validation du modèle de Cox	63
6.5 Conclusion	64

6.1 Les modèles à temps discret

6.1.1 L'événement de défaut

La probabilité de défaut décrit la vraisemblance de l'événement de défaut. Les organismes de crédit observent si un emprunteur va faire défaut dans le temps, et généralement cet événement est indiqué par un indicateur binaire de défaut :

$$D_{it} = \begin{cases} 1 & \text{si le client } i \text{ fait défaut à l'instant } t \\ 0 & \text{sinon} \end{cases}$$

avec $i = 1, \dots, n$ and $t = 1, \dots, T$. On suppose que l'événement de défaut est une variable aléatoire, la lettre "Grand" D définit la variable aléatoire, alors que le "Petit" d sa réalisation.

L'événement de défaut est peut être définie par l'un des événements suivants :

- Un nombre de jours impayés ou plus; en pratique les seuils sont de 30, 60, et 90 jours
- Faillite de l'emprunteur

D'autres définitions peuvent exister dans le monde du risque crédit, dans notre rapport on travaillera avec la définition par Bâle qui stipule qu'un client est en défaut si le nombre de jours impayés est 90 jours ou plus.

6.1.2 Le défaut conditionnel

Les banques et les organismes de crédit suivent l'emprunteur dans le temps et incluent des informations dynamiques dans leurs mesures du risque crédit. L'écoulement du temps enrichit les informations sur le client et sur le prêt (*informations idiosyncratiques* ou *propres*), ou des *informations macroéconomiques* (*informations systématiques*). En pratique l'évaluation du risque crédit peut être journalière (coûteux), mensuelle ou annuelle. Le choix de la période d'évaluation dépend essentiellement sur la disponibilité et la facilité d'obtention des informations, dépend aussi de l'objectif qu'on vise par exemple l'allocation du capital sous Bâle nécessite une estimation pour une période future courte, quant aux provisions pour perte dû au risque crédit sous les normes IFRS.9, l'estimation dans ce cas est faite sur plusieurs périodes et souvent sur toute la durée du crédit. Les défauts sont souvent des événements terminaux, l'événement de défaut sur une période est conditionnel par la survie du crédit (i.e Non défaut) sur les périodes précédentes.

Les banques mesurent la probabilité de défaut (PD) pour deux éléments principaux l'emprunteur et l'instrument financier utilisé. Comme mentionné précédemment (cf. Chapitre 3.1

p. 18), la terminologie de la probabilité de défaut est généralement utilisée pour référer à une période de probabilité de défaut. Autres termes utilisés, la probabilité de défaut conditionnelle (conditionnelle par la survie du crédit) ou intensité de défaut.

La probabilité de défaut inconditionnelle est souvent utilisée pour calculer la vraisemblance de cette probabilité suivant la perspective de la date origine du prêt, alors que la probabilité de défaut conditionnelle. Les deux probabilités sont égales pour la première période.

Supposons que chaque clients ont les mêmes probabilités conditionnelles de défaut $PD_{t-1,t}$ pour chaque période $[t, t - 1]$, on peut omettre l'indice i de chaque client. La probabilité de défaut inconditionnelle $PDI_{t-1,t}$, peut être écrite comme suit :

$$\begin{aligned}
 PDI_{t-1,t} &= S(t_1) - S(t_2) \\
 &= \prod_{t=1}^{t_1} (1 - PD_{t-1,t}) - \prod_{t=1}^{t_2} (1 - PD_{t-1,t}) \\
 &= \prod_{t=1}^{t_1} (1 - PD_{t-1,t}) - \prod_{t=1}^{t_1} (1 - PD_{t-1,t}) \prod_{t=t_1}^{t_2} (1 - PD_{t-1,t}) \\
 &= \prod_{t=1}^{t_1} (1 - PD_{t-1,t}) (1 - \prod_{t=t_1}^{t_2} (1 - PD_{t-1,t})) \\
 &= S(t_1) PD_{t_1,t_2}
 \end{aligned}$$

$S(t)$ c'est la fonction de survie à l'instant t (i.e pas de défaut jusqu'à l'instant t).

6.1.3 Les modèles non-linéaire :Probit,Cloglog

Spécification de la fonction de lien

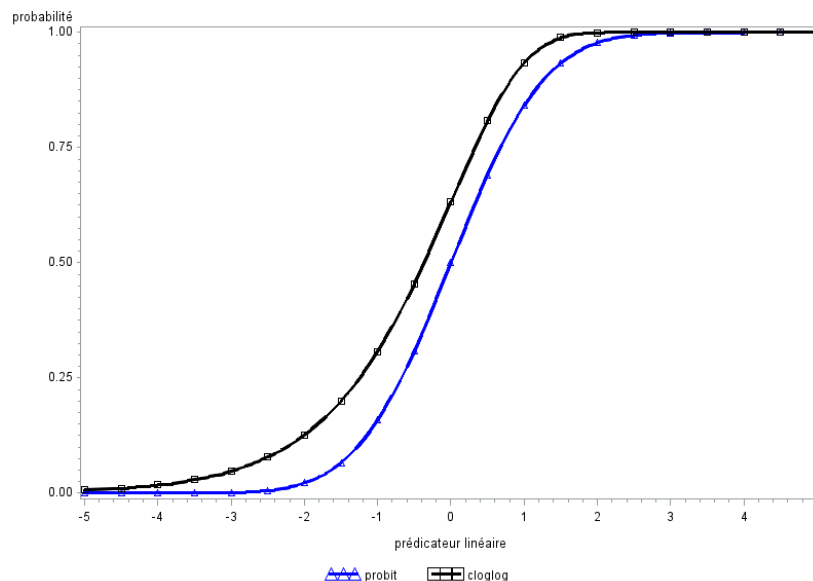


FIGURE 6.1 – Fonction de lien non linéaire :Probit et Cloglog

Comme cité précédemment, chaque fonction de lien à ses particularités, par exemple, la fonction **logit** permet de calibrer les PDs pour différents niveaux, la fonction **probit** quant à elle, est particulièrement utile pour l'estimation des paramètres dans le cas des modèles IRB internal ratings based (IRB) en respect avec la régulation Bâle. Dans ce qui suit on va comparer les deux modèles à savoir PROBIT et CLOGLOG(log log complémentaire), sur notre base de données.

L'équation du modèle probit est :

$$P(D_{it} = 1 | X_{it-1}) = \Phi(\beta' x_{it-1}) \quad \text{Avec } \Phi(.) \text{ la fonction densité cumulative de la loi normale centrale réduite, } \beta \text{ vecteur des paramètres et } x_{it-1} \text{ vecteur des co-variables.}$$

Estimation des paramètres par Maximum

La méthode utilisée dans SAS lors de l'estimation des paramètres est le score de Fisher via la maximisation du log de vraisemblance cf figure 6.2.

La vraisemblance est définie comme suit :

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

Avec $P(d_{it} = 1) = \Phi(\beta' x_{it-1})$

La fonction de vraisemblance est ensuite transformée par la fonction logarithme en :

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

TTC vs PIT

Généralement la modélisation de la probabilité de défaut est générée par deux méthodologie à savoir :

- Le modèle TTC '**Through The Cycle**' ne prend généralement pas en compte l'état de l'économie en excluant les variables macroéconomiques (**modèle Cloglog**).
- Le modèle PIT '**Point In Time**' par contre prend en compte les variables qui dépendent du temps (variables macroéconomiques) en plus des variables socioéconomiques des clients (**modèle Probit**).

En réalité la distinction entre les deux modèles est n'est pas facile à appréhender du fait que les facteurs du risque liés aux variables propres au client varient dans le temps et sont souvent corrélées avec l'état de l'économie.

Comme première comparaison, on va estimer les deux modèles **TTC** et **PIT**

Le modèle **TTC** comme mentionné précédemment il se base sur des variables qui ne dépendent pas du temps (i.e informations qui sont recueillies au moment origine).

Application de la méthodologie PIT

On se limite dans notre rapport à la méthodologie PIT, en utilisant le modèle probit, les figures suivantes sont les sorties du modèle probit sous SAS en modélisant la probabi-

lité que le client tombe en défaut par les variables score d'octroi du client, rapport valeur prêt et la variable macroéconomique PIB.

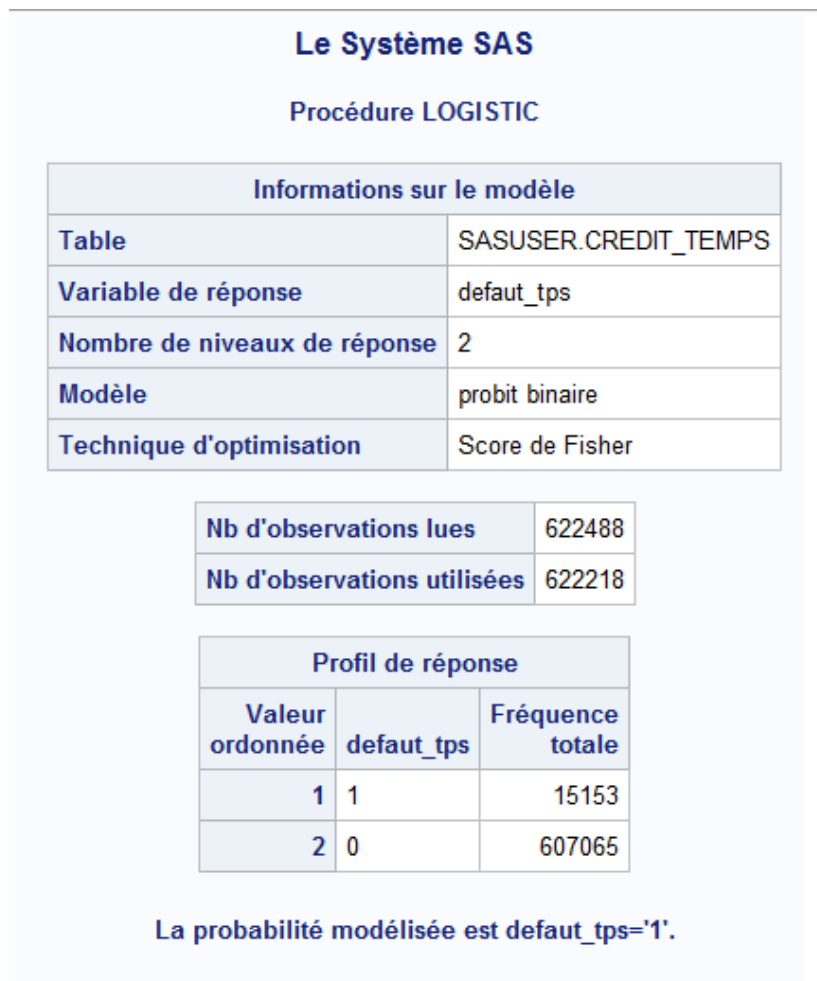


FIGURE 6.2 – modèle probit

Le tableau d'estimation des paramètres indique que notre modèle contient la constante qui égale à -1.2203 , ainsi que le score d'octroi, on remarque que le signe du paramètre estimé du score d'octroi 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, alors que le PIB n'affecte pas la probabilité de défaut dans notre modèle.

Quant à la figure 6.5 compare la probabilité de défaut moyenne sur une durée de 50 mois des modèles logit, PIT et TTC, on remarque que les modèles logit qui modélise la probabilité de défaut réel et PIT (probit) sont presque identiques, car le modèle PIT intègre un nombre de variables macroéconomiques donc la probabilité de défaut du modèle PIT va expliquer d'une manière plus précise et plus adéquate le taux de défaut réel.

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi-2	DDL	Pr > Khi-2
Rapport de vrais	6700.5450	3	<.0001
Score	6822.8919	3	<.0001
Wald	5648.9561	3	<.0001

Estimations par l'analyse du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > Khi-2
Intercept	1	-1.2203	0.0344	1255.7394	<.0001
Scor_Octroi	1	-0.00236	0.000050	2260.7557	<.0001
Rapp_pret_val_temps	1	0.00915	0.000149	3796.1322	<.0001
PIB_temps	1	-838E-13	1.678E-9	0.0025	0.9602

FIGURE 6.3 – Modèle probit : TTC

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	70.6	D de Somers	0.412
Pourcentage discordant	29.4	Gamma	0.412
Pourcentage lié	0.0	Tau-a	0.020
Paires	9198840792	c	0.706

FIGURE 6.4 – Modèle probit :TTC

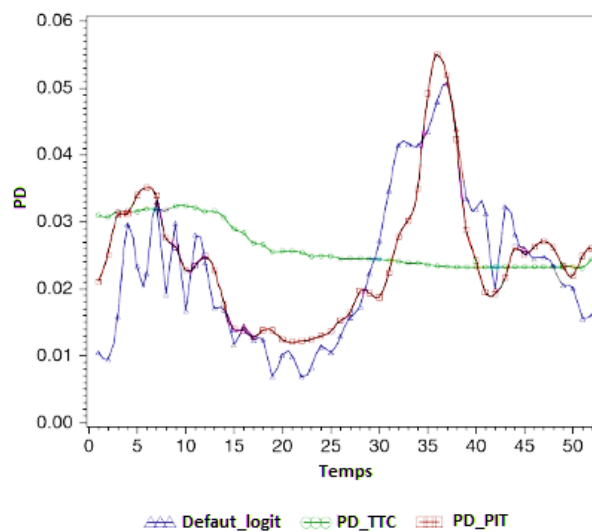


FIGURE 6.5 – Comparaison des probabilités de défaut moyenne pour les modèles : Logit, PPT et TTC

6.2 Les modèles à temps continu

6.2.1 Introduction

Les modèles à temps continu de survie dans le contexte du risque crédit décrit le temps de survie d'un prêt T_i ($T_i \geq 0$) avant de tomber en défaut, qu'on va considérer comme une variable aléatoire continue. Le temps de survie est généralement mesuré dès l'octroi du crédit ou depuis le temps d'observation initiale si cette donnée est disponible, dans la suite de notre étude on travaillera avec la seconde approche i.e le temps d'observation initiale et notre fenêtre d'observation s'étalera sur une période 44 mois.

