

Mémoire présenté devant l'Institut du Risk Management  
pour la validation du cursus à la Formation d'Actuaire  
de l'Institut du Risk Management  
et l'admission à l'Institut des Actuaire  
le 22 / 06 / 2020

Par : **Fares Boulaouad**

Titre : **Modélisation du processus de commutation  
de la réassurance en run-off au sein d'AXA LM**

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

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

*Membres présents du jury de l'Institut  
des Actuaire :*

*Entreprise : AXA Liabilities Managers*

*Nom : Hélène Gibello*

*Signature et cachet :*



**AXA LIABILITIES MANAGERS**

*Directeur du mémoire en entreprise :*  
61 rue Mstislav Rostropovitch - 75017 Paris

*Nom : Hélène Gibello N° Siren : 443 984 778*

*Signature :*

*Membres présents du jury de l'Institut  
du Risk Management :*

*Invité :*

*Nom :*

*Signature :*

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

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

*Secrétariat :*


Signature du responsable entreprise






*Bibliothèque :*

Signature du candidat



# Table des matières

Table des matières . . . . .	4
Remerciements . . . . .	5
Résumé . . . . .	6
Abstract . . . . .	7
Introduction . . . . .	8
<b>1 Contexte et données</b>	<b>11</b>
1.1 Activité d'AXA Liabilities Managers . . . . .	11
1.1.1 AXA Liabilities Managers acquiert une expertise en gestion de run-off . . . . .	11
1.1.2 Un modèle axé sur l'accélération de la finalité . . . . .	12
1.2 Présentation des données . . . . .	14
1.2.1 Données sur les cédantes . . . . .	14
1.2.2 Données sur les commutations . . . . .	14
1.2.3 Données sur les contrats . . . . .	21
1.3 Création de pas de temps pour les transitions . . . . .	21
1.4 Conclusion . . . . .	25
<b>2 Processus de commutation modélisé par chaînes de Markov</b>	<b>27</b>
2.1 Rappels théoriques sur les chaînes de Markov . . . . .	28
2.2 Vérification de la propriété de Markov . . . . .	30
2.2.1 Validation globale de la propriété de Markov sur l'ensemble des données . . . . .	31
2.2.2 Validation de la propriété de Markov projet par projet . . . . .	31
2.3 Estimation de la matrice de transition . . . . .	32
2.3.1 Estimateur de la matrice de transition . . . . .	32
2.3.2 Calcul sous  des matrices de transition . . . . .	33
2.3.3 Interprétation des matrices de transition . . . . .	37
2.4 État stationnaire . . . . .	38
2.4.1 Estimation de l'état stationnaire . . . . .	38
2.4.2 Interprétation des résultats de l'estimation de l'état stationnaire . . . . .	40
2.5 Conclusion . . . . .	40
<b>3 Modélisation logistique et segmentation des commutations</b>	<b>41</b>
3.1 Le modèle de régression logistique : Aspects théorique . . . . .	42
3.1.1 Présentation du modèle . . . . .	42
3.1.2 Les estimateurs dans les cas binaire et multinomial . . . . .	44
3.1.3 Tests statistiques . . . . .	46

3.1.4	Odds-ratio . . . . .	51
3.2	Variable d'intérêt et variables explicatives du modèle . . . . .	52
3.3	Mise en oeuvre sous  et éléments d'interprétation . . . . .	54
3.4	Modèle multinomial statique . . . . .	57
3.4.1	Estimation du modèle statique avec tous les statuts sans regroupement . . . . .	57
3.4.2	Estimation du modèle statique avec statuts regroupés . . . . .	58
3.5	Modèle multinomial dynamique . . . . .	61
3.6	Synthèse des critères de segmentation et estimation des matrices de transition avec segmentation . . . . .	66
3.6.1	Synthèse des critères de segmentation . . . . .	66
3.6.2	Estimation des matrices de transition segmentées du modèle de Markov . . . . .	66
3.7	Conclusion . . . . .	70
	Conclusion générale . . . . .	71
<b>A</b>	<b>Annexe : Résultats et code  pour le modèle de Markov</b>	<b>73</b>
A.1	Vérification de la propriété de Markov . . . . .	73
A.1.1	Lecture et concaténation des données pour obtenir une seule chaîne de Markov	73
A.1.2	Vérification de la propriété de Markov sur l'ensemble des données comme une seule chaîne . . . . .	73
A.1.3	Vérification de la propriété de Markov sur les données projet par projet . . . . .	74
A.2	Code  pour l'estimation de la matrice de transition . . . . .	75
A.2.1	Transitions pour 5 états et 4 pas de temps . . . . .	75
A.2.2	Transitions pour 4 états et 4 pas de temps . . . . .	76
A.3	Affichage des matrices de transition estimées par  . . . . .	77
A.3.1	Estimation de la matrice de transition mensuelle . . . . .	77
A.3.2	Estimation de la matrice de transition trimestrielle . . . . .	79
A.3.3	Estimation de la matrice de transition semestrielle . . . . .	81
A.3.4	Estimation de la matrice de transition annuelle . . . . .	83
A.3.5	État stationnaire . . . . .	85
<b>B</b>	<b>Annexe : Résultats détaillés du modèle logistique et estimation des matrices de transition avec segmentation</b>	<b>87</b>
B.1	Effectifs des modalités des variables qualitatives utilisées dans le modèle logistique . . . . .	87
B.1.1	Effectifs des modalités de la variable d'intérêt . . . . .	87
B.1.2	Effectifs des modalités des variables explicatives . . . . .	88
B.2	Résultats et code  du modèle logistique . . . . .	89
B.2.1	Choix de modèle par sélection pas à pas descendante . . . . .	89
B.2.2	Modèle logistique statique . . . . .	93
B.2.3	Modèle logistique dynamique . . . . .	96
B.2.4	Matrices de transition du modèle de Markov avec segmentation . . . . .	98

## Remerciements

Je tiens à remercier en premier lieu toute ma famille pour le soutien qu'elle m'apporte en toutes circonstances malgré l'éloignement.

Je tiens aussi à remercier particulièrement Hélène Gibello, ma responsable à *AXA Liabilities Managers* (AXA LM), qui m'a soutenu tout au long de ce travail par ses remarques très constructives et m'a aidé à franchir de multiples obstacles rencontrés lors de l'élaboration de ce document. Hélène m'a orienté, en effet, vers un contenu adapté à la fois au contexte d'AXA LM mais aussi aux exigences d'un mémoire d'actuariat.

Du point de vue académique, ce mémoire a été suivi par Christophe Dutang. Son implication à travers les remarques qu'il a apportées sur les différentes versions de ce document ainsi que les propositions de méthodes m'ont permis d'avancer en suivant un cadre théorique adapté à la problématique et aux données. Je tiens à le remercier vivement pour sa rigueur et son professionnalisme exemplaires.

Enfin, mes remerciements vont à tous ceux qui ont contribué de près ou de loin à ce travail : Amine Bouassida, Alice Charguéraud, Thérèse Chauvel, Rachid Hair, Thomas Lallement, Berthold Sobze, Adèle Asselain, Barnard Da Silva, Sandra McKenna et bien d'autres.

## Résumé

Les sociétés d'assurance ou de réassurance doivent être dotées d'un capital minimum pour leur permettre de faire face à des risques extrêmes. Dans le cadre de la gestion de run-off, la logique de création de valeur veut que la détention de capital soit optimisée en montant et dans le temps afin de rémunérer au mieux les investisseurs et de rester compétitif. Il convient ainsi de réduire le capital par le biais de différents vecteurs afin de remonter les dividendes aux investisseurs. Dans le cadre de la gestion de run-off en réassurance, un des vecteurs de réduction du risque, et donc du capital, est d'aboutir à des accords permettant la reprise des risques par la cédante contre la réception d'un montant correspondant à l'estimation des charges restantes sur les contrats repris : c'est le mécanisme de commutation. Dans ce mémoire, nous proposons l'application de modèles de Markov sur les données disponibles ainsi que le recours à la régression logistique afin d'analyser le processus de commutation au sein d'*AXA Liabilities Managers*. L'objectif principal est de construire un modèle qui permette d'apprécier les chances qu'un projet de commutation puisse aboutir à un horizon de temps donné. Il s'agit d'établir une segmentation selon des critères pertinents à appliquer, par exemple, au niveau contrat ou au niveau cédante, afin de mieux cibler les commutations réalisables. Ce travail permettra de tirer des conclusions qui aideront à améliorer la connaissance et la maîtrise des portefeuilles gérés, suggérer des adaptations au processus de commutation en place et enfin contribuer aux projets à venir d'acquisition de nouveaux portefeuilles de run-off.

**Mots-clés** — run-off, réassurance, cédante, réassureur, commutation, acquisition, chaînes de Markov, matrice de transition, régression logistique, segmentation

## Abstract

Insurance and Reinsurance companies need to hold a minimum amount of capital in order to face extreme risks. In the context of run-off management, value creation requires that the capital holdings be optimized in amount and over time in order to best compensate investors and remain competitive. It is therefore appropriate to reduce this capital through different possible ways in order to raise dividends for investors. More specifically, in reinsurance run-off management, one manner of risk reduction, and thus capital reduction, is to achieve an accord in which the cedant reassumes its risk in exchange for a cash amount representing an estimation of the remaining risks included in the contracts taken back – a mechanism known as commutation. This document aims to analyze the commutation process at *AXA Liabilities Managers* by applying Markov models and logistic regression on available data. The main goal is to build a model that allows to assess the chances of achievement for a commutation project at a given time horizon. A segmentation according to consistent criteria to be applied, for example, at contract or at cedant level, will allow for better commutation targeting. The conclusions of this analysis will enhance knowledge on the managed portfolios, lead to suggestions for improving the current commutation process and finally contribute to the targeting of new portfolio acquisitions.