6.3 La censure

- On rencontre habituellement ce type de données lorsque des unités statistiques (des individus ou autres) n'ont pas subi l'événement à la fin de l'étude ou sont pour une cause ou une autre, perdus de vue avant la fin prévue de l'expérience.
- Supposons qu'on s'intéresse à la durée de chômage. La durée de chômage est connue seulement pour une portion des individus étudiés; pour les autres une borne inférieure (qu'on appelle temps de censure) pour la durée de chômage est connue sans qu'il soit possible d'être plus précis. C'est le cas des individus qui sont encore à la recherche d'un emploi à la fin de la période d'observation, où manque d'informations sur certains individus, qui peut être dû à divers faits tels que le refus de réponse, les décès et les émigrations. Il s'agira de censure à droite.
- Supposons maintenant une situation où l'observation des durées ne se fait pas immédiatement lors de l'entrée des individus dans l'expérience. Il peut arriver que certains des individus trouvent un emploi entre leur entrée dans l'expérience et le moment où on commence à prendre les observations. Le fait que ces individus ont trouvé du travail dans cet intervalle de temps est alors connu, mais les temps précis de chômage, eux, sont inconnus; pour cette raison, on ne possédera qu'une borne supérieure pour la durée de chômage de ces individus. Il est aussi possible que certains individus soient au chômage au début de la période d'observation, sans connaître leurs durées de chômage avant le début de la période d'observation. Il est alors raisonnable de croire que la durée de chômage déclarée sous estime la durée réelle de recherche d'emploi. Il s'agira cette fois dans les deux cas de censure à gauche.

La figure 6.6 illustre les différents aboutissements d'un crédit.

La censure à droite est une spécificité du crédit pour une maturité donnée (crédit limité dans le temps). Par conséquent, le résultat observé pour un intervalle de temps donné est :

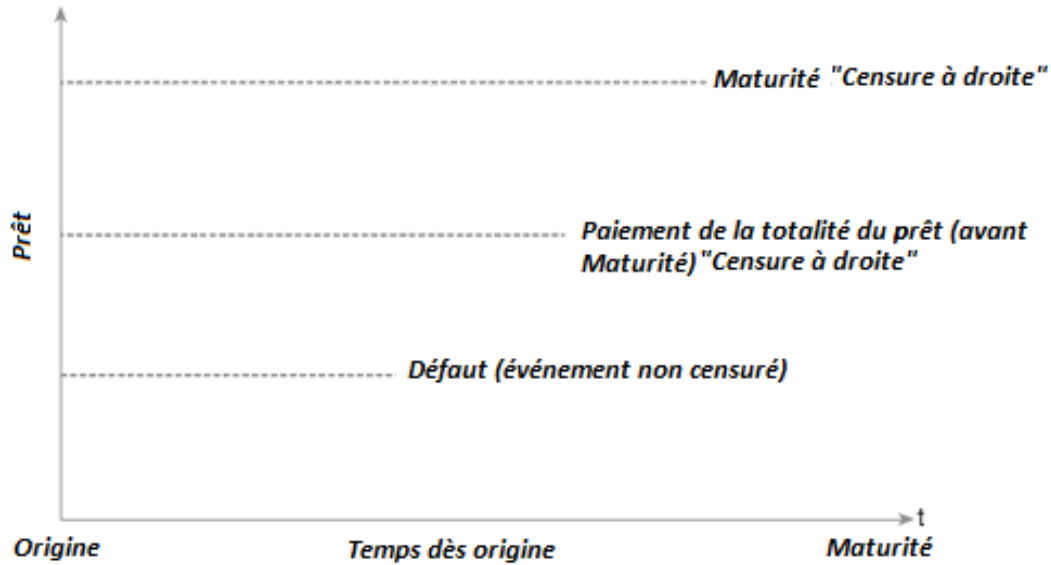


FIGURE 6.6 – Les différentes possibilités d'aboutissement d'un crédit : Défaut ou censure

$$\delta_{it} = 1_{\{T_i \leq t\}} = \begin{cases} 1 & \text{si } T_i \leq t, \text{ i.e défaut} \\ 0 & \text{sinon, i.e censure} \end{cases}$$

En général, les fonctions suivantes représentent la variable aléatoire T pour un client i :

- $F_i(t) = \int_{-\infty}^t f_i(u) du$
- La fonction de survie $S_i(t) = 1 - F_i(t)$

6.4 Application à notre portefeuille

6.4.1 Traitement des données

Avant l'étude de notre base de données on est amené à effectuer des traitements concernant la variable de censure ainsi que le calcul du moment de défaut des clients en mois, on doit aussi calculer le temps en mois dès le moment d'observation initiale. Par conséquent, l'indicateur de défaut i.e *défaut* égale à 0 si une observation est censurée, et 1 si un défaut survient.

On évite la censure à gauche, en considérant que tous les prêts commencent dès notre point d'observation, intitulé *temps_obs_initiale*.

La tableau 6.7 montre les quatre premières observations de clients de notre portefeuille. La variable *temps_obser* correspond au temps de défaut du client dès notre point d'observation. Contrairement, aux données en panel utilisées lors du modèle discret, où

	id	temps	Prem_observ_temps	defaut_tps	Pay_av_Mat	temps_obser
1	1	48	25	1	0	24
2	2	26	25	0	1	2
3	3	29	25	0	1	5
4	4	60	25	0	0	36

FIGURE 6.7 – Table des observations après réorganisation des données panels

la variable *temps_obser* indique le temps absolu et où les prêts sont octroyés dans des instants différents.

Le premier client est tombé en défaut après 24 mois de notre point d'observation, le second et le troisième client quant à eux ont payé la totalité du prêt avant leurs maturités respectivement à 2, 5 mois de notre point d'observation, alors que le quatrième client est suivi jusqu'à la fin de notre observation i.e sans qu'il tombe en défaut ou être censuré.

Le 2ème, 3ème clients sont considérés des observations censurées à droite de notre tableau 6.7.

6.4.2 Modèle semi-paramétrique : Modèle de Cox

Le modèle de *Cox* est un modèle à risques proportionnels étudié par Cox (1972), il fait partie des modèles de régression qui relie le temps de survie aux covariables i.e variables explicatives. Par conséquent il est plus flexible comme modèle comparé à l'analyse de KM ou la méthode actuarielle. Le modèle de *Cox* est donné par :

$$\lambda(t | x_i) = \lambda_0(t) \exp(\beta' x_i)$$

où

β : Vecteur des paramètres estimés

x_i : Vecteur des covariables

λ_0 : risque de référence (baseline hazard)

Le phénomène que l'on se propose d'analyser est l'étude de la durée de survie d'un client c-à-d avant sa tombée en défaut, ou plus précisément comment la fonction de risque de tomber en défaut est influencée par des caractéristiques de l'économie (PIB, Tx chômage), du client *scor_octroi* et de l'affaire (*type produit*, *rapp_pret_val*).

La table de la sortie obtenue par SPSS indique qu'il y a 49982 événements dont 34833 censurés, 5352 observations manquantes.

Régression de Cox

[Ensemble_de_données6]

Récapitulatif du traitement des observations

		N	Pourcentage
Observations disponibles dans l'analyse	Événement ^a	15149	27,4%
	Censurée	34833	63,0%
	Total	49982	90,3%
Observations enlevées	Observations avec valeurs manquantes	5352	9,7%
	Observations avec durée négative	0	0,0%
	Observations censurées avant l'événement le plus ancien dans une strate	0	0,0%
	Total	5352	9,7%
Total		55334	100,0%

a. Variable dépendante : duree_default

FIGURE 6.8 – Tableau récapitulatif des données

Recodage des variables catégorielles

Comme dans la régression classique, les variables explicatives catégorielles sont admises mais doivent être recodées sous forme d'un ensemble de variables auxiliaires.

Codages des variables nominales^a

		Fréquence	(1)
Type_produit ^b	EDM	31170	1
	PP	18812	0

a. Variable nominale : Type_produit

b. Codage de paramètres Indicatrice

FIGURE 6.9 – Codage de la variable type produit

Le recodage est essentiel. En effet, si l'on introduit une variable catégorielle sans la recoder, elle sera traitée comme quantitative ce qui implique en particulier que d'une part l'on tient compte de l'ordre de ses modalités, ce qui n'a par exemple aucun sens pour des variables purement nominales comme l'état-civil ou le domaine d'activité, et d'autre part, que les intervalles de valeurs sont comparables ce qui signifie, si l'on considère par exemple les catégories célibataire, marié, veuf, divorcé, que la différence entre marié et célibataire équivaut à celle entre divorcé et veuf, et que l'écart entre divorcé et célibataire vaut 3 fois celui entre célibataire et marié. Ceci n'est évidemment pas défendable. C'est le cas de notre variable type du produit qui comporte deux modalités à savoir le produit **PP** pour Prêt personnel et le produit **EDM** pour produit Équipement de ménage. L'avantage

de la solution du recodage dans la procédure COXREG est que l'on obtient ainsi une statistique de Wald pour la variable catégorielle en plus de celles pour les variables indicatrices.

6.4.3 Estimation des paramètres du modèle

Évaluation du modèle

Les indications sur l'ajustement global du modèle sont données dans le tableau intitulé "Tests de spécification du modèle". On y trouve :

- le $-2LV$ qui correspond ici à $-2foislelogarithme$ du maximum atteint de la vraisemblance partielle.
- la statistique du khi-deux du Score Test
- la statistique du khi-deux du rapport de vraisemblance

L'ajustement du modèle est d'autant meilleur que sa vraisemblance est forte, et que donc son $-2LV$ est petit. Dans notre cas le meilleur modèle est trouvé à la 8ème étape la méthode utilisée est la méthode pas à pas de Wald tel que $-2LV(\text{Modèle}_{estimé}) = 293289$

Bloc 1 : Méthode = Ascendante pas à pas (Wald)

Tests de spécification du modèleⁱ

Etape	-2log-vraisemblance	Global (note)			Changement de l'étape précédente			Changement du bloc précédent		
		Khi-deux	ddl	Signif.	Khi-deux	ddl	Signif.	Khi-deux	ddl	Signif.
1 ^a	299779,702	5582,691	1	,000	3761,912	1	,000	3761,912	1	,000
2 ^b	299262,123	8441,238	2	,000	517,579	1	,000	4279,491	2	,000
3 ^c	296118,684	10961,178	3	,000	3143,439	1	,000	7422,930	3	,000
4 ^d	293970,426	12182,485	4	,000	2148,258	1	,000	9571,188	4	,000
5 ^e	293306,565	13273,122	5	,000	663,861	1	,000	10235,050	5	,000
6 ^f	293299,423	13273,160	6	,000	7,142	1	,008	10242,191	6	,000
7 ^g	293293,840	13273,362	7	,000	5,583	1	,018	10247,774	7	,000
8 ^h	293289,798	13274,907	8	,000	4,042	1	,044	10251,816	8	,000

- a. Variable(s) entrée(s) à l'étape numéro 1: Rapp_pret_val
 b. Variable(s) entrée(s) à l'étape numéro 2: Tx_Interet
 c. Variable(s) entrée(s) à l'étape numéro 3: PIB
 d. Variable(s) entrée(s) à l'étape numéro 4: Scor_Octroi
 e. Variable(s) entrée(s) à l'étape numéro 5: Tx_chom
 f. Variable(s) entrée(s) à l'étape numéro 6: Type_produit
 g. Variable(s) entrée(s) à l'étape numéro 7: AGE_CL
 h. Variable(s) entrée(s) à l'étape numéro 8: Anc_Emploi
 i. Bloc de départ numéro 1. Méthode = Ascendante pas à pas (Wald)

FIGURE 6.10 – Méthode ascendante pas à pas

Sélection du modèle

Comme indiqué précédemment, le meilleur modèle est trouvé à la 8ème étape, L'estimation proprement dite du modèle est donnée dans le tableau 'Variables dans l'équation'. Sous B on y lit le coefficient estimé de chaque facteur explicatif. Le modèle ainsi

trouvé comprend 8 variables explicatives de la durée avant que le client tombe en défaut. Les variables montant du prêt et le nombre crédit ne sont pas prises en compte par le modèle estimé.

Variables dans l'équation

		B	E.S.	Wald	ddl	Signif.	Exp(B)
Étape 1	Rapp_pret_val	,011	,000	9785,830	1	,000	1,011
Étape 2	Rapp_pret_val	,011	,000	9859,799	1	,000	1,011
	Tx_Interet	,017	,000	1443,972	1	,000	1,017
Étape 3	Rapp_pret_val	,011	,000	5681,676	1	,000	1,011
	Tx_Interet	,017	,000	1231,068	1	,000	1,017
	PIB	-,197	,003	3530,483	1	,000	,821
Étape 4	Rapp_pret_val	,012	,000	6070,014	1	,000	1,012
	Tx_Interet	,016	,001	813,450	1	,000	1,016
	PIB	-,200	,003	3617,758	1	,000	,819
	Scor_Octroi	-,005	,000	2180,713	1	,000	,995
Étape 5	Rapp_pret_val	,012	,000	7974,988	1	,000	1,012
	Tx_Interet	,016	,001	783,958	1	,000	1,017
	PIB	-,253	,004	4001,832	1	,000	,777
	Tx_chom	-,151	,006	627,107	1	,000	,860
	Scor_Octroi	-,005	,000	2028,535	1	,000	,995
Étape 6	Rapp_pret_val	,012	,000	7916,674	1	,000	1,012
	Tx_Interet	,016	,001	772,698	1	,000	1,016
	PIB	-,252	,004	3993,892	1	,000	,777
	Tx_chom	-,151	,006	629,241	1	,000	,860
	Scor_Octroi	-,005	,000	2018,102	1	,000	,995
	Type_produit	-,046	,017	7,167	1	,007	,955
Étape 7	Rapp_pret_val	,012	,000	7730,056	1	,000	1,012
	Tx_Interet	,016	,001	777,882	1	,000	1,017
	PIB	-,252	,004	3988,758	1	,000	,777
	Tx_chom	-,151	,006	632,296	1	,000	,860
	Scor_Octroi	-,005	,000	2013,757	1	,000	,995
	Type_produit	-,046	,017	7,200	1	,007	,955
	AGE_CL	-,002	,001	5,578	1	,018	,998
Étape 8	Rapp_pret_val	,012	,000	7693,644	1	,000	1,012
	Tx_Interet	,016	,001	775,125	1	,000	1,017
	PIB	-,252	,004	3987,497	1	,000	,777
	Tx_chom	-,152	,006	633,683	1	,000	,859
	Scor_Octroi	-,005	,000	2012,348	1	,000	,995
	Type_produit	-,046	,017	7,279	1	,007	,955
	AGE_CL	-,002	,001	5,404	1	,020	,998
	Anc_Emploi	-,002	,001	4,021	1	,045	,998

FIGURE 6.11 – Estimation des paramètres du modèle de Cox

6.4.4 Estimation de la fonction de survie

Pour un modèle à risque proportionnel, la fonction de survie $S(t, x)$ d'un individu avec profil x se déduit de la fonction de survie $S_0(t)$, correspondant au profil virtuel de référence $x=0$ par la relation :

$$S(t, x) = S_0(t) \exp(x'\beta)$$

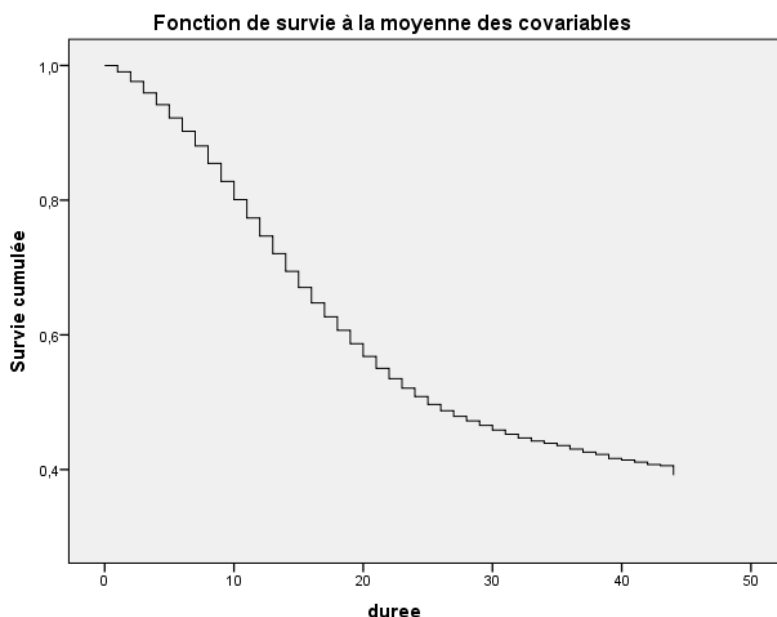


FIGURE 6.12 – Courbe de survie

La courbe de survie de base ainsi obtenue en sortie est un affichage visuel de la durée de survie pour un client dont les caractéristiques sont les moyennes de chaque variable avant de tomber en défaut prédite par le modèle . L'axe horizontal indique la durée jusqu'à l'événement. L'axe vertical indique la probabilité de survie. Ainsi, tous les points de la courbe de survie indiquent la probabilité que le client "moyen" ne tombe pas en défaut durant cette durée passée. Après 30 mois, la courbe de survie devient plus régulière. Plus de clients sont tombés en défaut donc plus d'informations disponibles sur le défaut , ce

qui crée une courbe plus lisse contrairement en dents de scie au début de l'observation. Ainsi un client dont les caractéristiques sont les moyennes de variables explicatives va faire défaut avec une probabilité 0.6 dans les 40 mois qui suivent notre observation.

6.4.5 Modèle stratifié

COXREG permet d'estimer des modèles de Cox stratifiés. Un modèle stratifié est un modèle où le risque de référence (baseline hazard) est propre à chaque strate. Il ne comporte cependant qu'un jeu de coefficients β qui sont supposés être les mêmes pour toutes les strates.

Lorsqu'on a précisé une variable de stratification, le modèle de Cox est donc estimé en admettant un risque de référence $h_{s,0}(t)$ différent pour chaque strate s . L'hypothèse générale de proportionnalité est alors remplacée par les deux hypothèses suivantes :

1. Risques proportionnels pour individus d'une même strate.
2. Effets (β) des covariables indépendants de la strate.

Considérons dans notre cas l'estimation de notre modèle avec une stratification selon le type de produit (prêt personnel PP, Équipement de ménage EDM).

On remarque que par rapport au modèle sans stratification, on gagne un degré de li-

Statut de strate^a

Strate	Événement	Censurée	Pourcentage de censurés
PP	5711	13101	69,6%
EDM	9438	21732	69,7%
Total	15149	34833	69,7%

a. La variable de strate est : Type_produit

FIGURE 6.13 – Nombre événement par produit

berté puisque l'on estime un paramètre de moins, celui de type_produit. Il est naturel alors que les statistiques du khi-deux soient plus faibles. Elles restent cependant toujours aussi significatives. Bien que l'on estime un seul jeu de coefficients β pour les deux strates, les fonctions de survie $S_s(t, x)$ sont, elles, différentes puisque l'on postule des fonctions risques $h_{s,0}(t)$ de référence propres à chaque strate s . D'après la figure 6.15 on voit bien que les clients qui ont un crédit du type prêt personnel **PP** ont tendance à tomber en défaut avant ceux qui ont contracté un crédit de type équipement de ménage **EDM**. cette conclusion va dans le même sens que l'interprétation de -0.046 de l'exponentiel du coefficient de la variable **type_produit** nous dit en effet que le risque de tomber en défaut est de 0.63 fois plus important pour un client qui a contracté un prêt personnel comparé à

Tests de spécification du modèle^a

-2log-vraisemblance	Global (note)			Changement de l'étape précédente			Changement du bloc précédent		
	Khi-deux	ddl	Signif.	Khi-deux	ddl	Signif.	Khi-deux	ddl	Signif.
273235,362	13256,845	9	,000	10239,183	9	,000	10239,183	9	,000

a. Bloc de départ numéro 1. Méthode = Entrée

Variables dans l'équation

	B	E.S.	Wald	ddl	Signif.	Exp(B)
Rapp_pret_val	,012	,000	7599,101	1	,000	1,012
Tx_Interet	,016	,001	762,322	1	,000	1,016
PIB	-,252	,004	3981,265	1	,000	,777
Tx_chom	-,152	,006	634,376	1	,000	,859
Scor_Octroi	-,005	,000	2010,502	1	,000	,995
AGE_CL	-,002	,001	5,402	1	,020	,998
Anc_Emploi	-,002	,001	3,790	1	,052	,998
MT_PRET	,000	,000	,274	1	,601	1,000
Nombr_Cred	,000	,001	,022	1	,883	1,000

FIGURE 6.14 – Estimation du modèle de Cox avec stratification par type du produit

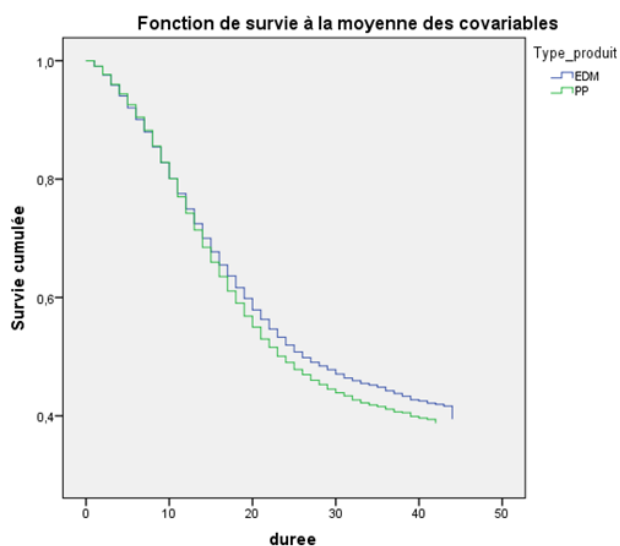


FIGURE 6.15 – Courbes de survie stratifiées selon le type de produit

un prêt de type EDM.

6.4.6 Test de proportionnalité

Le modèle semi paramétrique de Cox est très général puisqu'il ne suppose aucune hypothèse sur la distribution des durées des épisodes. Il suppose cependant la proportionnalité des risques. Il convient donc de vérifier que cette hypothèse est raisonnable. Une première approche consiste à vérifier la proportionnalité graphiquement. Une seconde repose sur des tests statistiques. Les tests graphiques examinent les effets, c'est-à-dire les covariables introduites dans le modèle, individuellement. Les tests statistiques portent également sur les effets individuels.

Tests graphiques

Une première évaluation graphique exploite une propriété des courbes $\ln(-\ln S)$ où S est la fonction de survie. Ces courbes sont appelées **LML** pour **Log Minus Log**. Sous l'hypothèse de proportionnalité, on a en effet l'équivalence

$$S(t, x) = S_0^{exp(x'\beta)} \Leftrightarrow \ln(-\ln S(t, x)) = \ln(-\ln S_0(t) + x'\beta)$$

Selon la partie de droite, les courbes **LML** sont parallèles, pour deux profils x_1 et x_2 , la différence entre les courbe **LML**, est tout t, égale à $(x_2 - x_1)'\beta$ est donc indépendante du temps.

Le test graphique consiste alors à tracer les courbes LML pour chaque valeur d'une covariable, les autres étant maintenues constantes, et de vérifier si cette translation se vérifie. Dans notre étude, on a testé si les courbes du type produit (**PP,EDM**) vérifier ce parallélisme, et effectivement d'après la figure à gauche de 6.16 on peut affirmer que les courbes LML des produits **PP** et **EDM** sont parallèles, par contre la proportionnalité est peu affirmée dans le cas de la variables **NINQ : Nombre de recherche** surtout pour les valeurs 0 et 12

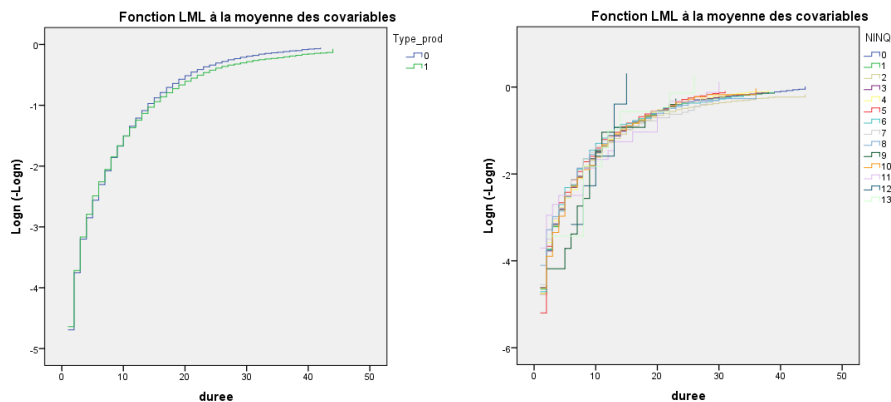


FIGURE 6.16 – Comparaison des courbes $\log(-\log S)$

Analyse des résidus

Les résidus partiels de **Schoenfeld** du modèle de Cox permettent de confirmer l'hypothèse de proportionnalité. Les résidus sont calculés pour les cas non censurés seulement. Un résidu pour chaque coefficient β_j .

Ces résidus partiels ne sont pas des écarts entre valeurs observées et prédites de la variable réponse. estimé.

Dans notre cas on a calculé ces résidus par l'outil **XLSTAT** pour les variables type de produit et score d'octroi et la variable taux de chômage. On remarque que les résidus sont distribués sur les deux cotés de l'axe X donc on peut affirmer le test de proportionnalité pour les variables score d'octroi et taux de chômage.