**Keywords** — run-off, reinsurance, cedant, reinsurer, commutation, acquisition, Markov chains, transition matrix, logistic regression, segmentation

## Introduction

L'objectif de ce mémoire est de modéliser le processus de commutation des contrats de réassurance en run-off au sein d'AXA LM. Les commutations constituent un vecteur majeur d'accélération de la finalité dans le cadre de la gestion de run-off en réassurance. Un accord de commutation permet la reprise par une société d'assurance ou de réassurance des risques qu'elle avait cédés à un autre réassureur. Cette reprise des risques par la cédante est accompagnée par la réception d'un montant correspondant à l'estimation des charges restantes de sinistres. Le réassureur se voit ainsi réduire les risques dans les portefeuilles qu'il détient et pourra réduire le capital correspondant. Cela lui permet d'avoir une meilleure visibilité sur son résultat. Il pourra ainsi rémunérer ses investisseurs dans les meilleurs délais. Un accord de commutation n'est signé que si les deux parties estiment avoir un intérêt : pour le réassureur l'intérêt est de réduire ses risques ; quant à la cédante cela lui permettra d'obtenir des liquidités et de faire face, par exemple, à des difficultés de trésorerie.

Étant donné l'importance des commutations comme vecteur d'accélération de la finalité, une étude approfondie du processus permettant d'aboutir (ou non) à un accord de commutation sera proposée. La démarche générale consiste à analyser les données passées afin de déterminer les trajectoires générales et les éléments explicatifs des différences de trajectoires entre les projets de commutation. Les trajectoires générales sont définies par les transitions entre les différents états possibles d'un projet de commutation de façon globale ; autrement dit, sans tenir compte des caractéristiques spécifiques à chaque projet de commutation, des cédantes impliquées ou des contrats concernés par le projet. Les éléments explicatifs des différences, sont ceux qui influent sur les transitions et font dévier la trajectoire prise par un projet de commutation particulier, de la trajectoire de tendance générale.

Cette étude s'intéressera notamment à la possibilité de segmentation des projets de commutation selon différents critères afin de pouvoir cibler les contrats ou les cédantes qui permettent en priorité d'aboutir à un accord de commutation concernant des portefeuilles gérés actuellement par AXA LM ; mais aussi afin de réaliser des projections de commutation sur les acquisitions de nouveaux portefeuilles. Dans le cadre de cette étude, nous tenterons d'évaluer la probabilité de succès d'un projet ou de son annulation.

Afin de parvenir à ces objectifs nous proposons d'abord, dans le premier chapitre, une présentation détaillée du contexte de l'étude ainsi qu'un exposé des statistiques descriptives et des analyses exploratoires réalisées sur les données existantes. En fonction de ces analyses, nous affinerons nos objectifs et nous proposerons un cadre d'étude plus précis, par exemple, la construction de pas de temps pour les transitions d'un état à un autre nous aidera à utiliser le modèle de Markov.

Le deuxième chapitre sera consacré à une modélisation du processus de commutation par des chaînes de Markov. Les données étant datées, nous utiliserons plusieurs pas de temps (mensuel, trimestriel, semestriel et annuel) afin de déterminer la pertinence du pas choisi. Nous listerons les arguments favorables et opposés quant à l'utilisation de ces différents pas de temps. Nous retiendrons le pas de temps le plus pertinent au vu des données et du contexte afin de l'utiliser dans la suite de ce document.

Le troisième et dernier chapitre analysera les processus de commutation sous un autre angle, celui de la régression logistique par l'introduction de variables explicatives. Il s'agit d'appliquer un



modèle multinomial sur les variables explicatives afin de voir leur influence sur la variable d'intérêt choisie. Nous distinguerons selon la variable d'intérêt choisie, un modèle logistique statique et un modèle logistique dynamique. Ce dernier permettra de faire le lien avec le modèle de Markov et d'évaluer des matrices de transition segmentées.

En conclusion, une synthèse des critères de segmentation sera présentée et un bilan des résultats de cette étude sera dressé ainsi que les points qui restent en suspens. Nous récapitulerons notamment, les conclusions qui peuvent servir pour répondre à la problématique de ce mémoire d'une part, et celles à prendre avec précaution ainsi que les extensions possibles de modèle, d'autre part.



# Chapitre 1

## Contexte et données

Nous aborderons ce premier chapitre par un bref historique d'AXA LM, une présentation de son activité de gestion de run-off, ainsi que son business model.

Ensuite, un aperçu global sur les données existantes sera exposé. Afin de faciliter la présentation, les données seront organisées en trois groupes se rapportant chacun à un sujet particulier. D'abord, le groupe des données sur les cédantes sera présenté, ensuite les données sur les commutations, pour finir par un exposé sur les données relatives aux contrats. Au fur et à mesure de cet exposé, des statistiques descriptives sur chacun de ces groupes accompagneront la présentation.

Avant de conclure ce chapitre, le travail de préparation des données sera décrit. Cette préparation vise à présenter les données sous une forme adaptée aux modèles utilisés dans les chapitres suivants. En particulier, plusieurs durées de pas de temps séparant deux états successifs seront définies et analysées.

### 1.1 Activité d'AXA Liabilities Managers

#### 1.1.1 AXA Liabilities Managers acquiert une expertise en gestion de run-off

*AXA Liabilities Managers* (AXA LM) est une société de gestion de passifs de réassurance mis en run-off. Elle a été créée en 2001 pour gérer des portefeuilles de run-off pour le compte du *Groupe AXA*.

En 2008, AXA LM ouvre son premier fonds d'investissement dans le domaine de la gestion de run-off. Ce fonds est ouvert à des investisseurs privés externes qui s'engagent à apporter leurs capitaux afin de financer les acquisitions futures. Les fonds levés auprès de ces investisseurs, entre 2008 et 2014, ont permis à AXA LM de réaliser plusieurs opérations d'acquisition de portefeuilles de run-off mis en vente par des compagnies de réassurances autres que le *Groupe AXA*. Les investisseurs du fonds en sont les propriétaires et AXA LM gère ces portefeuilles pour le compte des investisseurs. AXA LM mise sur l'expertise acquise au fil des années afin de tirer de la valeur de la gestion de passifs de réassurance à travers le monde. Ainsi, plusieurs acquisitions ont été conclues dans le cadre de ce premier fonds, par exemple, le rachat de plusieurs portefeuilles de Global Re.

Les performances visées ayant été respectées, le premier fonds est un succès. Cette première réussite dans la gestion de passifs pour le compte de tiers permet à AXA LM d'obtenir une notoriété sur le marché et surtout la confiance des investisseurs. Ceci ouvre la voie à AXA LM pour la création en 2016 d'un deuxième fonds d'investissement. De nouvelles acquisitions sont conclues dans le cadre de ce deuxième fonds et, en février 2018, AXA LM fait son entrée sur le marché des Lloyd's au Royaume-Uni en finalisant son premier RITC (ReInsurance To Close). En septembre 2018, AXA LM a finalisé avec succès l'acquisition d'une participation majoritaire à hauteur de 92,68% dans Emirates Re, qui était autrefois le plus grand réassureur islamique international indépendant.

### 1.1.2 Un modèle axé sur l'accélération de la finalité

Dans un contexte Solvabilité 2, les assureurs sont amenés à accroître la connaissance de leurs risques afin de mieux les maîtriser. Les portefeuilles gérés en run-off doivent répondre à des contraintes réglementaires au même titre que les affaires actives. Parmi ces contraintes, l'obligation de détenir un capital suffisant, pour faire face à un scénario exceptionnel, immobilise des ressources et peut empêcher des voies de développement.

Pour une société qui continue à souscrire des contrats d'assurance ou de réassurance, l'immobilisation du capital peut être justifiée par le fait que la branche soit rentable. Cependant, la gestion de portefeuilles en run-off ne fonctionne pas selon les mêmes schémas que ceux des portefeuilles avec souscription.

Dans le cadre de ce mémoire, nous nous focaliserons sur le modèle d'AXA LM qui suppose, à la différence de la réassurance active, une liquidation la plus rapide possible des affaires détenues et mises en run-off. Ainsi, plusieurs voies possibles permettent cette liquidation rapide que l'on peut désigner aussi par *accélération de la finalité*. Nous pouvons citer la gestion proactive des sinistres en cours, les commutations, les novations, etc. comme le montre la FIGURE 1.1. Le but final étant de pouvoir libérer du capital le plus rapidement possible en réduisant le volume des affaires gérées afin de rémunérer les investisseurs avec une meilleure rentabilité.