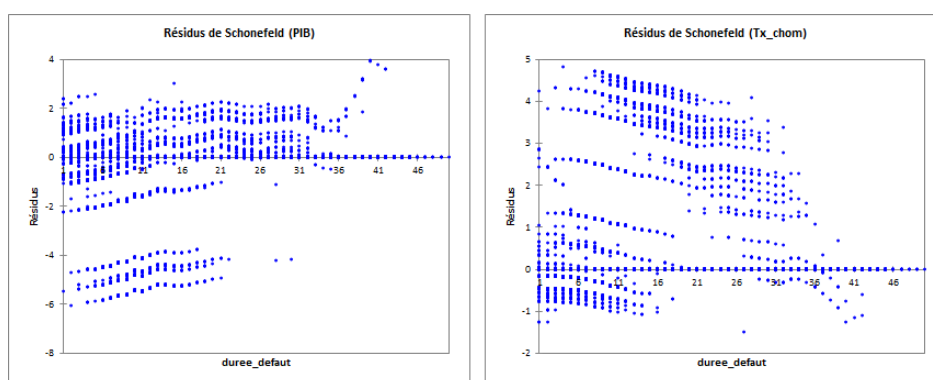


FIGURE 6.17 – Résidus partiels : Score d'octroi et taux de chômage

6.4.7 Validation du modèle de Cox

Finalement, pour valider le modèle de Cox on est amené à mesurer sa performance via la courbe ROC, pour cela on a calculé la probabilité de défaut en se basant sur la fonction de survie :

$$PD(t) = 1 - S(t)$$

D'après le tableau ci-dessous la zone sous la courbe est de 65%, donc on peut considérer que notre modèle est moyennement bon.

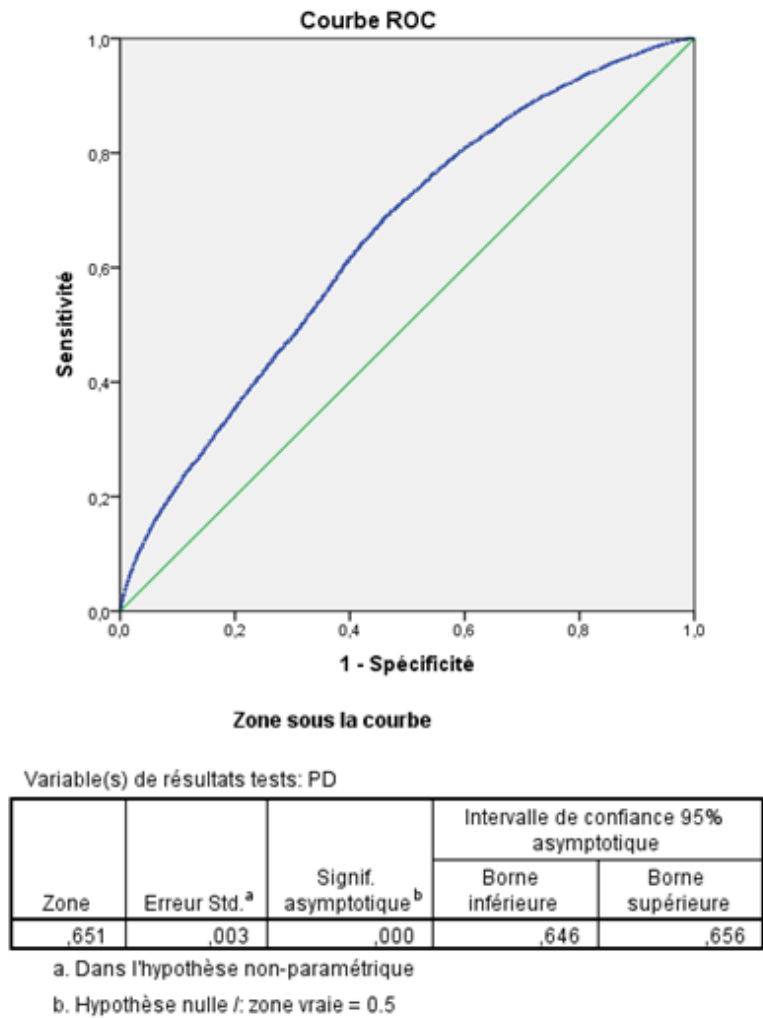


FIGURE 6.18 – Courbe Roc du modèle de Cox

6.5 Conclusion

Dans la pratique, les modèles discrets de survie (en particulier les modèles logit et cloglog) sont beaucoup plus utilisés que les modèles continus de survie. Ces modèles estiment la probabilité de défaut pour une périodicité des données donnée. Récemment des progrès dans la régulation, en particulier les provisions pour perte dû au risque crédit sous les normes IFRS.9 requies des estimations sur toute la durée de vie du crédit (i.e) Les modèles à temps continu dans l'analyse de la survie permettent d'estimer d'une manière beaucoup plus poussée le risque de défaut, mais comme mentionné précédemment, leur coût et leur implémentation est leur défaut majeur, ainsi que la perte de transparence, cela dit, ils sont utilisés pour des études avancées dans le domaine du risque crédit.

Chapitre 7

Le modèle Bayésien pour le calcul de la probabilité de défaut

Sommaire

7.1 Le modèle bayésien	66
7.1.1 Généralités	66
7.1.2 Simulation de Monte Carlo	66
7.1.3 MCMC	66
7.2 Application à notre base de données	67
7.2.1 Estimation de la probabilité de défaut avec l'approche Bayésienne	67
7.2.2 Comparaison entre le modèle classique et le modèle bayésien . . .	71

7.1 Le modèle bayésien

7.1.1 Généralités

Le modèle bayésien est une méthode statistique qui permet de faire une inférence sur des quantités d'intérêt (peuvent être des paramètres d'un modèle, des données manquantes ou des prédictions) en combinant un ensemble d'informations connues sur la base de données observées. Généralement le modèle bayésien se compose de deux niveaux : un modèle statistique qui décrit la distribution des données (\mathbf{y}) en l'occurrence la probabilité de défaut **PD** en se donnant des quantités inconnus, et une distribution à priori qui décrit les quantités inconnus indépendamment de notre base de données. Généralement, les quantités inconnus d'intérêt sont les paramètres θ dans notre modèle, et le modèle statistique c'est la fonction de vraisemblance $L(\theta; \mathbf{y})$, après la collecte des données, on peut actualiser nos prédictions avancées sur le paramètre θ et calculer la distribution à posteriori $p(\theta|\mathbf{y})$, le passage de la distribution à priori à la distribution à posteriori est calculée par le théorème de Bayes :

$$p(\theta|\mathbf{y}) \propto f(\theta|\mathbf{y}) = L(\theta; \mathbf{y})\pi(\theta)$$

où $f(\theta|\mathbf{y})$ distribution a posteriori et $\pi(\theta)$ densité a priori de θ .

Le tableau suivant détaille la différence entre l'approche bayésienne et l'approche classique.

	L'approche fréquence	L'approche bayésienne
Probabilité	Objective	Subjective
Base de données	Un échantillon aléatoire	Un échantillon réalisé
Critère d'estimation	Non biais	Minimisation de la fonction de perte
Intervalle	Intervalle de confiance	Intervalle de crédibilité
test d'hypothèse	Via p-Value	Via la vraisemblance marginale

7.1.2 Simulation de Monte Carlo

c'est une techniques d'estimation s'appuyant sur la simulation d'un grand nombre de variables aléatoires, ses avantages principaux sont domaine d'application très vaste, peu d'hypothèses de mise en œuvre et facile à implémenter.

7.1.3 MCMC

La méthode MCMC a pour but échantillonner selon une loi cible de densité $\pi(z)$ connue à une constante multiplicative près

$$\pi = \text{loi}(Z)$$

où $Z = Z_{1:n}$ est une variable réelle n-dimensionnelle

Principe

- On construit une chaîne de Markov $(Z_{k \geq 0}^k)$ sur \mathbb{R}^n dont la loi limite est π
- Lorsque la chaîne devienne stationnaire, on extrait un échantillon de la chaîne (Z^{k1}, \dots, Z^{kN}) pour estimer les quantités d'intérêt : c'est le théorème ergodique
- proposer une approximation de la mesure cible : $\pi(f) = \int_E f(z)\pi(dz)$ où f est une fonction réelle mesurable et π intégrable

7.2 Application à notre base de données

7.2.1 Estimation de la probabilité de défaut avec l'approche Bayésienne

Tout d'abord on commence par l'introduction de l'approche bayésienne dans l'estimation de la probabilité de défaut **PD**.

Comme l'approche bayésienne nécessite généralement l'évaluation de la distribution bayésienne à posteriori avec la **Méthode de Monte-Carlo par chaînes de Markov (MCMC)**.

La première table de notre sortie de la procédure **MCMC** montre le nombre d'observation de notre échantillon et la méthode d'échantillonnage appliquée ainsi que la distribution simulée a priori de nos paramètres.

Le tableau des *agrégations à posteriori* figure 7.2 rapporte quant à lui la moyenne a posteriori, l'écart type ainsi que 1er, 2ème et 3ème centiles d'intérêt. On remarque que les moyennes des β ont les signes attendus et leurs valeurs sont similaires à ceux du modèle fréquentiste (après calibration), notons aussi que les écarts types sont similaires aux erreurs standards.

Procédure MCMC				
Nb d'observations lues		6000		
Nb d'observations utilisées		5996		
Paramètres				
Bloc	Paramètre	Méthode d'échantillonnage	Valeur initiale	Distribution a priori
1	beta0	N-Metropolis	0	normal(0,var=1000)
	beta1		0	normal(0,var=1000)
	beta2		0	normal(0,var=1000)
	beta3		0	normal(0,var=1000)

FIGURE 7.1 – Résumé des paramètres MCMC pour un modèle probit

Contrairement à l'approche fréquentiste, l'approche bayésienne calcule les intervalles de confiance directement des distributions a posteriori, qui sont exposés dans le tableau

Procédure MCMC						
Agrégations a posteriori						
Paramètre	N	Moyenne	Ecart-type	Centiles		
				25	50	75
beta0	5000	-0.1991	0.1938	-0.3310	-0.1990	-0.0684
beta1	5000	-0.0375	0.00279	-0.0394	-0.0375	-0.0356
beta2	5000	2.8109	0.1025	2.7424	2.8113	2.8804
beta3	5000	-13.9495	1.1272	-14.7403	-13.9639	-13.1763

FIGURE 7.2 – Procédure MCMC pour un modèle probit

ci-dessous (figure 7.3).

La matrice de corrélation des paramètres est donnée dans le tableau suivant 7.4, à l'ex-

Intervalles a posteriori					
Paramètre	Alpha	Intervalle de confiance symétrique		Intervalle de confiance à plus haute densité (HPD)	
		beta0	0.050	-0.5709	0.1798
beta1	0.050	-0.0431	-0.0324	-0.0429	-0.0323
beta2	0.050	2.6093	3.0055	2.6088	3.0041
beta3	0.050	-16.1359	-11.7387	-16.1579	-11.8040

FIGURE 7.3 – Procédure MCMC pour le modèle probit

ception de la constante β_0 , la matrice contient des valeurs de corrélation petites donc il n'y a pas de problème de corrélation entre les paramètres de l'échantillon.

Matrice de corrélation a posteriori				
Paramètre	beta0	beta1	beta2	beta3
beta0	1.0000	-0.8683	-0.3188	-0.3577
beta1	-0.8683	1.0000	-0.1766	0.1106
beta2	-0.3188	-0.1766	1.0000	0.3693
beta3	-0.3577	0.1106	0.3693	1.0000

FIGURE 7.4 – Procédure MCMC pour le modèle probit

Afin de vérifier la convergence de la simulation de Monte Carlo, les auto-corrélations sont importantes comme montrer dans le tableau 7.6, l'existence d'une importante auto-corrélation pour le petits lags et décroît vite pour les longs lags affirme qu'il y a convergence de la simulation c'est ce qui est souhaité (propriété de convergence de la simulation

Matrice de covariance a posteriori				
Paramètre	beta0	beta1	beta2	beta3
beta0	0.0375	-0.00047	-0.00633	-0.0781
beta1	-0.00047	7.775E-6	-0.00005	0.000348
beta2	-0.00633	-0.00005	0.0105	0.0426
beta3	-0.0781	0.000348	0.0426	1.2705

FIGURE 7.5 – Procédure MCMC pour le modèle probit

est vérifiée).

Pour plus de détail sur la convergence, on a calculé quelques tests à savoir ceux de **Ge-**

Autocorrélations a posteriori				
Paramètre	Lag 1	Lag 5	Lag 10	Lag 50
beta0	0.7197	0.1635	0.0225	-0.0191
beta1	0.7345	0.1962	0.0504	-0.0299
beta2	0.7404	0.2277	0.0150	0.0325
beta3	0.7493	0.2148	0.0105	0.0208

FIGURE 7.6 – Procédure MCMC pour le modèle probit

weke, Raftery lewis, concernant **Geweke**, on compare les moyennes de début et de fin de la chaîne de Markov le test est comme suit tableau 7.7 :

H_0 : la moyenne estimée converge

H_1 : la moyenne estimée ne converge pas

Pour des grandes valeurs de z on rejette H_0 . Dans notre exemple la convergence est satisfaite.

Le test de **Raftery-Lewis** évalue la précision des centiles estimés (désirés) en rapportant le nombre d'échantillon pour atteindre la précision souhaitée. L'échec signifie qu'on a besoin d'une **chaîne de Markov** plus longue. Mais si le total du nombre d'échantillon n'excède le nombre minimal il y a rejet de convergence sinon si le nombre d'échantillon est supérieur au nombre minimal donc la convergence est satisfaite c'est le cas de notre tableau A.4.

Le test de **Heidelberger-Welch** est semblable au test de **Raftery-Lewis**, à la seule différence que le test de **Heidelberger-Welch** est divisé en deux sous tests, un test stationnaire et un test de demi-largeur, les deux tests ony pour objectifs d'évaluer la stationnarité de la covariance de la chaîne de Markov. L'échec signifie qu'on a besoin d'une chaîne de Markov plus longue. Dans notre cas (Voir **Annexe**), en se basant sur la statistique de **Cramer Von Mises**, le p -value est supérieur à la statistique de **Cramer Von Mises** donc la covariance de notre chaîne de Markov est stationnaire.

Diagnostics de Geweke		
Paramètre	z	Pr > z
beta0	0.4379	0.6614
beta1	-0.8870	0.3751
beta2	1.1114	0.2664
beta3	0.0715	0.9430

Diagnostics de Raftery et Lewis				
Quantile=0.025 Exactitude=+/-0.005 Probabilité=0.95 Epsilon=0.001				
Paramètre	Nombre d'échantillons			Facteur de dépendance
	Ignoré	Total	Minimum	
beta0	13	13499	3746	3.6036
beta1	14	14906	3746	3.9792
beta2	13	14429	3746	3.8518
beta3	12	12954	3746	3.4581

FIGURE 7.7 – Procédure MCMC pour le modèle probit

Pour le test de demi-largeur, il informe si on aura besoin d'une taille d'échantillon plus grande pour atteindre la précision requise pour la moyenne estimée. Dans notre cas l'échantillon est suffisant.

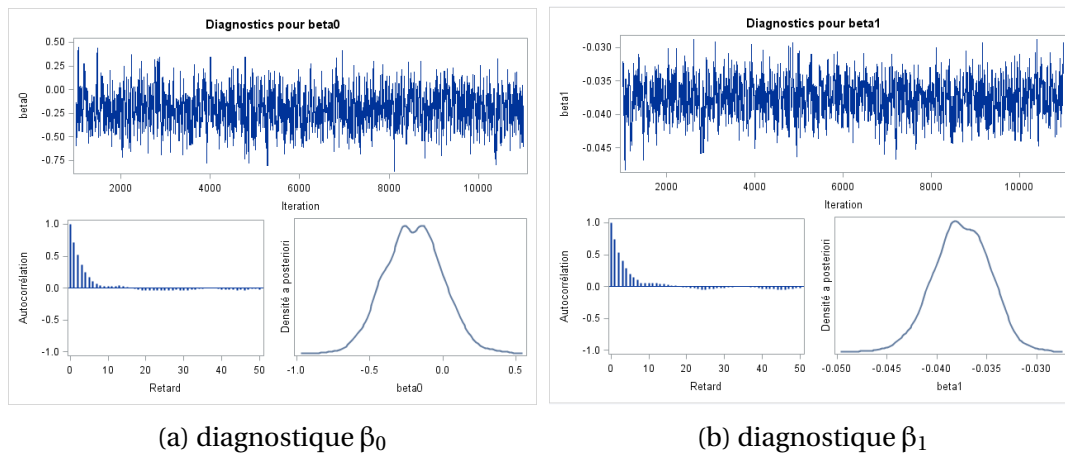


FIGURE 7.8 – Diagnostic des paramètres β_0 et β_1

En plus des statistiques, PROC MCMC affiche des graphes de diagnostics 7.8. Pour chaque paramètre l'historique de la simulation est affichée en haut, si le graphe ressemble à un **bruit blanc sans drift** comme le cas de notre exemple, il indique une bonne convergence, la figure du bas à gauche montre les auto-corrélations jusqu'au lag 50, comme déjà vu précédemment et dans les tableaux plus haut les autocorrélations sont longues pour les petits lags et petites pour les longues lags, la propriété souhaité est donc vérifiée. Fina-

lement la figure de bas à gauche est la distribution a posteriori simulée pour chacune des paramètres qui suivent une loi normale.

7.2.2 Comparaison entre le modèle classique et le modèle bayésien

Le modèle bayésien est plus adapté pour des portefeuilles comportant un nombre de clients important puisqu'il permet de travailler sur un échantillon puis d'inférer sur toute la population, donc c'est une méthode rapide et moins coûteuse en terme du temps cependant son problème majeur est de l'implémenter puisqu'il nécessite la programmation de la technique de simulation de Monte Carlo par la Chaîne de Markov contrairement au modèle classique qui est simple comme on a déjà vu plus haut.

Chapitre 8

Stress Testing

Sommaire

8.1 Objectif du Stress Tests	74
8.2 Les types du Stress Testing	74
8.2.1 Stress Testing basé sur les scénarios	75
8.3 Le modèle risque sous la régulation Bâle	76
8.3.1 Pire scénario du taux de défaut	76
8.4 Stress Test de la probabilité de défaut basé sur les scénarios	79

Introduction

Dans ce chapitre, on va aborder l'étape du *Stress Testing*, le Stress Testing est une activité qui est généralement effectuée après que les modèles de *PD*, *LGD* et *EAD* sont élaborés. Son objectif est d'analyser le comportement de ces modèles sous certaines variations ou changements internes ou externes. On va discuter plusieurs manières d'élaborer un Stress Test comme celui basé sur la sensibilité ou encore celui basé sur les scénarios historiques ou hypothétiques.

8.1 Objectif du Stress Tests

La question qu'on se pose généralement, Pourquoi a-t-on besoin des Stress tests? La réponse c'est que les Stress tests sont imposés par les régulateurs, en l'occurrence par le comité Bâle. Cependant, de plus les Stress Tests sont d'une valeur capitale d'un point de vue management, premièrement ils permettent d'anticiper les scénarios les plus pesants en terme de capital.

8.2 Les types du Stress Testing

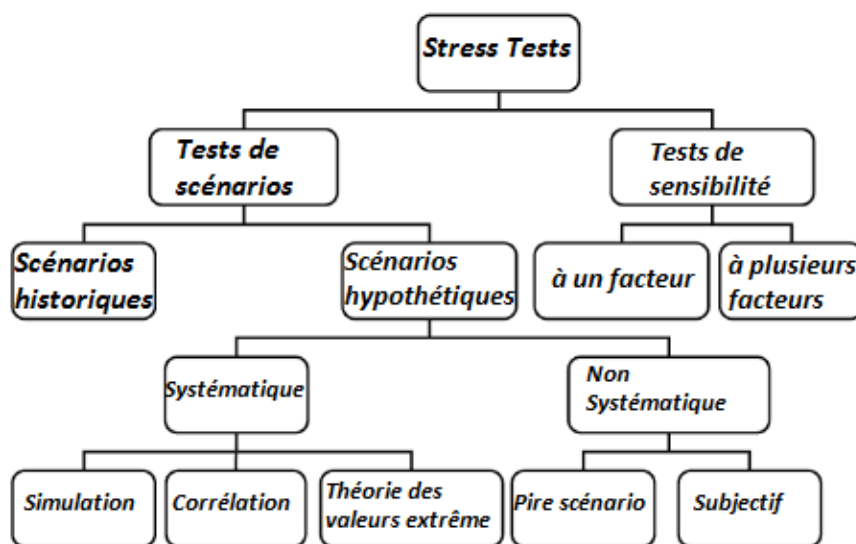


FIGURE 8.1 – Les différents type du Stress testing

Sur la figure 8.1 on peut voir les différents approches du stress testing, qui peuvent généralement être divisées en deux grands niveaux :

- **Tests de scénarios** : comportent *les scénarios historiques* qui sont des scénarios qui ont déjà eu lieu dans le passé et qui ont engendrés de lourdes pertes financières

(exemple : la crise économique de 1929 ou la crise financière de 2007) ou des *scénarios hypothétiques* qui sont des scénarios imaginaires mais ils ont une probabilité de se réaliser et comportent à leur tour des simulations comme des scénarios systématiques ou bien le pire scénario comme non systématique.

- **Tests de sensibilité** sont des tests qui se basent sur l'étude de variation d'un facteur ou plusieurs facteurs simultanément (exemple : hausse de salaire, changement de l'état matrimonial du client).

8.2.1 Stress Testing basé sur les scénarios

Dans cette approche, le stress test est basé sur différents scénarios. L'idée est de travailler sur des événements de choc possible puis jauger leurs impact sur les différentes variables affectées. La figure 8.2 expose les trois scénarios possibles (hausse, normale ou baisse) du taux de croissance du PIB, taux d'inflation ou taux de chômage.

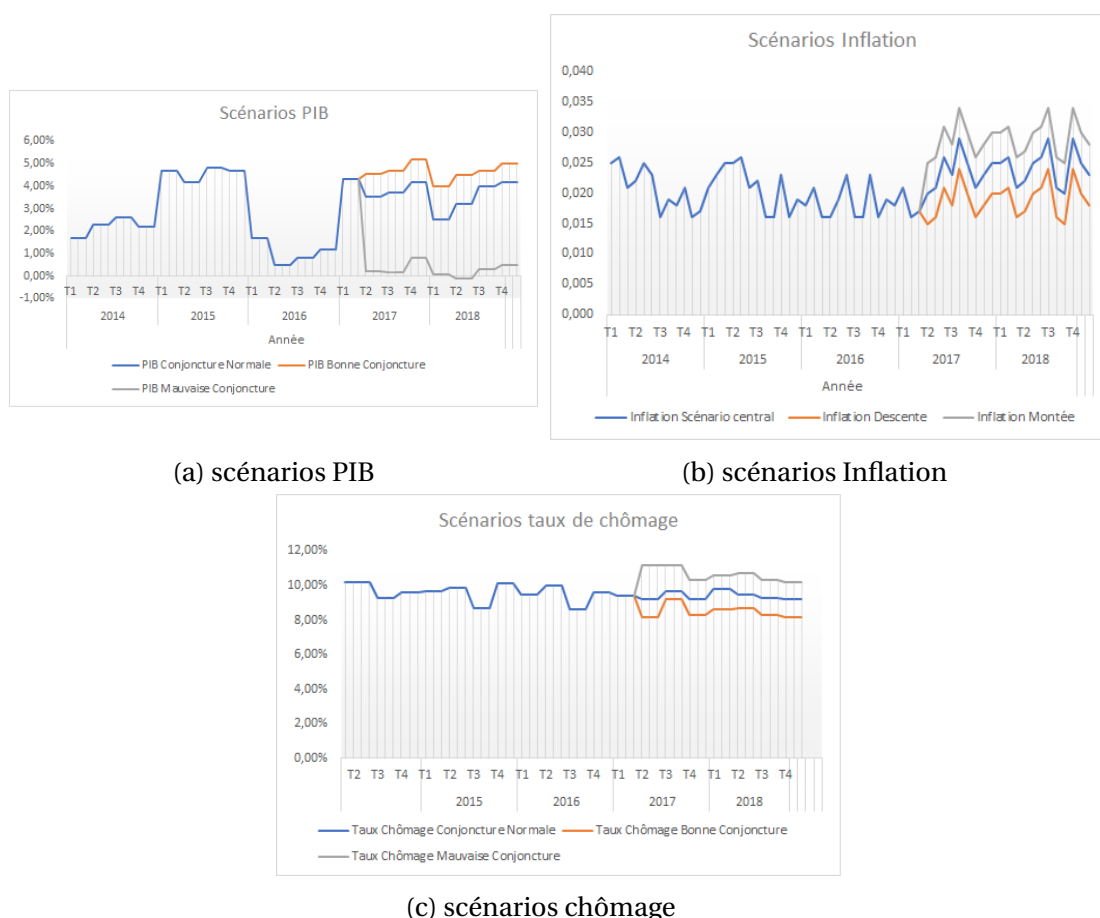


FIGURE 8.2 – Les différents scénarios pour les différents variables macroéconomiques : Cas du Maroc sur la période T2 de 2017 à T4 2018 source :Banque Mondiale

8.3 Le modèle risque sous la régulation Bâle

Le comité de Bâle veut que les institutions financières aient assez de fonds propres (Tiers I et Tiers II) pour faire face aux pires scénarios de pertes dans leurs portefeuilles de crédit.