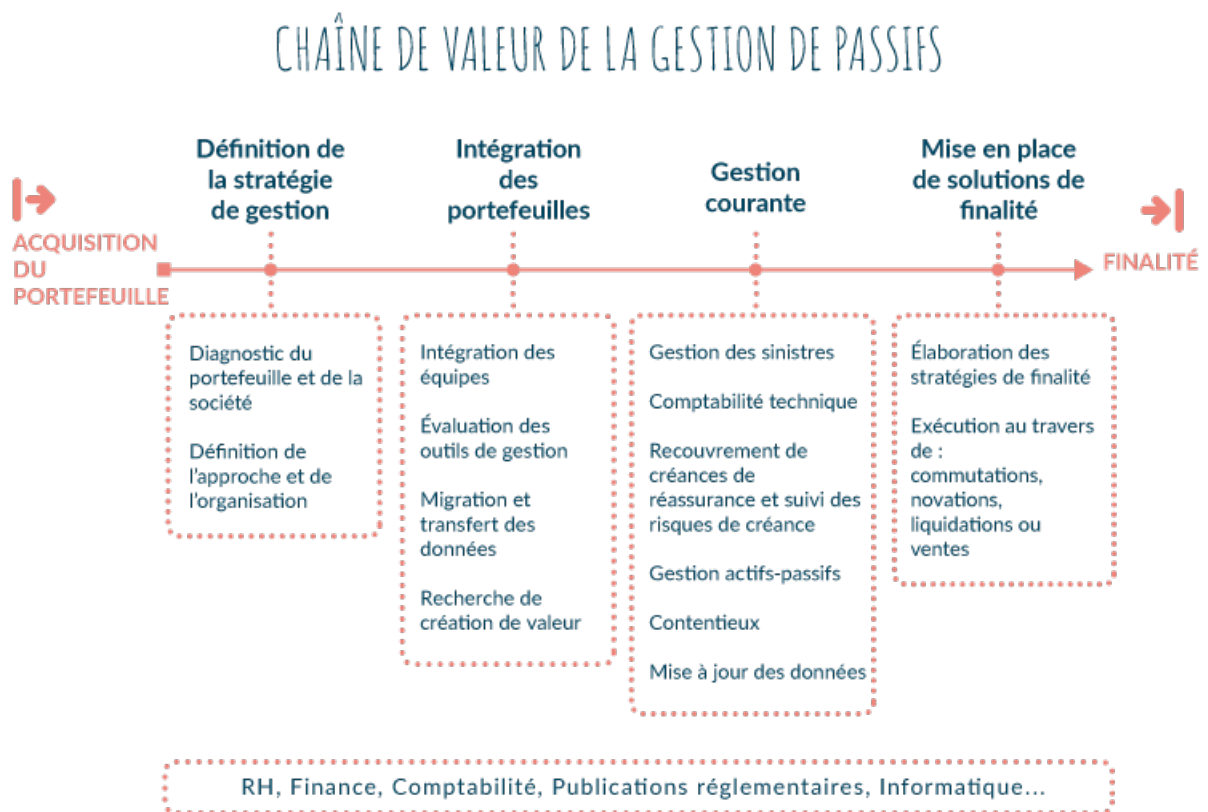


FIGURE 1.1 – Chaîne de valeur en gestion de run-off

Source : documentation interne AXA LM

## 1.2 Présentation des données

### 1.2.1 Données sur les cédantes

Il n'est pas aisé de suivre dans le temps les évolutions de noms des réassureurs ainsi que les fusions et acquisitions de certains d'entre eux. Une base de données permettant l'identification des compagnies de réassurance est mise à jour régulièrement au sein d'AXA LM. Cette base est utilisée par les équipes rétrocession et aide à l'évaluation du risque de contrepartie. La mise à jour la plus récente date de décembre 2018. Nous pouvons utiliser la même base pour identifier les cédantes intervenant dans les commutations.

Plus de 3800 cédantes ont été approchées par les responsables commutation à AXA LM pour leur proposer des accords de commutation. Il peut s'agir de différentes sociétés d'un même Groupe mais cela ne devrait pas poser de problème pour notre analyse. En effet, lorsque deux codes différents sont appliqués à la même cédante, les caractéristiques analysées devraient révéler des similarités entre les profils de risque.

Les variables présentes dans cette base et qui peuvent servir dans notre étude sont notamment : le pays d'origine de la cédante ainsi que le rating de celle-ci. Ce rating est un score de crédit (de solvabilité) créé par le département Credit Control à AXA LM. Il est issu d'une synthèse de ratings marché.

### 1.2.2 Données sur les commutations

Les données de commutations sont organisées par projet. Chaque projet de commutation est identifié par un code et possède un nom de projet. Une ou plusieurs cédantes peuvent faire partie d'un même projet de commutation.

Dans le processus actuel de commutation, une fois que les cédantes concernées sont choisies, le périmètre d'étude doit être défini (étape de « Scope definition »). Il s'agit de recenser l'ensemble des contrats qui entrent dans le cadre de la commutation. Ces contrats peuvent provenir d'un ou de plusieurs portefeuilles gérés par AXA LM (par exemple : Colisée Re, Portman, DBU, . . . , etc.).

*Comtool* est l'outil permettant d'enregistrer les informations sur les commutations. La base de données extraite de cet outil met à notre disposition la chronologie des différents statuts par lesquels chaque projet de commutation passe. C'est-à-dire le nom du statut ainsi que la date à laquelle le projet est entré dans ce statut.

Le nombre de portefeuilles gérés par AXA LM depuis sa création en 2001 jusqu'à juin 2019 est de 27 comprenant les portefeuilles AXA et les portefeuilles investisseurs (c'est-à-dire appartenant aux fonds d'investissement). Certains portefeuilles sont complètement liquidés ou vendus, d'autres sont encore en cours de gestion. Dans la même période, le nombre de projets de commutation initiés s'élève à 4316, dont :

- 4157 initiés par AXA LM soit la plus grande partie des projets (96%)
- 82 initiés par les cédantes (2%)
- 77 initiés par une tierce partie (2%)

Les projets de commutation initiés sont à différents stades d'évolution : certains sont déjà achevés, d'autres projets sont en cours, mais il y a aussi des projets annulés. La répartition par dernier statut constaté de l'ensemble de ces projets est illustrée par la FIGURE 1.2.

Le libellé complet de chacun des statuts de la FIGURE 1.2 est indiqué dans la TABLE 1.1.

Répartition des projets par dernier statut constaté à juin 2019

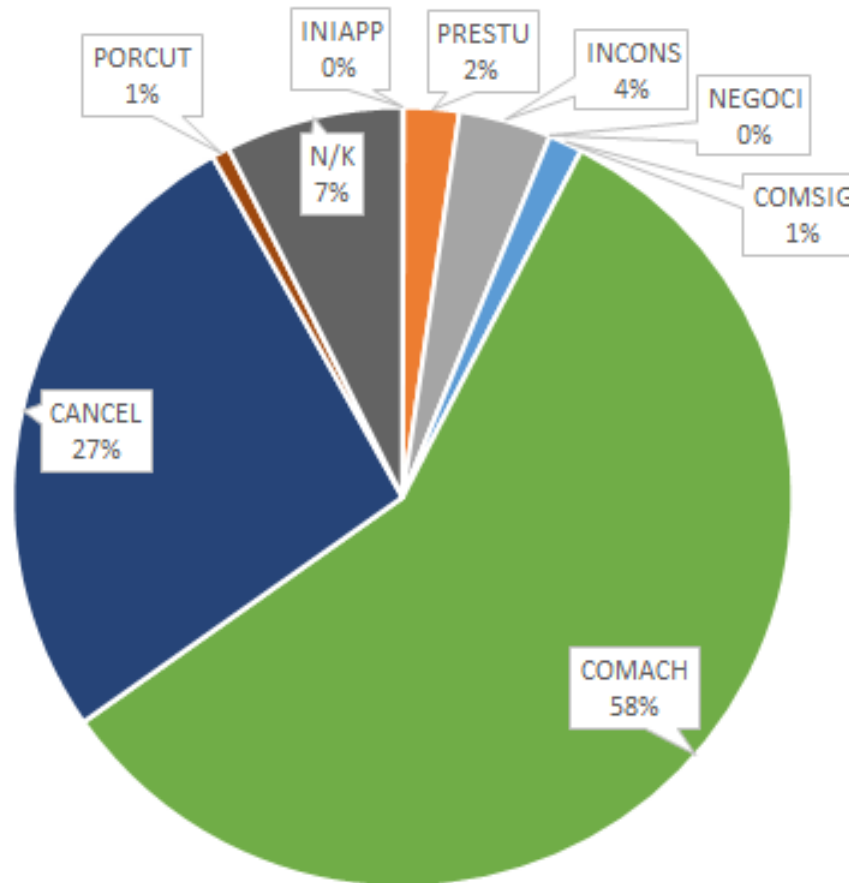


FIGURE 1.2 – Répartition des projets de commutation par dernier statut constaté  
Date : fin juin 2019

<i>Code Statut</i>	<i>Statut original (Comtool)</i>	<i>Code Statut</i>	<i>Statut original (Comtool)</i>
SCODEF	Scope Definition	COMPRE	Post Commutation in Preparation
INIAPP	Initial Approach	COMACH	Commutation Achieved
PRESTU	Pre-Study	COMSIG	Commutation Signed
FIGREC	Figures Reconciliation	CANCEL	Cancelled
INCONS	In Consultation	PORCUT	Portfolio Withdrawal / Cut-Off
NEGOCI	Negotiation in Course	N/K	N/K

TABLE 1.1 – Libellé des statuts de commutation



## Détection des statuts non utilisés

Les responsables commutation confirment que certains statuts ne sont plus utilisés. Afin de détecter ces statuts inutilisés, il est plus aisé d'agréger les dates des différents statuts au niveau annuel. Ceci permettra, en effet, de détecter une baisse soudaine de la fréquence annuelle d'un statut particulier.

La FIGURE 1.3 fait un zoom sur les cinq dernière années d'évolution de l'ensemble des statuts.

Ensuite, nous avons retracé dans le temps, l'évolution de chaque statut de commutation afin de détecter les statuts dont l'utilisation s'arrête et à partir de quelle année.

### *Exemple :*

Sur la FIGURE 1.4, nous constatons que le statut « Initial Approach » (en bleu clair) est très peu utilisé durant l'année 2015 ensuite n'est plus utilisé à partir de 2016 alors que le statut « Commutation Achieved » (en violet) continue d'être utilisé après 2015. L'utilisation de trois autres statuts s'arrête à la même année (2015). Il s'agit des statuts suivants :

- Scope Definition
- Figures Reconciliation
- Post Commutation in Preparation

Après plusieurs échanges avec l'équipe chargée des commutations au sein d'AXA LM, nous sommes parvenus à établir les correspondances entre les statuts qui ont cessé d'être utilisés et leurs équivalents parmi les statuts qui continuent d'être utilisés après 2015. La TABLE 1.2 récapitule ces correspondances.

<i>Code Statut</i>	<i>Statut original (Comtool)</i>	<i>Statut modifié</i>
SCODEF	Scope Definition	Pre-Study
INIAPP	Initial Approach	Pre-Study
PRESTU	Pre-Study	Pre-Study
FIGREC	Figures Reconciliation	In Consultation
INCONS	In Consultation	In Consultation
NEGOCI	Negotiation in Course	Negotiation in Course
COMPRE	Post Commutation in Preparation	Commutation Achieved
COMACH	Commutation Achieved	Commutation Achieved
COMSIG	Commutation Achieved	Commutation Achieved
CANCEL	Cancelled	Cancelled
PORCUT	Portfolio Withdrawal / Cut-Off	Non-Significant - To be excluded
N/K	N/K	Non-Significant - To be excluded

TABLE 1.2 – Correspondance entre statuts dans Comtool et statut retenu

La colonne « Statut original (Comtool) » de la TABLE 1.2, affiche chacun des statuts tels qu'ils apparaissent, dans l'outil de gestion des commutations (Comtool). La colonne « Statut modifié » contient le nouveau statut attribué suite à la consultation de l'équipe commutations pour les statuts qui ne sont plus utilisés à partir de 2015.

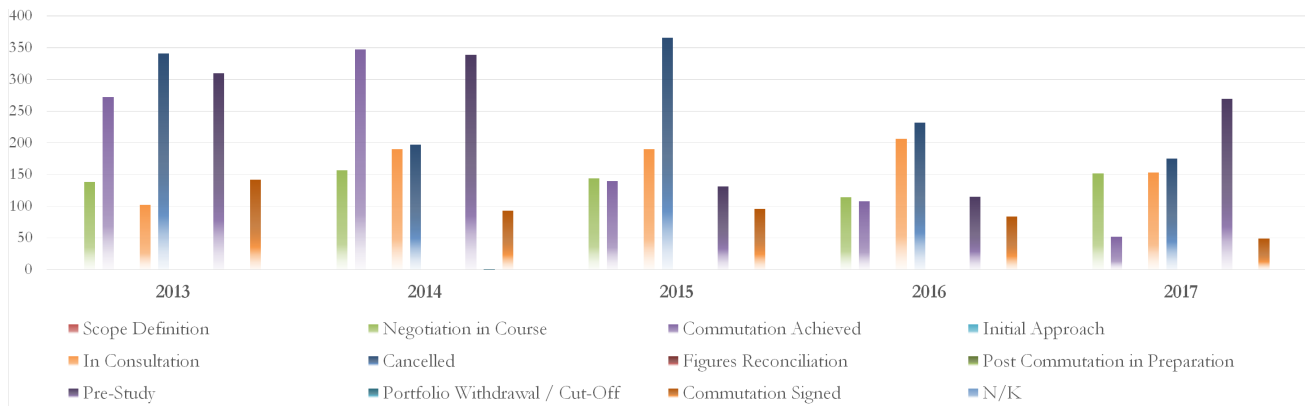


FIGURE 1.3 – Nombre de statuts de commutation par an



FIGURE 1.4 – Comparaison de l'évolution de deux statuts de commutation  
 Pour chaque année, le nombre de projets ayant le statut représenté est affiché au dessus du rectangle

Si le statut continue d'être utilisé après 2015, nous le gardons tel quel. Le statut « Portfolio Withdrawal / Cut-Off » a rarement été utilisé et ne correspond à aucun autre statut alors que le statut « N/K » serait une anomalie. Ces deux statuts (« Portfolio Withdrawal / Cut-Off » et « N/K ») représentent moins de 2% de l'ensemble des données, nous pouvons les écarter de l'étude.

Quant aux statuts retenus, nous proposons de les ordonner chronologiquement dans la mesure du possible. Cela nous offrira par la suite la possibilité d'utiliser un modèle ordinal dans la régression logistique. Cet ordre chronologique se présente comme suit :

① **Pre-Study** (*PRESTU*) : Premier contact et rapprochement de la cédante pour savoir son appétit à commuter, définition du périmètre des contrats à commuter, et études préalables sans proposer de prix.

② **In Consultation** (*INCONS*) : Réconciliation des chiffres AXA et comparaison avec ceux de la cédante, analyse plus complète, définition d'une fourchette de prix en interne (actuariat) qui servira comme base pour l'équipe commutation.

③ **Negotiation in Course** (*NEGOCI*) : L'équipe commutation définit un prix sur la base de la proposition de l'actuariat et négocie sur cette même base avec la cédante.

④ **Commutation Achieved** (*COMACH*) : Fixation du prix après les négociations, signature du contrat de commutation et transfert des fonds (ce transfert est désigné par le terme *post-commutation*).

⑤ **Cancelled** (*CANCEL*) : Ce statut peut intervenir à n'importe quel moment, notamment à cause d'un désaccord avec la cédante sur le périmètre, le prix ou tout autre élément de la commutation. Ainsi, le positionnement chronologique en ⑤ est arbitraire.

Le statut *CANCEL* (annulé) prend la valeur ⑤ mais peut intervenir après chacun des statuts précédents. Par exemple, nous pouvons nous retrouver avec un projet pour lequel les études préalables et les consultations sont finalisées (statuts ① et ② finalisés) et mêmes des négociations sont entamées (statut ③) alors que la commutation n'aboutit pas (statut ⑤), d'où un passage direct du statut ② au statut ⑤ ou, si les négociations sont entamées, un passage du statut ③ au statut ⑤.

Dans certains cas, nous pouvons avoir des passages d'un statut *CANCEL* ⑤ à un statut qui peut être interprété comme antérieur vu que son ordre chronologique est plus bas. Par exemple, nous pouvons avoir un projet annulé (statut ⑤), qui, quelques années plus tard passe au statut *COMACH* (statut ④) parce que la cédante a été relancée et que cette fois-ci elle voudrait finaliser la commutation.

L'analyse des passages d'un statut à un autre fait ressortir plusieurs cas où le projet ne passe pas par tous les statuts. Ce ne sont pas nécessairement des projets annulés. Il se peut, par exemple, qu'un projet passe par le statut ② (*INCONS*) sans avoir été préalablement passé par le statut ① (*PRESTU*). Après l'étape ② (*INCONS*), le même projet pourrait directement passer au statut ④ (*COMACH*) sans passer par le statut intermédiaire ③ (*NEGOCI*).

Nous remarquons que le volume des commutations initiées est beaucoup plus élevé dans l'intervalle de temps allant de 2010 à 2018 en comparaison avec l'intervalle allant de 2001 à 2009.

Cela est dû, entre autres, à la mise en place progressive d'une stratégie de commutation comme solution de finalité après la mise en run-off des portefeuilles AXA de réassurance.

En réassurance active, les traités de réassurance peuvent prévoir des clauses de commutation mais cela reste rare. Cela explique en partie le peu de commutations réalisées avant la mise en run-off des portefeuilles AXA de réassurance et la création d'AXA LM en 2001.

Par ailleurs, à partir de 2010, si nous excluons les portefeuilles des fonds d'investissement, nous obtenons moins de commutations en nombre. Cela veut dire que les portefeuilles des fonds d'investissement impactent directement les commutations parce qu'ils sont au cœur de la stratégie de commutation. Cela s'explique aussi par le degré de maturité atteinte par les portefeuilles en run-off du Groupe AXA, les commutations deviennent de plus en plus difficiles à réaliser.

### Passages possibles entre les statuts d'un projet de commutation

La FIGURE 1.5 montre les différents passages possibles entre deux statuts pour un projet de commutation donné. Par exemple, un projet de commutation en statut ① en  $t = 0$  peut se retrouver dans les statuts ① (rester dans le même état de départ), ②, ③, ④ ou ⑤ en  $t = 1$ .