Le stress de la probabilité de défaut PD est basé sur un modèle non linéaire d'un facteur (One Factor) où le facteur égale à 99.9

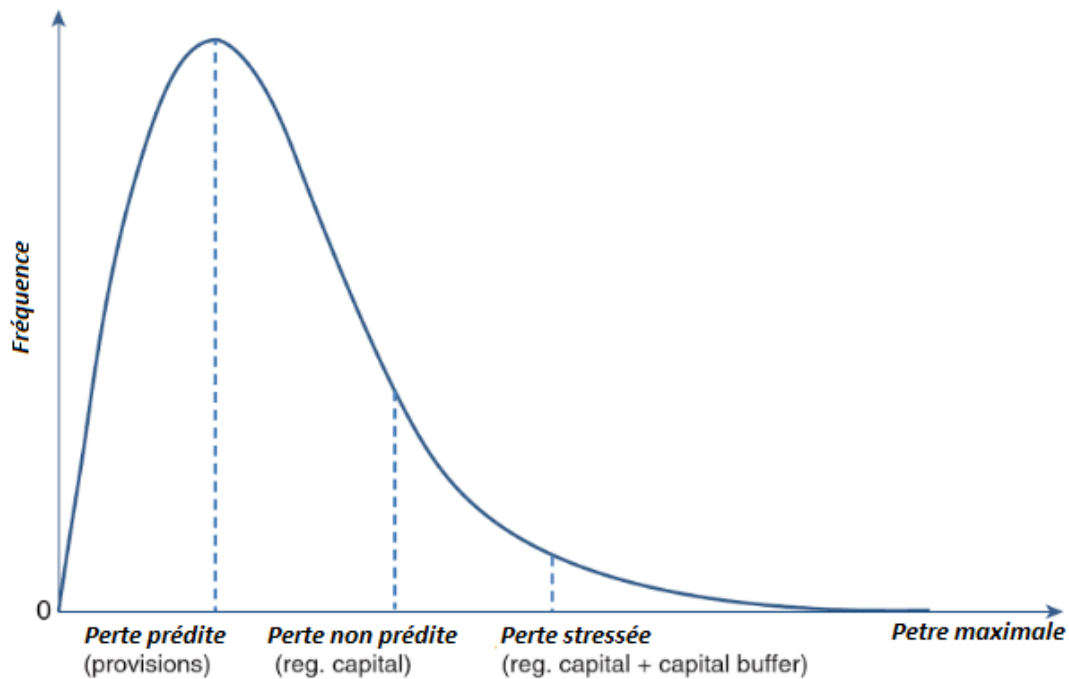


FIGURE 8.3 – Perte prédite, non prédite et stressée et leurs capitaux correspondants

8.3.1 Pire scénario du taux de défaut

Le pire scénario du taux de défaut (i.e the Worst-Case Default Rate WCDR) est un élément central dans le modèle de risque bâlois et aussi un exemple de stress test pour la probabilité de défaut. On se réfère au modèle de défaut présent dans notre portefeuille de crédit. Le modèle bâlois est défini comme suit :

$$R_{it} = -\sqrt{\rho}X_t + \sqrt{1-\rho}\epsilon_{it}$$

La sensibilité au facteur systémique est $\sqrt{\rho}$. On a choisi un signe négatif en respect avec la formule de Bâle mais dans ce cas il est Irrévérencieux puisque X_t suit une loi normale. La sensibilité du facteur Idiosyncratique quant à elle est $\sqrt{1-\rho}$. L'événement de défaut se réalise si le rendement tombe en dessous d'un seuil spécifié c_{it} . La probabilité de défaut PD dans le cas où X_t suit une loi normale est :

$$PD_{it} = P(D_{it} = 1) = P(R_{it} < c_{it}) = \Phi(c_{it})$$

Φ est fonction de densité cumulative de la loi normale centrale réduite. Cette probabilité correspond à une probabilité de défaut inconditionnelle concernant le facteur de risque systémique non observé X_{it} et elle est estimée par les modèles de durée de vie comme on a déjà vu au chapitre précédent. La probabilité de défaut conditionnelle qui a pour réalisation le facteur risque systémique est :

$$WCDR = P(D_{it} = 1 | x_t) = P(R_{it} < c_{it|X_t}) = \Phi\left(\frac{c_{it} + \sqrt{\varrho} X_t}{\sqrt{1 - \varrho}}\right) \quad (*)$$

Avec $c_{it} = \Phi^{-1}(PD_{it})$. Le comité de Bâle a spécifié que la corrélation entre les actifs et la réalisation x_t de la variable aléatoire X_t considérée comme le pire scénario entre 1000 scénarios économiques (i.e, Comme X_t est une normale centrée réduite : $x_t = \Phi^{-1}(0.999)$). On remplace dans (*), on trouve l'équation bien connue comme résultat au Worst-Case-Default-Rate (pire scénario du taux de défaut) sous l'approche IRB :

$$WCDR = P(D_{it} = 1 | x_t) = P(R_{it} < c_{it|X_t}) = \Phi\left(\frac{\Phi^{-1}(PD_{it}) + \sqrt{\varrho} \Phi^{-1}(0.999)}{\sqrt{1 - \varrho}}\right)$$

Les régulations prudentielles de Bâle ne permettent pas aux institutions financières d'estimer le paramètre de corrélation des actifs, par contre ils ont spécifié une constante à ce paramètre (15% dans le cas hypothécaire et 4% pour les crédits revolving (Permanents)) ou une fonction de la probabilité de défaut dans le cas des crédits pour entreprises et pour les autres crédits de détail.

Dans le cas du crédit de détail le paramètre de corrélation des actifs est calculé comme suit :

$$\varrho = 0.03 \times \frac{1 - \exp(-35PD)}{1 - \exp(-35)} + 0.16 \times \left(1 - \frac{1 - \exp(-35PD)}{1 - \exp(-35)}\right) \cong 0.03(1 + \exp(-35PD))$$

On peut remarquer que le paramètre de corrélation est une fonction décroissante de la probabilité de défaut qui est égale à 16% dans le cas où $PD=0$ et 3% dans le cas où $PD=1$.

Dans le cas des crédits pour entreprises le paramètre de corrélation des actifs est calculé comme suit :

$$\varrho = 0.12 \times \frac{1 - \exp(-50PD)}{1 - \exp(-50)} + 0.24 \times \left(1 - \frac{1 - \exp(-50PD)}{1 - \exp(-50)}\right) \cong 0.12(1 + \exp(-50PD))$$

De même que dans le cas du crédit de détail pour les ménages, le paramètre de corrélation dans le cas du crédit pour entreprises est une fonction décroissante de la probabilité de défaut qui est égale à 24% dans le cas où $PD=0$ et 12% dans le cas où $PD=1$, que la figure 8.4 illustre clairement.

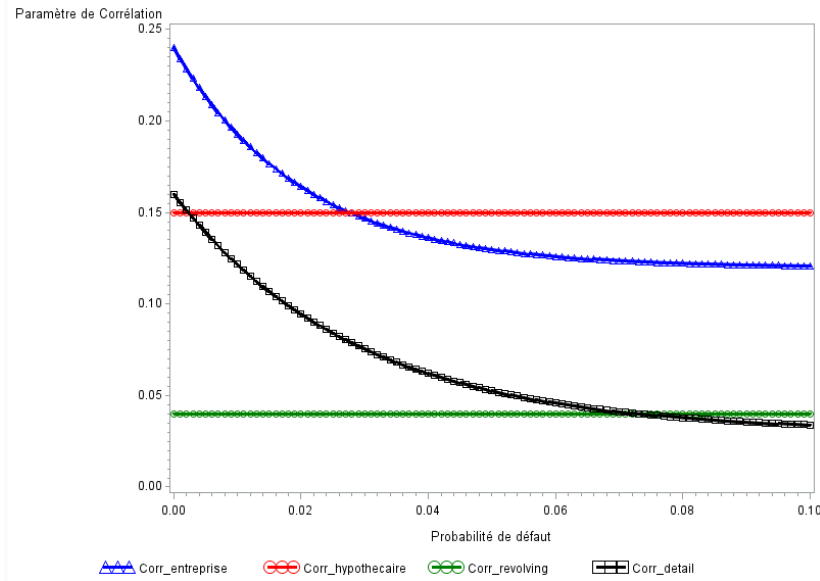


FIGURE 8.4 – Le paramètre de corrélation en fonction de la probabilité de défaut pour les différents types de crédit

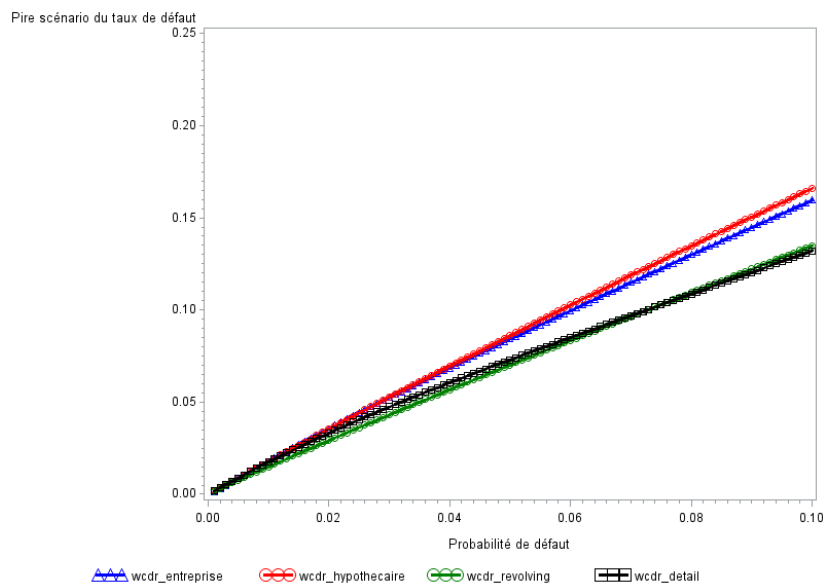


FIGURE 8.5 – Le pire scénario du taux de défaut en fonction de la probabilité de défaut pour les différents types de crédit

8.4 Stress Test de la probabilité de défaut basé sur les scénarios

On va étudier le cas simple de l'impact d'une mauvaise conjoncture économique sur l'évolution de la probabilité de défaut de notre portefeuille en l'occurrence le cas du PIB et le Taux de chômage en mauvaise conjoncture en se basant sur les scénarios de la figure 8.2.

On inclut en plus des variables macroéconomiques *PIB_temps* et *Taux_chômage* les variables liés au clients et au produit à savoir *Scor_octroi* et *Rapp_pret_valeur_temps*, le modèle utilisé est le modèle **probit**, on obtient les résultats suivants :

Estimations par l'analyse du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > Khi-2
Intercept	1	-0.1152	0.0675	2.9068	0.0882
Scor_Octroi	1	-0.00406	0.000097	1760.2215	<.0001
PIB_temps	1	-0.1394	0.00400	1211.4029	<.0001
Tx_chom	1	0.0902	0.00452	397.2631	<.0001
Rapp_pret_val_temps	1	0.0232	0.000366	4017.1519	<.0001

FIGURE 8.6 – Sortie du modèle Logistique

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	83.4	D de Somers	0.669
Pourcentage discordant	16.6	Gamma	0.669
Pourcentage lié	0.0	Tau-a	0.282
Paires	527685117	c	0.834

FIGURE 8.7 – validation du modèle

Les paramètres estimés sont significatifs montrent que la hausse du taux de croissance du PIB réduit la probabilité de défaut *PD*, par contre la hausse du taux de chômage augmente la probabilité de défaut chez le client, ce résultat peut être expliqué par le fait qu'une bonne conjoncture économique (hausse du taux de croissance du PIB) est souvent accompagné par une hausse des salaires ainsi les clients par contre la hausse du chômage reflète une mauvaise conjoncture économique ce qui va être impacté les traites des clients d'où la hausse de la probabilité de défaut (exemple perte de travail chez les personnes qui ont déjà des crédits).

On va maintenant estimer les probabilités de défaut des clients d'ici 12 mois (1 ans) dans

le cas d'un conjoncture normale appelées **PD_base** à travers le modèle probit , ainsi que dans le cas du pire scénario appelées **PD_wcdr** via la formule (*).

La figure 8.8 est un extrait des probabilités par ordre croissant calculées pour chaque client, on remarque que les probabilités **PD_wcdr** sont supérieures que celle du **PD_base** ce qui est prévisible puisque le cas du **wcdr** est un cas

PD_temps_base	PD_temps_wcdr
0.1103730954	0.1819066249
0.1103875804	0.1819286411
0.1104007643	0.1819486793
0.110415539	0.181971135
0.1104843699	0.1820757438
0.1104949501	0.1820918227
0.110499009	0.182097991
0.11051075	0.1821158337
0.1105400346	0.182160336
0.1106190787	0.182280447
0.1106250255	0.182289483
0.1106849485	0.1823805299
0.110805744	0.1825640453
0.1108416208	0.1826185449

█
█
█
█
█
█

FIGURE 8.8 – Comparaison des probabilités : PD base et PD wcdr

Le cas stressé est par définition l'étude de l'impact des variables extérieures sur l'aggravation du taux de défaut sous des hypothèses réalistes, en effet la probabilité stressée est calculée dans notre cas sous les hypothèses suivantes :

- Un taux de croissance négatif égale à -0.3% (Mauvaise Conjoncture T3/2018) 8.2
- Un taux de chômage de 11% (projections Mauvaise Conjoncture T3/2018)

Finalement, la figure 8.9 retrace les trois scénarios à savoir le scénario normal (base), pire scénario (wcdr) et le scénario stressé, on remarque que les probabilités de défaut dans le cas stressé qui est un cas plus grave que le cas du **WCDR** sont supérieures comparées au deux autres scénarios.