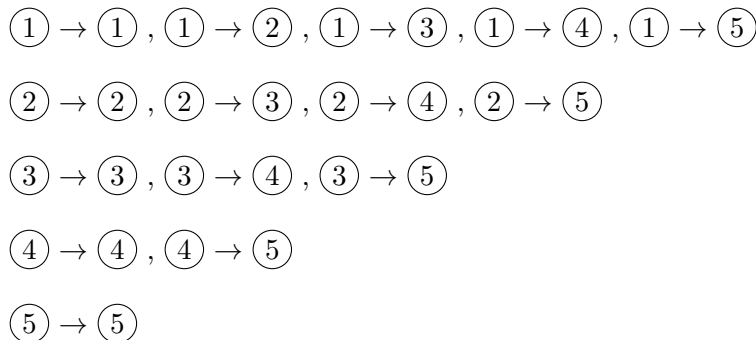


FIGURE 1.5 – Passages possibles entre statuts d'un projet de commutation

### Durées moyennes de passage entre statuts

Afin d'avoir un aperçu global sur les durées de passage d'un statut à un autre, nous avons calculé des moyennes globales sur l'ensemble des projets. Ces moyennes excluent les valeurs aberrantes : si nous supprimons les durées supérieures à 3 ans, nous perdons moins de 5% des données. Le schéma dans la FIGURE 1.6 montre que pour qu'une commutation aboutisse, il faudra environ 18 mois en moyenne. La durée moyenne pour qu'un projet soit annulé est de 12 mois.

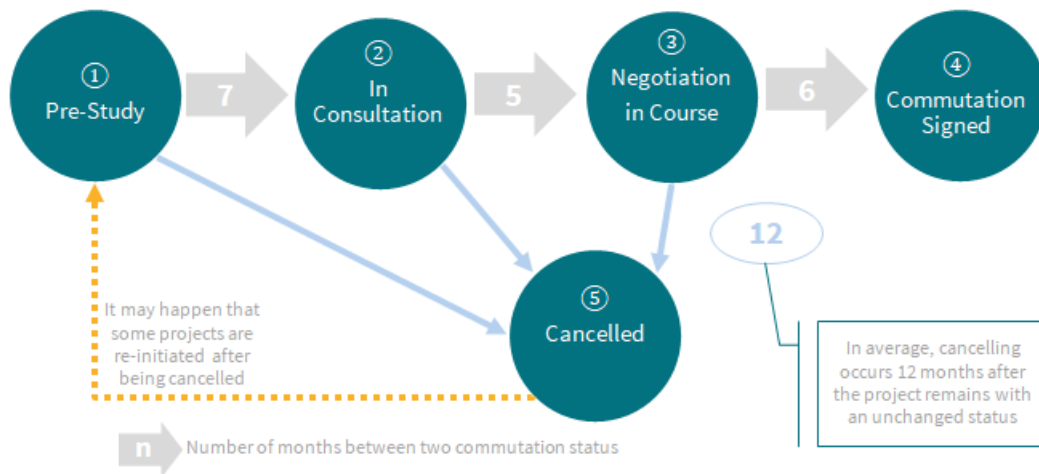


FIGURE 1.6 – Durée moyenne de passage d'un projet de commutation d'un statut à un autre

### 1.2.3 Données sur les contrats

Le système d'information offre la possibilité d'extraire une base de données détaillées par contrat pour chaque portefeuille. Cette base est appelée *IBNR1 Sub*. Une autre base de données permet de faire le lien entre les contrats et les cédantes. Cette base de données est enregistrée dans deux fichiers : un pour les traités proportionnels et un autre pour les traités non proportionnels. Ce dernier contient aussi les limites et les priorités de chaque contrat.

Les extractions *IBNR1 Sub* contiennent les données comptables sur le développement de sinistres par années de souscription. Il s'agit notamment des montants de payés et des montants de provisions dossier/dossier (que nous désignerons par OS : Outstanding). Étant donné que les années de souscription sont très anciennes, certains montants de payés des premières années de développement sont manquants. En effet, cette ancienneté des données entraîne des changements de système d'information et donc des migrations de données ; et il arrive que les montants les plus anciens ne soient pas migrés vers le nouveau système d'information. Ainsi, les payés cumulés ne sont pas fiables à 100%.

## 1.3 Création de pas de temps pour les transitions

Les données sur les commutations sont organisées par dossier ou projet de commutation. Dans le tableau des données brutes (dont la TABLE 1.3 montre un extrait), apparaissent les statuts de commutation et leurs dates pour chaque projet de commutation.

IDFILE	DATE STATUT	NOM STATUT	CODE STATUT
122	01/01/2009	Pre-Study	1
122	05/02/2010	In Consultation	2
122	17/03/2011	Negotiation in Course	3
122	05/07/2013	Cancelled	5
646	01/01/2009	Pre-Study	1
646	12/06/2009	In Consultation	2
646	25/09/2012	Cancelled	5
646	02/01/2014	Pre-Study	1
646	17/08/2015	Cancelled	5
653	13/11/2009	In Consultation	2
653	21/02/2012	Pre-Study	1
653	05/07/2013	Cancelled	5
977	17/12/2012	Negotiation in Course	3
977	20/12/2012	Commutation Achieved	4
1321	01/01/2009	Pre-Study	1
1321	07/09/2012	Cancelled	5
1321	18/02/2013	Pre-Study	1
1321	20/02/2013	In Consultation	2
1321	17/06/2013	Negotiation in Course	3
1321	16/12/2013	Commutation Achieved	4
3053	08/06/2010	Pre-Study	1
3053	18/12/2012	Cancelled	5
3053	11/07/2013	Pre-Study	1
3053	07/12/2016	Cancelled	5
3053	20/01/2017	Pre-Study	1
3053	12/05/2017	In Consultation	2
3053	24/05/2017	Negotiation in Course	3
3090	01/01/2009	Pre-Study	1
3090	01/03/2009	In Consultation	2
3090	01/04/2011	Pre-Study	1
3090	24/08/2011	Negotiation in Course	3
3090	27/02/2013	Commutation Achieved	4
3102	22/06/2012	Pre-Study	1
3102	13/06/2013	Cancelled	5
3102	08/07/2015	In Consultation	2
3102	19/10/2015	Negotiation in Course	3
3102	18/12/2015	Commutation Achieved	4
3105	19/01/2011	Pre-Study	1
3105	22/12/2014	Cancelled	5
3141	27/10/2010	Pre-Study	1
3141	11/04/2011	In Consultation	2
3141	29/04/2011	Negotiation in Course	3
3141	17/02/2012	Cancelled	5

TABLE 1.3 – Extrait des données d'évolution des statuts de commutation par date

Sur la TABLE 1.4 se trouve un extrait de la transformation appliquée sur les données datées. Le tableau affiche le dernier statut du projet de commutation atteint avant la date indiquée en colonne. Les dates en colonnes sont séparées par laps de temps de 1 an (pas annuel). Cette même transformation est réalisée avec des laps de temps de 1 mois, 3 mois et 6 mois pour obtenir respectivement des pas de temps mensuel, trimestriel et semestriel.

IDFILE	Déc2010	Déc2011	Déc2012	Déc2013	Déc2014	Déc2015	Déc2016	Déc2017	Déc2018
122	2	3	3	5	5	5	5	5	5
646	2	2	5	5	1	5	5	5	5
653	2	2	1	5	5	5	5	5	5
977	1	1	4	4	4	4	4	4	4
1321	1	1	5	4	4	4	4	4	4
3043	1	1	5	1	1	1	5	3	3
3090	2	3	3	4	4	4	4	4	4
3102	1	1	1	5	5	4	4	4	4
3104	2	1	1	1	5	5	5	5	5
3141	1	3	5	5	5	5	5	5	5
3201	3	2	5	5	5	5	5	5	5
3331	2	1	5	5	5	5	5	5	5
3401	2	2	2	1	5	5	5	5	5
3421	1	2	2	4	4	4	4	4	4
3444	1	5	5	5	5	5	2	2	5
3648	1	5	5	1	4	4	4	4	4
3770	3	4	4	4	4	4	4	4	4
3808	2	2	2	5	1	5	5	5	5
3840	1	2	4	4	4	4	4	4	4
3845	1	3	3	5	5	5	5	5	5
3852	1	1	1	3	3	5	5	5	5
3853	1	1	1	3	5	5	5	5	5
3907	1	2	5	5	5	5	5	5	5
3923	1	1	1	2	1	1	5	5	5
4015	2	4	4	4	4	4	4	4	4
4105	1	4	4	4	4	4	4	4	4
4123	1	1	3	3	4	4	4	4	4
4149	2	2	4	4	4	4	4	4	4
4150	1	1	5	1	4	4	4	4	4
4153	3	3	1	5	1	1	5	5	5
4182	1	4	4	4	4	4	4	4	4
4259	2	1	1	5	5	5	5	5	5
4314	2	3	2	3	4	4	4	4	4
4375	3	3	1	5	5	5	5	5	5
4413	1	3	5	5	5	5	5	5	5
4421	1	1	2	5	5	5	5	5	5
4433	1	3	3	4	4	4	4	4	4
4449	2	3	4	4	4	4	4	4	4
4461	1	2	5	5	5	5	5	5	5
4480	1	4	4	4	4	4	4	4	4
4493	1	1	1	2	2	1	5	5	5
4503	2	3	4	4	4	4	4	4	4
4510	2	4	4	4	4	4	4	4	4
4597	1	3	3	5	5	5	5	5	5

TABLE 1.4 – Création de pas de temps annuel



## 1.4 Conclusion

Les portefeuilles en run-off du *Groupe AXA* ont atteint une certaine maturité et les commutations sur ces portefeuilles sont de plus en plus difficiles à réaliser. Les portefeuilles des fonds d'investissement présentent plus de potentiel pour des commutations à venir. Ce mémoire vise à tirer profit de l'expertise d'*AXA LM* en gestion de run-off en analysant les données historiques sur le processus de commutation.

Le raisonnement en termes de passage, d'un statut à un autre, pour un projet de commutation a l'intérêt d'analyser dans le détail chacune des étapes pour déceler les éventuels freins à commuter à une étape donnée et proposer en conséquence des améliorations. Par exemple, si un projet de commutation reste bloqué longtemps dans le statut *INCONS* (In Consultation), quelles seraient les mesures à prendre afin de réduire la durée de ce type de blocage ou de l'éliminer ?

Les données disponibles concernent notamment les cédantes, les contrats et les projets de commutation. La création de pas de temps pour les passages des projets de commutation d'un statut à un autre, permettra d'appliquer un modèle de Markov qui fera l'objet du chapitre 2.

L'ajout des caractéristiques de cédantes et de contrats au modèle permettra de réaliser des segmentations pertinentes (par exemple, par périodes de souscription, ou par type de contrats). Ces segmentations aideront à la compréhension du processus actuel de commutation et à proposer des axes d'amélioration. La prise en compte de ces caractéristiques fera l'objet du chapitre 3, et ce, par l'application d'un modèle logistique.



## Chapitre 2

# Processus de commutation modélisé par chaînes de Markov

Ce chapitre est consacré à l'application d'un modèle simplifié de chaînes de Markov aux données de commutation.

Les chaînes de Markov constituent un formalisme qui permet d'analyser et de caractériser des systèmes dynamiques dans lesquels des éléments évoluent d'état en état [1].

Elles ont un très grand nombre d'applications aussi bien dans la recherche que dans l'industrie. Elles sont souvent considérées comme un instrument non seulement utile mais même indispensable pour la caractérisation de systèmes hautement multiparamétriques. Elles font donc partie des méthodes mises en œuvre dans le cadre du big data. Leur application sans doute la plus emblématique est la fameuse matrice *Google* utilisée par la firme pour hiérarchiser les pages du web. C'est l'idée de cette matrice qui a permis la création d'un moteur de recherche révolutionnaire à l'époque et donc la naissance de *Google* [6].

Pour commencer ce chapitre, quelques rappels théoriques sur les chaînes de Markov seront présentés dans la première section.

Ensuite, les vérifications de la *propriété de Markov* sur les données seront détaillées, dans la deuxième section, selon deux variantes : la première portera sur l'ensemble des données tous projets confondus ; la deuxième sera appliquée sur les données projet par projet.

La troisième section synthétise les estimations des différentes matrices de transition selon plusieurs pas de temps ainsi que leurs commentaires.

Intuitivement, une fois qu'un projet de commutation passe à une étape donnée, il est très rare de revenir à une étape antérieure. Nous verrons si la matrice de transition estimée confirme cette intuition ou non.

Par ailleurs, la matrice de transition estimée apportera une information sur la proportion de projets de commutation qui ont été réalisés avec succès et l'intervalle de temps correspondant. Nous tenons à préciser que cette analyse reste encore globale et ne tient pas compte des différences relatives aux caractéristiques des projets de commutation, ni des cédantes sous-jacentes.

Enfin, la notion d'état stationnaire sera introduite dans la quatrième et dernière section de ce chapitre. L'estimation de cet état stationnaire sera réalisée et interprétée.

## 2.1 Rappels théoriques sur les chaînes de Markov

Les chaînes de Markov sont composées [6] :

- d'éléments. Les éléments sont les entités qui évoluent à l'intérieur du système. Il peut s'agir de personnes, de molécules, de végétaux, de biens manufacturés, d'argent, etc. Dans le cadre de ce mémoire, il s'agit de projets de commutations
- d'états. Les états sont les différentes localisations ou les différentes formes, qualités, etc., atteintes successivement par les éléments. Dans notre étude, les différents états sont les statuts décrits dans le chapitre précédent (section 1.2.2).
- de probabilités de transitions. Une transition se définit comme le passage d'un état à un autre. Les éléments se trouvant dans un état donné ont des probabilités données de rejoindre d'autres états ou de rester dans le même. Ces probabilités dépendent uniquement de l'état initial et de l'état final de la transition, pas de l'élément. Elles ne varient pas avec le temps.
- d'un pas de temps. Un pas de temps est un laps de temps constant qui se répète. A chaque pas de temps, l'élément peut, ou non, passer d'un état à un autre. Quand le flux de passage d'un état à un autre est continu, le formalisme s'applique toujours ; il suffit alors de considérer que le pas de temps devient infiniment court. Dans notre cas, plusieurs pas sont définis, mensuel, trimestriel, semestriel et annuel. Nous retiendrons le pas le plus adapté aux données et à la nature de l'activité de commutation.

Une propriété fondamentale des chaînes de Markov est que leurs éléments doivent être *sans mémoire* ; c'est-à-dire que, lorsqu'ils ont atteint un certain état, les probabilités d'évoluer vers un autre état ne dépendent pas de leur passé [6].

Les processus en temps discret à valeurs discrètes ont été utilisés par Markov en 1913 pour modéliser la fréquence d'apparition des consonnes et des voyelles dans l'œuvre « *Eugene Onegin* » composée par Alexandre Pouchkine. Ainsi, sur une suite de 20000 caractères, il estime les probabilités suivantes :

$$\begin{pmatrix} 12.8\% & 87.2\% \\ 66.3\% & 33.7\% \end{pmatrix}$$

C'est-à-dire qu'en russe, la probabilité qu'une voyelle soit suivie d'une consonne est de 87.2%.

Les définitions et rappels théoriques dans la présente section sont extraits de l'ouvrage [5] traitant des chaînes de Markov et simulations.

**Définition 2.1.** (Processus aléatoire). *Un processus aléatoire  $(X_n)_n$  à temps discret est une famille  $\{X_n, n \in \mathbb{N}\}$  de variables aléatoires indicées par des entiers positifs. Il est caractérisé par la loi de chacun des vecteurs  $(X_0, \dots, X_n)$  pour tout  $n \in \mathbb{N}$ . Les valeurs sont dans un ensemble  $E$  fini ou dénombrable.*

**Définition 2.2.** (Chaîne de Markov). *Un processus aléatoire  $(X_n)_n$  à temps discret est une chaîne de Markov homogène si pour tout  $n \in \mathbb{N}$ , on a :*

$$\forall i_0, i_1, \dots, i_{n-2}, i, j \in E,$$

$$\mathbb{P}(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \dots, X_0 = i_0) = \mathbb{P}(X_n = j | X_{n-1} = i) = p_{ij}$$

*La probabilité  $p_{ij}$  représente la probabilité de transition du processus vers l'état  $j$  sachant qu'il est dans l'état  $i$ .*

**Remarque 2.1.** (Une autre formulation de la Définition 2.2). *La loi de  $X_n$  ne dépend que de la valeur de  $X_{n-1}$  et non de la totalité du passé  $X_{n-1}, X_{n-2}, \dots, X_0$ . Cette propriété est appelée **propriété de Markov**.*

**Remarque 2.2.** *Cette chaîne de Markov est homogène car la probabilité  $\mathbb{P}(X_n = j | X_{n-1} = i) = p_{ij}$  est indépendante de l'état  $n$ .*

**Remarque 2.3.** *Puisque les  $p_{ij}$  sont des probabilités et que le processus doit nécessairement aller dans un état sachant qu'il est en  $i$ , on a :*

$$\forall i, j \in E, p_{ij} \in [0, 1] \text{ et } \forall i \in E, \sum_{j \in E} p_{ij} = 1$$

*La matrice  $P = (p_{ij})_{ij}$  est appelée **matrice de transition**.*

**Exemple 2.1.** (avec trois états). *Considérons une chaîne de Markov avec la matrice de transition suivante*

$$P = \begin{pmatrix} 0.25 & 0.25 & 0.5 \\ 0.67 & 0.33 & 0 \\ 0 & 0.4 & 0.6 \end{pmatrix}$$

On suppose que l'état initial de la chaîne est 1. La FIGURE 2.1 (a) représente un exemple de trois trajectoires de la chaîne, tandis que la FIGURE 2.1 (b) est le diagramme de la chaîne, i.e. une représentation graphique de la matrice de transition.