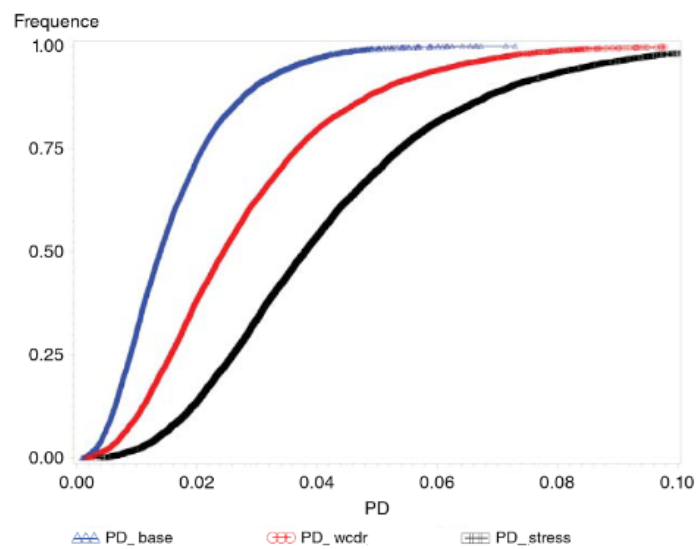


FIGURE 8.9 – Fonction de distribution cumulative pour les probabilités de défaut :base, pire scénario, stressée

Chapitre 9

Développement d'une application VBA

Sommaire

9.1 Motivation	84
9.2 Composantes de l'application VBA	84
9.3 Automatisation du calcul de la PD et du score de comportement	85
9.4 Calcul de la perte attendue	85

Introduction

Dans ce dernier chapitre on va mettre au point une application VBA qui permet l'automatisation et l'implémentation du score de comportement ainsi que la probabilité de défaut prédite pour un client donné, en se basant sur le modèle logistique obtenu au chapitre 5, et comme étape finale cette application permet d'estimer la perte attendue du client en question.

9.1 Motivation

Pour faciliter le calcul de la probabilité de défaut et le score de comportement et afin d'émettre une décision dans un délai réduit, on a pensé qu'il est judicieux de les implanter dans une application VBA .

9.2 Composantes de l'application VBA

FIGURE 9.1 – Interface de l'application VBA :calcul de la probabilité de défaillance future

La figure 9.1 nous montre les différentes composantes de notre application, il comporte comme inputs les variables explicatives retenues par le modèle logistique et comme résultat il retourne la probabilité que le client dont correspond les caractéristiques entrées fasse défaut dans le futur ainsi que son score et finalement la décision sur le client en question : profil risque ou non risqué.

9.3 Automatisation du calcul de la PD et du score de comportement

Après le calcul de la probabilité de défaut par le modèle logistique, et via la méthode de calibration en utilisant la formule vu au chapitre 5

$$\text{Score crédit} = \sum_{i=1}^N (\text{WOE}_i * \beta_i + \frac{\beta_0}{N}) * \frac{\text{offset}}{N}$$

Avec :

β_0 : la constante dans le modèle logistique

β_i : l'estimation des paramètres des variables significatives

WOE : Weight of evidence (poids d'évidence)

N : le nombre de paramètres significatifs

On obtient les score de comportement d'un client donné.

Décision

Finalement, l'application VBA retourne une décision basé sur le score de comportement calculé et le cutoff de l'expert (le seuil au dessous duquel le client est considéré comme profil risqué), donc la décision serait de type : **profil risqué** ou **non risqué**.

FIGURE 9.2 – Interface de l'application VBA : Prise de décision

9.4 Calcul de la perte attendue

L'intérêt du score de comportement c'est d'anticiper les clients qui vont tomber en défaut, et par suite constituer des provisions pour faire face au risque crédit. La formule pour le calcul des provisions nécessite la connaissance de deux autres paramètres à savoir :

- LGD : correspond à l'évaluation la perte encourue en cas de défaut de la part d'une contrepartie. Le Loss Given défaut est exprimé en pourcentage.
- EAD est l'exposition encourue par un créancier en cas de défaut de son débiteur. En d'autres termes, c'est le montant du prêteur exposé au risque de défaut de l'emprunteur. qui est le risque que le débiteur ne remplisse pas son obligation de remboursement au créancier.

La perte Attendue est ainsi obtenue par la formule :

$$EL = PD \times LGD \times EAD$$

La perte attendue doit être calculée pour chaque client EL_i , puis être agrégée pour tout le portefeuille telle que $EL_{totale} = \sum_{i=1}^n EL_i$ pour un portefeuille de n clients.

Dans notre cas on va se limiter au calcul de la perte attendue pour chaque client.

The screenshot shows a VBA UserForm2 with a green dotted background. It features three input fields on the left, each with a dark teal label: 'Probabilité de défaut' (0,72), 'LGD' (40%), and 'EAD' (4000). To the right of these fields is a red button labeled 'Perte attendue'. Below the button, a text box displays the calculated result '1152'. The top left corner contains the INSEA logo, and the top right corner contains the Wafasalaf logo.

FIGURE 9.3 – Calcul de la perte attendue d'un client donné

Conclusion générale

Mon expérience chez **Wafasalaf** m'a permis de me familiariser avec un nouveau monde, c'est celui du **risque crédit** qui est en constante évolution surtout avec l'introduction des nouvelles **normes IFRS.9**. J'ai pu acquérir à travers ce stage les fondements du développement des scores et le suivi du processus de renouvellement de la grille de score.

Ce rapport qui expose des techniques classiques de développement d'un score crédit comme la **modélisation logistique** mais adaptée dans notre cas à un **score de comportement** via l'intégration des nouvelles variables de comportement, aussi des techniques utilisés en **machine learning** comme le **réseau bayésien** ou encore les arbres de décision dans notre cas on a étudié l'**arbre CART**, on a vu qu'ils peuvent donner des modèles plus performants mais leur interprétation et leur implantation les rendent moins utilisable par les risk managers à moins d'utiliser des techniques poussées pour l'extraction des coefficients comme le système fuzzy dans le cas du SVM. Ainsi que l'utilisation des modèles poussés comme le **modèle PIT** qui fait introduire en plus des variables propres au client des **variables macroéconomiques** et tester si ces dernières ont un impact sur le taux de défaut des clients.

On a étudié toujours dans le cadre des modèles poussés, le **modèle de Cox**, qui en plus de prédire les probabilités de défaillance du client, intègre la dimension temporelle puisqu'il permet de prédire le temps de défaut, ou encore étudié le taux de défaut par strate qu'on a vu pour la variables type de produit.

Finalement, on a étudié la modélisation par inférence bayésienne qui appartient à un tout autre référentiel de modélisation totalement différent du paradigme fréquentiel. La technique utilisée dans notre cas est celle du Monte Carlo Chaine de Markov **MCMC**, cette technique repose sur un estimation de la probabilité de défaut sur un échantillon puis inférer sur toute la population.

L'étape qui a suivi la modélisation était bien sur de tester la rigidité de cette modélisation vis à vis les variations extérieures notamment le calcul de la probabilité dans le cas du pire scénario ou dans un cas plus hypothétique appelé la probabilité stressée calculée dans le cas ou le scénario est considéré plus pire que celui du **WCDR :pire scénario**.

Finalement, on a automatisé le calcul de la probabilité de défaut ainsi que le score qui lui correspond à travers une application **VBA**, par la suite et grâce à cette probabilité calculée on estime la perte attendue pour un client donné.

Annexe A

Annexes

A.1 Figures annexes

Scorecard

		Groupe	Points du scorecard	Poids d'évidence	Taux d'événement BIN_DEFAULT = 1	Pourcentage de population	Coefficient
Anc_Prem_Crd	Anc_Prem_Crd < 69.13	1.00	58	-1.08	41.98	4.99	-0.97
	69.13 <= Anc_Prem_Crd < 172.5, _MISSING_	2.00	79	-0.33	25.59	45.00	-0.97
	172.5 <= Anc_Prem_Crd < 240.69	3.00	101	0.47	13.34	30.00	-0.97
	240.69 <= Anc_Prem_Crd < 312.74	4.00	113	0.88	9.26	15.00	-0.97
	312.74 <= Anc_Prem_Crd	5.00	96	0.30	15.49	5.02	-0.97
MT_PRET	MT_PRET < 8000	1.00	63	-1.02	40.53	9.70	-0.86
	8000 <= MT_PRET < 15200, _MISSING_	2.00	87	-0.05	20.51	34.90	-0.86
	15200 <= MT_PRET < 21600	3.00	94	0.23	16.34	25.36	-0.86
	21600 <= MT_PRET < 37700	4.00	99	0.44	13.75	25.01	-0.86
	37700 <= MT_PRET	5.00	85	-0.13	21.96	5.04	-0.86
NINQ	NINQ < 1, _MISSING_	1.00	96	0.33	15.06	46.13	-0.87
	1 <= NINQ < 2	2.00	93	0.20	16.84	24.61	-0.87
	2 <= NINQ < 3	3.00	83	-0.18	22.74	14.60	-0.87
	3 <= NINQ < 4	4.00	79	-0.33	25.62	7.63	-0.87
	4 <= NINQ	5.00	54	-1.34	48.49	7.04	-0.87
Val_prop	Val_prop < 43600	1.00	64	-1.25	46.23	4.99	-0.67
	43600 <= Val_prop < 88303, _MISSING_	2.00	86	-0.12	21.78	40.00	-0.67
	88303 <= Val_prop < 121781	3.00	97	0.46	13.42	30.00	-0.67
	121781 <= Val_prop < 177819	4.00	86	-0.11	21.66	15.00	-0.67
	177819 <= Val_prop	5.00	95	0.35	14.82	10.01	-0.67
nombr_impay	nombr_impay < 1, _MISSING_	1.00	100	0.43	13.80	79.68	-0.95
	1 <= nombr_impay < 2	2.00	69	-0.68	32.70	11.23	-0.95
	2 <= nombr_impay	3.00	42	-1.65	56.22	9.09	-0.95

FIGURE A.1 – Les scorecards crédit du modèle logistique de notre portefeuille avec SAS Miner