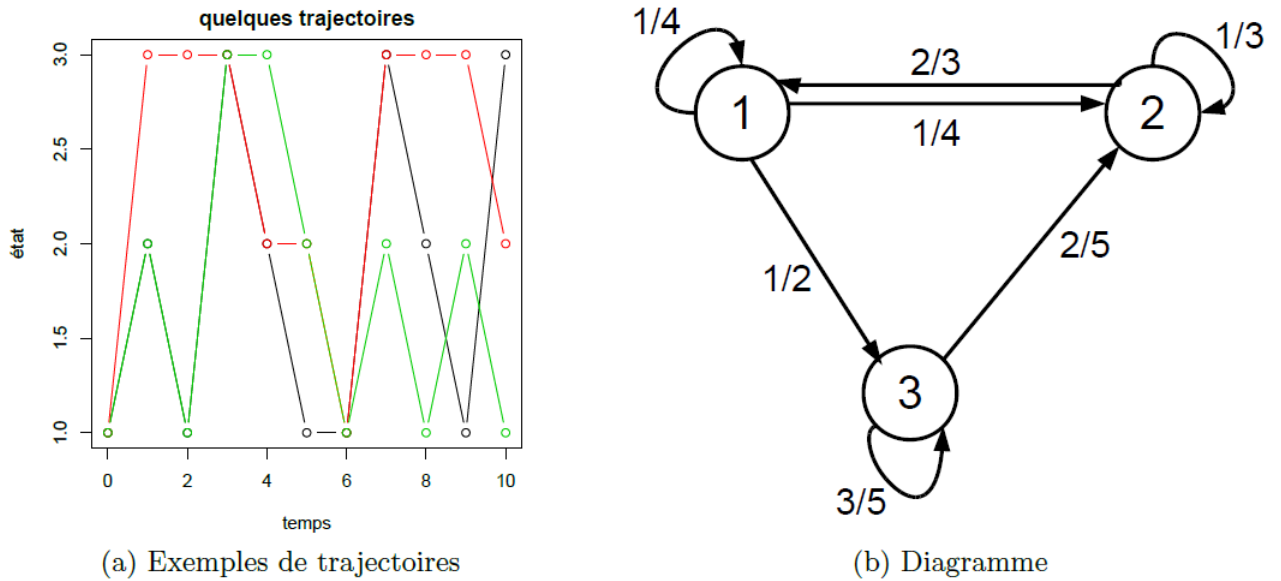


FIGURE 2.1 – Exemple de chaîne de Markov  $3 \times 3$

**Exemple 2.2.** (*bon/mauvais conducteur*). Un modèle simplifié de bonus/malus en assurance est de considérer une chaîne de Markov à deux états  $\{B; M\}$ , respectivement pour bon et mauvais conducteur. Si à la date  $t$  le conducteur est dans l'état  $B$ , alors il y reste avec une probabilité  $\alpha$ , tandis que s'il est dans l'état  $M$ , il y reste avec une probabilité  $\beta$ . La matrice de transition est donc

$$P = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix}$$

Le diagramme est donné en FIGURE 2.2.

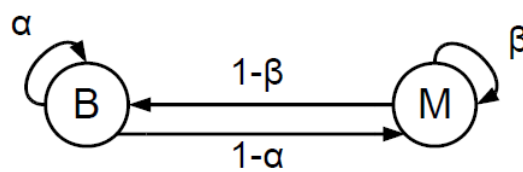


FIGURE 2.2 – Exemple de chaîne de Markov  $2 \times 2$

## 2.2 Vérification de la propriété de Markov

L'objectif de cette section est de vérifier la propriété de Markov dite aussi d'absence de mémoire à l'aide d'un test de *Khi-deux*. La vérification est faite projet par projet de commutation. Sous il convient d'utiliser la fonction `VerifyMarkovProperty` du package `markovchain`.

Intuitivement, une commutation passe par plusieurs stades. Le premier en général étant l'étude initiale. Elle évolue ensuite dans plusieurs autres stades jusqu'à une éventuelle signature de l'accord de commutation ou alors une annulation du projet de commutation. Dès que la commutation

atteint un stade particulier, on peut considérer que toute l'information sur les stades précédents est incluse dans le stade que la commutation aura atteint. Par exemple, si la commutation atteint le stade de négociation avec la cédante, elle est déjà passée par le stade de réconciliation des données entre les différentes sources possibles et cette information est connue et incluse dans la négociation. Cela répond à la propriété de Markov qui stipule que les états futurs ne dépendent que de l'état actuel qui, lui, contient toute l'information déjà connue dans les états précédents.


Cependant, afin de pouvoir utiliser les conclusions qui découlent de la *propriété de Markov* dite aussi d'*absence de mémoire*, il convient de procéder à une validation statistique de cette propriété. Le test de Khi-deux d'absence de mémoire répond à ce besoin :


$$\begin{cases} H_0 : \text{Propriété de Markov vérifiée} \\ H_1 : \text{Propriété de Markov non vérifiée} \end{cases}$$

L'hypothèse nulle du test stipule que la propriété de Markov est vérifiée. Ainsi, une *p-value* inférieure à un certain seuil que l'on pourrait se fixer (le seuil de 5% étant généralement admis dans la communauté scientifique) indique que l'on pourrait rejeter l'hypothèse nulle. Donc, plus la *p-value* est proche de 0, plus on sera orienté vers un rejet de la propriété de Markov. À l'inverse, plus la *p-value* dépasse le seuil choisi (5%) et se rapproche de 1, plus il sera légitime de ne pas rejeter l'hypothèse nulle d'absence de mémoire et donc de valider la propriété de Markov.

### 2.2.1 Validation globale de la propriété de Markov sur l'ensemble des données

Dans ce paragraphe, le test de l'hypothèse d'absence de mémoire (*propriété de Markov*) est effectué sur l'ensemble des données. C'est-à-dire que nous considérons, en quelques sortes, que l'ensemble des données disponibles, pour l'ensemble des projets de commutation, constitue une seule chaîne de Markov.


Pour créer cette unique chaîne, une concaténation des données détaillées projet par projet est réalisée à l'aide de , après lecture des données (voir ANNEXE A.1.1).

Ainsi, il est possible de réaliser le test de vérification de la propriété de Markov en utilisant la fonction **VerifyMarkovProperty**. Les résultats détaillés sous  de cette vérification sont en ANNEXE A.1.2.

Le  $\chi^2$  calculé par le test est de 2694.599 pour une *p-value* de 0. Au vu de ces résultats (*p-value* nulle), l'hypothèse  $H_0$  d'absence de mémoire est rejetée, la propriété de Markov pour l'ensemble des données n'est pas vérifiée. Ainsi, en concaténant les transitions de l'ensemble des données, nous n'obtenons pas une seule chaîne de Markov.

### 2.2.2 Validation de la propriété de Markov projet par projet

Dans ce paragraphe, le test envisagé est différent du précédent par le fait que la propriété de Markov sera testée ligne par ligne, *i.e.* projet de commutation par projet de commutation.

La même fonction de  est appliquée mais cette fois-ci, elle est itérée sur les données pour ne vérifier la propriété de Markov que pour un projet de commutation à la fois avant de passer au projet suivant.

Les résultats de cette vérification sont en ANNEXE A.1.3. Vu le grand nombre de tests à réaliser (plus de 5000), nous nous limitons à afficher en ANNEXE A.1.3 un extrait de quelques lignes seulement.

Les valeurs calculées de la statistique du  $\chi^2$  par le test sont très petites. Dans l'extrait en ANNEXE A.1.3, ces valeurs sont comprises entre 0.02 et 1.03 pour une *p-value* systématiquement égale à 1. La *p-value* pour le reste des projets est égale à 1 ou très proche de 1. Au vu de ces résultats, l'hypothèse  $H_0$  d'absence de mémoire ne peut pas être rejetée, et donc la propriété de Markov peut être considérée comme étant validée pour chacun des projets de commutation.

## 2.3 Estimation de la matrice de transition

### 2.3.1 Estimateur de la matrice de transition

Soient  $X_t$  les variables aléatoires décrivant l'état d'un projet de commutation à chaque pas de temps  $t$ . À chaque instant  $t$ , la variable  $X_t$  peut prendre la valeur :

- 1 : PRESTU (Pre-Study)
- 2 : INCONS (In Consultation)
- 3 : NEGOCI (Negotiation in Course)
- 4 : COMACH (Commutation Achieved)
- 5 : CANCEL (Cancelled)

La modélisation de ces variables peut être réalisée au moyen de chaînes de Markov homogènes ou non homogènes. La non homogénéité signifie que la matrice de transition dépend de  $t$  (*i.e.* du moment du lancement d'un projet de commutation). Étant donné que l'on peut se ramener au cas homogène, nous nous intéresserons dans notre étude uniquement au cas homogène. La non homogénéité sera prise en compte, au chapitre 3, par l'introduction d'une variable indiquant le moment de l'observation du statut de départ d'un projet de commutation donné.

Les éléments de la matrice de transition sont estimés par la formule suivante :

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i.}} = \frac{\sum_{t=0}^{T-1} n_{ij}(t)}{\sum_{t=0}^{T-1} n_{i.}(t)} = \frac{\sum_{t=0}^{T-1} n_{ij}(t)}{N \sum_{k=1}^{T-1} \sum_{t=0}^{T-1} n_{ik}(t)}$$

où :

- $n_{i.}(t)$  le nombre de fois où le projet de commutation est dans l'état  $i$  à la date  $t$  ( $t = 0, 1, \dots, T - 1$ )
- $n_{ij}(t)$  le nombre de fois où le projet de commutation est passé de l'état  $i$  à la date  $t$  à l'état  $j$  à la date  $t + 1$  ( $t = 0, 1, \dots, T - 1$ )



### 2.3.2 Calcul sous $\mathbb{R}$ des matrices de transition

Partant de l'hypothèse d'absence de mémoire validée dans la section précédente pour chaque projet de commutation, une matrice de transition pourra être estimée. Sous  $\mathbb{R}$ , la fonction `markovchainFit` permet d'estimer la matrice de transition à partir des données observées.

L'objectif de cette sous-section est d'estimer les probabilités de transition de la matrice ainsi que des intervalles de confiance. Ces derniers ne sont pas présentés dans cette section, mais mis en annexe, et ce, afin de faciliter la lecture des matrices de transition. L'annexe correspondante est indiquée pour chaque matrice.

Dans la suite de cette sous-section, les matrices estimées pour chacun des pas de temps créés au chapitre 1 seront présentées ainsi que les commentaires sur les probabilités de transition. Pour chacun des pas de temps, deux matrices seront affichées : une première matrice pour les 5 états possibles de commutation, et une deuxième matrice en éliminant les projets avec statut « annulé ». Cette dernière nous renseignera sur les transitions des processus qui permettent d'aboutir à un succès du projet de commutation (statut ④).