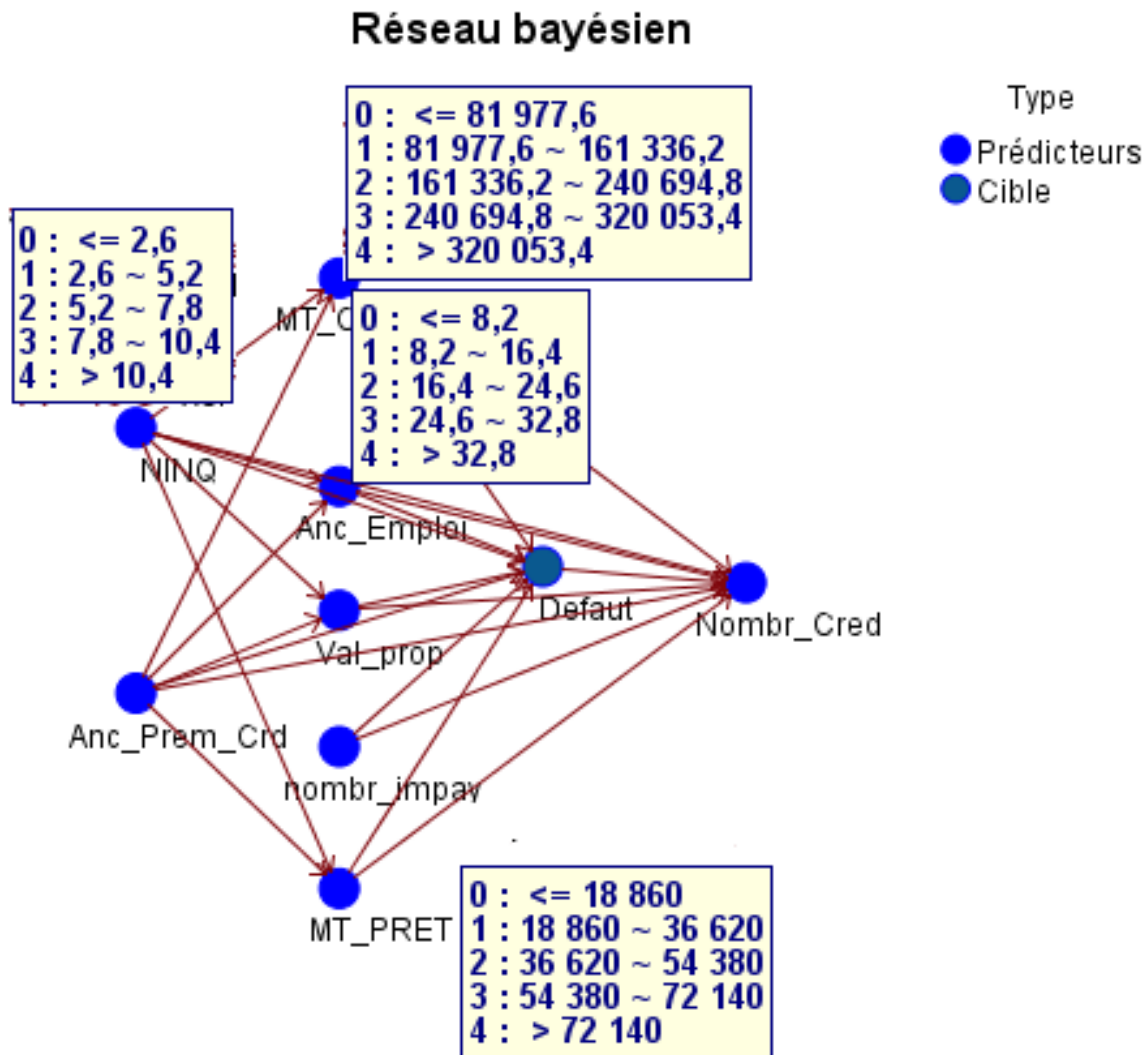


FIGURE A.2 – Réseau bayésien :structure couverture de Markov

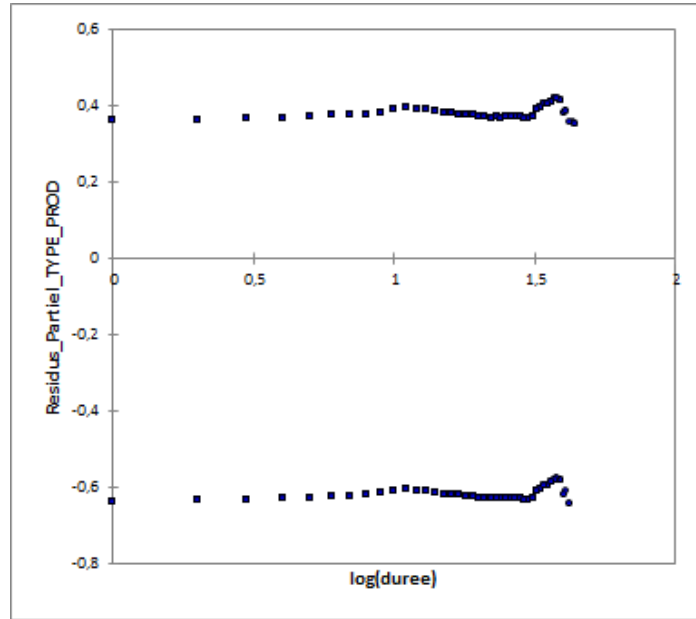
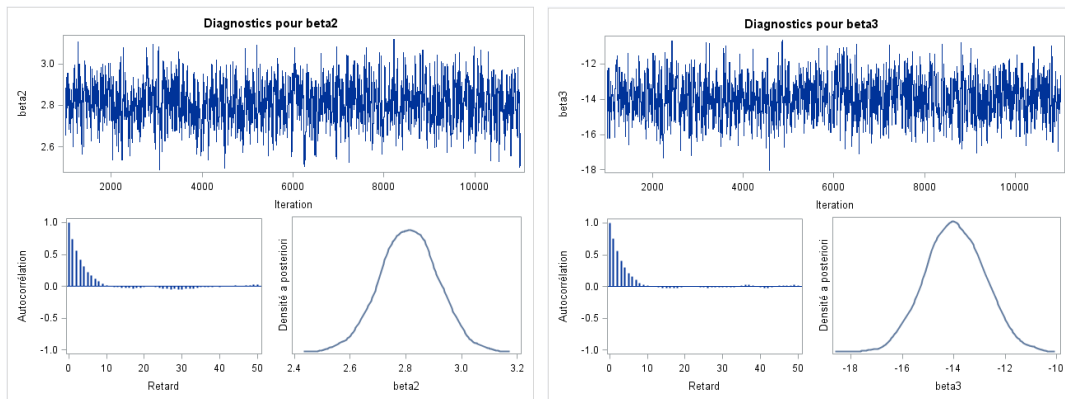


FIGURE A.3 – Résidus partiels : type produit

Diagnostics de Heidelberg-Welch								
Paramètre	Test de stationnarité				Test de demi-largeur			
	Stat de Cramer-von Mises	p-Value	Résultat du test	Itérations supprimée	Demi-largeur	Moyenne	Demi-largeur relative	Résultat du test
beta0	0.1268	0.4688	Succès	0	0.0113	-0.1991	-0.0565	Succès
beta1	0.1445	0.4067	Succès	0	0.000179	-0.0375	-0.00476	Succès
beta2	0.0612	0.8067	Succès	0	0.00650	2.8109	0.00231	Succès
beta3	0.0749	0.7218	Succès	0	0.0831	-13.9495	-0.00596	Succès

FIGURE A.4 – Procédure MCMC pour le modèle probit



(a) diagnostique β_2

(b) diagnostique β_3

FIGURE A.5 – Diagnostique des paramètres β_2 et β_3

```
Private Sub ComboBox1_Change()  
  
End Sub  
  
Private Sub CommandButton1_Click()  
TextBox1.Value = ""  
ComboBox1.Value = ""  
ComboBox2.Value = ""  
ComboBox3.Value = ""  
ComboBox4.Value = ""  
End Sub  
  
Private Sub CommandButton2_Click()  
If ComboBox1.Value = "]0-69]" And (ComboBox2.Value = "]1-5]" Or ComboBox2.Value = "]5-10]"  
    TextBox1.Value = 3733  
End If  
If ComboBox1.Value = "]70-172]" And (ComboBox2.Value = "]1-5]" Or ComboBox2.Value = "]5-10]"  
    TextBox1.Value = 3733  
End If  
If ComboBox1.Value = "]173-240]" And (ComboBox2.Value = "]1-5]" Or ComboBox2.Value = "]5-10]"  
    TextBox1.Value = 3733  
End If  
If ComboBox1.Value = "]241-312]" And (ComboBox2.Value = "]1-5]" Or ComboBox2.Value = "]5-10]"  
    TextBox1.Value = 3191  
End If  
If ComboBox1.Value = "]313-]" And (ComboBox2.Value = "]1-5]" Or ComboBox2.Value = "]5-10]"  
    TextBox1.Value = 3191
```

FIGURE A.6 – Extrait du code VBA pour l'automatisation du calcul du score

A.2 Tableaux annexes

Probabilités conditionnelles de Default							
Lignes 1 de 100 à 253							
Parents							Probabilité
MT_PRET	MT_Cred_exist	Val_prop	Anc_Emploi	Anc_Prem_Crd	nombr_impay	NINQ	1
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	<= 2,6	0,23
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	2,6 ~ 5,2	0,36
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	5,2 ~ 7,8	0,80
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	7,8 ~ 10,4	0,38
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	<= 3	> 10,4	1,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	3 ~ 6	<= 2,6	0,75
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	3 ~ 6	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	3 ~ 6	5,2 ~ 7,8	1,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	<= 233,646	6 ~ 9	<= 2,6	1,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	233,646 ~ 467,292	<= 3	<= 2,6	0,09
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	0,33
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	233,646 ~ 467,292	3 ~ 6	<= 2,6	1,00
<= 18 880	<= 81 977,6	<= 112 530	<= 8,2	233,646 ~ 467,292	3 ~ 6	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	<= 3	<= 2,6	0,31
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	<= 3	2,6 ~ 5,2	0,31
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	<= 3	5,2 ~ 7,8	0,75
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	<= 3	7,8 ~ 10,4	1,00
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	<= 3	> 10,4	1,00
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	3 ~ 6	<= 2,6	0,60
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	3 ~ 6	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	<= 233,646	6 ~ 9	<= 2,6	1,00
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	233,646 ~ 467,292	<= 3	<= 2,6	0,05
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	0,12
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	233,646 ~ 467,292	<= 3	5,2 ~ 7,8	1,00
<= 18 880	<= 81 977,6	<= 112 530	8,2 ~ 16,4	233,646 ~ 467,292	3 ~ 6	<= 2,6	1,00
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	<= 233,646	<= 3	<= 2,6	0,16
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	<= 233,646	<= 3	2,6 ~ 5,2	0,57
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	<= 233,646	<= 3	7,8 ~ 10,4	0,00
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	<= 233,646	3 ~ 6	<= 2,6	0,50
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	<= 2,6	0,03
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	0,00
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	7,8 ~ 10,4	0,00
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	> 10,4	0,00
<= 18 880	<= 81 977,6	<= 112 530	16,4 ~ 24,6	233,646 ~ 467,292	3 ~ 6	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	<= 112 530	24,6 ~ 32,8	<= 233,646	<= 3	<= 2,6	0,06
<= 18 880	<= 81 977,6	<= 112 530	24,6 ~ 32,8	<= 233,646	<= 3	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	<= 112 530	24,6 ~ 32,8	<= 233,646	3 ~ 6	<= 2,6	1,00
<= 18 880	<= 81 977,6	<= 112 530	24,6 ~ 32,8	<= 233,646	> 12	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	<= 112 530	24,6 ~ 32,8	233,646 ~ 467,292	<= 3	<= 2,6	0,07
<= 18 880	<= 81 977,6	<= 112 530	24,6 ~ 32,8	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	<= 112 530	24,6 ~ 32,8	233,646 ~ 467,292	3 ~ 6	<= 2,6	1,00
<= 18 880	<= 81 977,6	<= 112 530	> 32,8	<= 233,646	<= 3	2,6 ~ 5,2	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	<= 8,2	<= 233,646	<= 3	<= 2,6	0,50

FIGURE A.7 – Tableau des probabilités conditionnelles de défaut :Réseau bayésien Markovien

<= 18 880	<= 81 977,6	112 530 ~ 212 580	<= 8,2	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	<= 8,2	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	0,00	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	8,2 ~ 16,4	<= 233,646	<= 3	<= 2,6	1,00	0,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	8,2 ~ 16,4	<= 233,646	<= 3	7,8 ~ 10,4	1,00	0,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	8,2 ~ 16,4	<= 233,646	<= 3	> 10,4	1,00	0,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	8,2 ~ 16,4	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	8,2 ~ 16,4	233,646 ~ 467,292	3 ~ 6	<= 2,6	1,00	0,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	16,4 ~ 24,6	<= 233,646	<= 3	<= 2,6	0,00	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	16,4 ~ 24,6	<= 233,646	<= 3	2,6 ~ 5,2	1,00	0,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	0,00	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	24,6 ~ 32,8	<= 233,646	<= 3	<= 2,6	0,00	1,00
<= 18 880	<= 81 977,6	112 530 ~ 212 580	24,6 ~ 32,8	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	<= 81 977,6	212 580 ~ 312 580	<= 8,2	<= 233,646	<= 3	<= 2,6	1,00	0,00
<= 18 880	<= 81 977,6	212 580 ~ 312 580	8,2 ~ 16,4	<= 233,646	<= 3	7,8 ~ 10,4	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	<= 8,2	<= 233,646	<= 3	<= 2,6	0,17	0,83
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	<= 8,2	<= 233,646	<= 3	2,6 ~ 5,2	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	<= 8,2	<= 233,646	3 ~ 6	<= 2,6	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	<= 8,2	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	<= 8,2	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	8,2 ~ 16,4	<= 233,646	<= 3	<= 2,6	0,20	0,80
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	8,2 ~ 16,4	<= 233,646	<= 3	2,6 ~ 5,2	0,12	0,88
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	8,2 ~ 16,4	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	16,4 ~ 24,6	<= 233,646	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	16,4 ~ 24,6	<= 233,646	<= 3	2,6 ~ 5,2	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	16,4 ~ 24,6	233,646 ~ 467,292	<= 3	2,6 ~ 5,2	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	24,6 ~ 32,8	<= 233,646	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	24,6 ~ 32,8	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	<= 112 530	24,6 ~ 32,8	467,292 ~ 817,761	<= 3	<= 2,6	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	<= 8,2	<= 233,646	<= 3	<= 2,6	0,15	0,85
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	<= 8,2	<= 233,646	<= 3	2,6 ~ 5,2	0,64	0,36
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	<= 8,2	<= 233,646	<= 3	5,2 ~ 7,8	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	<= 8,2	<= 233,646	3 ~ 6	2,6 ~ 5,2	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	<= 8,2	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	<= 8,2	> 817,761	<= 3	<= 2,6	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	8,2 ~ 16,4	<= 233,646	<= 3	<= 2,6	0,15	0,85
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	8,2 ~ 16,4	<= 233,646	<= 3	2,6 ~ 5,2	0,06	0,94
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	8,2 ~ 16,4	<= 233,646	3 ~ 6	<= 2,6	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	8,2 ~ 16,4	<= 233,646	3 ~ 6	5,2 ~ 7,8	1,00	0,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	8,2 ~ 16,4	233,646 ~ 467,292	<= 3	<= 2,6	0,00	1,00
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	16,4 ~ 24,6	<= 233,646	<= 3	<= 2,6	0,19	0,81
<= 18 880	81 977,6 ~ 161 336,2	112 530 ~ 212 580	16,4 ~ 24,6	<= 233,646	<= 3	5,2 ~ 7,8	1,00	0,00

FIGURE A.8 – Tableau des probabilités conditionnelles de défaut :Réseau bayésien Markovien (suite)

A.3 Références

- [1] Wafasalaf. Rapport annuel. *2015*.
- [2] CHATER Mohammed. Cours Monnaie et banque *2016*.
- [3] Naeem Siddiqi. Credit Risk Scorecards Developing *2007*.
- [4] Bart Baesens. Developing Intelligent Systems for Credit Scoring Using Machine Learning Techniques *2003*.
- [5] Ricardo Cao, Juan M. Vilar and Andrés Devia. Modeling consumer Credit risk Via survival analysis *Universidade da Coruna ,2009*.
- [6] CHAOUBI Abdelaziz. Cours de l'analyse de durée de vie *2016*.
- [7] CHAOUBI Abdelaziz. Cours sur les modèles GLM *2016*.
- [8] Comité de Bâle. Réglementation du comité de Bâle *2013*.
- [9] Comité de bâte. Normes IFRS.9.
- [10] CHAOUBI Mohammed. Analyse statistique des durées de vie *Modèle de régression de Cox avec SPSS ,2009*.
- [11] Gilbert Ritschard. Estimer un modèle de Cox en temps continu avec SPSS *Université de Genève ,2004*.
- [12] CAMPILLO Fabien et ROSSI Vivien. les méthodes de Monte Carlo *INRIA ,2006*.