*Estimation sous  $\mathbb{R}$  de la matrice de transition mensuelle* (voir Annexe A.3.1)

Cas 5 états

$$P_5^M = \begin{pmatrix} 0.957 & 0.009 & 0.006 & 0.006 & 0.020 \\ 0.003 & 0.973 & 0.009 & 0.005 & 0.010 \\ 0.002 & 0.003 & 0.928 & 0.047 & 0.020 \\ 0.002 & 0.001 & 0.000 & 0.997 & 0.000 \\ 0.003 & 0.001 & 0.000 & 0.000 & 0.996 \end{pmatrix}$$

Comme pour les exemples fournis en section 2.1, la lecture de la matrice de transition se fait ligne par ligne. La première ligne de la matrice de transition mensuelle estimée fait apparaître 95.7% de projets qui restent dans le même état (statut ①) pendant une période d'un mois. Les autres éléments de la même ligne indiquent que :

- 0.9% des projets passent du statut ① au statut ②
- 0.6% des projets passent du statut ① au statut ③
- 0.6% des projets passent du statut ① au statut ④ donc sont commutés un mois plus tard
- 2% des projets passent du statut ① au statut ② donc sont annulés au bout d'un mois

Les autres lignes se lisent de façon similaires. Cependant, nous remarquons que les valeurs de la diagonale sont proches de 1, notamment les deux dernières valeurs de la diagonale (99.7% et 99.6%). Ces deux valeurs peuvent être interprétées par le fait que les statuts ④ (commuté) et ⑤ (annulé) sont des états absorbants, c'est-à-dire qu'un projet commuté ne peut changer de statut, et idem pour un projet annulé.

Le constat global sur cette matrice est que les valeurs en dehors de la diagonale sont très proches de zéro. Ceci, s'explique par le fait que, durant une période aussi courte qu'un mois, il ne peut pas y avoir beaucoup d'évolution dans les projets de commutation, donc les projets ne changent pas souvent de statut durant une période d'un mois.

Enfin, la plus faible valeur de la diagonale se situe en troisième ligne (92.8%). Juste à droite

de cette valeur, on trouve la valeur hors diagonale la plus élevée (4.7%). Ceci pourrait être le signe que le fait d'arriver au statut (3) permet d'accélérer la commutation relativement aux statuts précédents. En effet, 4.7% des projets en statut (3) sont commutés un mois plus tard, contre 0.6% et 0.5% pour les statuts (1) et (2) respectivement.

#### Cas 4 états

$$P_4^M = \begin{pmatrix} 0.964 & 0.012 & 0.008 & 0.016 \\ 0.002 & 0.985 & 0.007 & 0.006 \\ 0.001 & 0.001 & 0.937 & 0.061 \\ 0.002 & 0.000 & 0.000 & 0.998 \end{pmatrix}$$

La matrice estimée en se restreignant à 4 états confirme que le statut (4) (commuté) est absorbant avec une valeur très proche de 1 (99.8%) encore plus élevée que la valeur estimée dans la matrice avec 5 états. Elle confirme aussi le constat que les projets en statut (3) présentent une accélération du passage vers le statut commuté. En effet, 6.1% des projets en statut (3) passent au bout d'un mois vers le statut commuté contre 1.6% et 0.6% pour les statuts (1) et (2).

*Estimation sous  $\mathbb{R}$  de la matrice de transition trimestrielle* (voir Annexe A.3.2)

#### Cas 5 états

$$P_5^T = \begin{pmatrix} 0.890 & 0.019 & 0.015 & 0.018 & 0.058 \\ 0.007 & 0.931 & 0.017 & 0.016 & 0.029 \\ 0.005 & 0.008 & 0.823 & 0.109 & 0.054 \\ 0.004 & 0.002 & 0.001 & 0.993 & 0.001 \\ 0.007 & 0.003 & 0.002 & 0.001 & 0.988 \end{pmatrix}$$

Comparée à la matrice à 5 états avec transition mensuelle, cette matrice trimestrielle montre globalement les mêmes tendances. Cependant, bien que les valeurs sur la diagonale soient proches de 1, elles sont plus faibles de 5 à 10 centièmes sauf les deux dernières valeurs qui restent encore très proches de 1. Ceci confirme encore une fois le fait que les statuts (4) et (5) soient absorbants.

Les faibles valeurs sur la diagonale sont compensées par des valeurs plus élevées en dehors de la diagonale. Ceci indique qu'il existe plus de projets de commutation qui changent de statut. Cela est dû au fait qu'une période d'un trimestre permet plus d'évolution pour un projet donné qu'une période de seulement un mois.

Arrivé au statut (3), un projet de commutation a désormais encore plus de chance d'être commuté (on atteint les 10.9%) dans le trimestre qui suit. Et les statuts (1) et (2) permettent toujours moins de commutation que le statut (3).

#### Cas 4 états

$$P_4^T = \begin{pmatrix} 0.913 & 0.024 & 0.018 & 0.045 \\ 0.005 & 0.964 & 0.013 & 0.019 \\ 0.004 & 0.004 & 0.850 & 0.143 \\ 0.003 & 0.001 & 0.001 & 0.995 \end{pmatrix}$$

La comparaison avec la matrice à 4 états avec transition trimestrielle montre globalement le même comportement des projets de commutation dû à l'allongement de la durée d'analyse (pas de

temps) que pour la matrice à 5 états : plus le pas d'analyse est élevé, plus les valeurs de la diagonale sont faibles et compensées par des valeurs plus élevées en dehors de la diagonale.

La dernière valeur de la diagonale reste encore proche de 1 et la proportion de projets en statut ③ qui passent en commuté un trimestre plus tard s'élève à 14.3%.

*Estimation sous  $\mathbb{R}$  de la matrice de transition semestrielle* (voir Annexe A.3.3)

Cas 5 états

$$P_5^S = \begin{pmatrix} 0.815 & 0.022 & 0.019 & 0.034 & 0.111 \\ 0.012 & 0.888 & 0.017 & 0.030 & 0.055 \\ 0.008 & 0.010 & 0.727 & 0.164 & 0.090 \\ 0.006 & 0.003 & 0.002 & 0.987 & 0.002 \\ 0.011 & 0.006 & 0.003 & 0.004 & 0.977 \end{pmatrix}$$

En transition semestrielle, la tendance est confirmée encore une fois, plus le pas de temps est élevé, plus les valeurs de la diagonale sont faibles. Bien qu'elles soient légèrement plus faibles que les cas précédents, les deux dernières valeurs de la diagonale (98.7% et 97.7%) restent proches de 1 confirmant encore une fois que les statuts commuté et annulé sont absorbants. Désormais, 16.4% des projets en statut ③ passent en commuté au bout d'un semestre.

Cas 4 états

$$P_4^S = \begin{pmatrix} 0.864 & 0.029 & 0.026 & 0.081 \\ 0.007 & 0.943 & 0.012 & 0.037 \\ 0.006 & 0.006 & 0.772 & 0.216 \\ 0.004 & 0.002 & 0.002 & 0.992 \end{pmatrix}$$

Globalement, la tendance est très similaire. La réalisation de ces différentes estimations permet de corroborer la tendance et de la confirmer.

*Estimation sous  $\mathbb{R}$  de la matrice de transition annuelle* (voir Annexe A.3.4)

Cas 5 états

$$P_5^A = \begin{pmatrix} 0.701 & 0.021 & 0.019 & 0.065 & 0.194 \\ 0.018 & 0.846 & 0.014 & 0.046 & 0.082 \\ 0.014 & 0.011 & 0.631 & 0.209 & 0.134 \\ 0.009 & 0.005 & 0.004 & 0.978 & 0.005 \\ 0.016 & 0.008 & 0.005 & 0.010 & 0.961 \end{pmatrix}$$

Le matrice avec transition annuelle à 5 états, affiche les valeurs en diagonale les plus faibles et donc en dehors de la diagonale les plus élevées. La plus grande valeur en dehors de la diagonale (20.9%) correspond à la proportion de projets en statut ③ qui passent en statut commuté. Sur la même ligne on constate la valeur en diagonale la plus faible (63.1%).

Cas 4 états

$$P_4^A = \begin{pmatrix} 0.804 & 0.027 & 0.023 & 0.147 \\ 0.011 & 0.918 & 0.012 & 0.059 \\ 0.010 & 0.010 & 0.691 & 0.289 \\ 0.005 & 0.004 & 0.004 & 0.988 \end{pmatrix}$$

De la même manière que la transition annuelle avec 5 états, les différentes tendances citées précédemment sont confirmées, notamment la proportion de projets en statut (3) qui passent vers le statut commuté est de 28.9%.

Pour toutes les matrices précédentes, les écarts-types et intervalles de confiance à 95% des probabilités de transitions estimées (mensuelle, trimestrielle, semestrielle et annuelle) se trouvent dans chacune des annexes (A.3.1, A.3.2, A.3.3, A.3.4) citées précédemment, sous forme de sorties R.

Pour le pas annuel, les FIGURES 2.3 et 2.4 représentent les transitions sous forme de graphes associés aux matrices de transitions respectivement pour 5 états (incluant l'état (5) CANCEL) et 4 états (excluant l'état (5) CANCEL).

**Annual Transition Probability Diagram for Commutations 5 states**

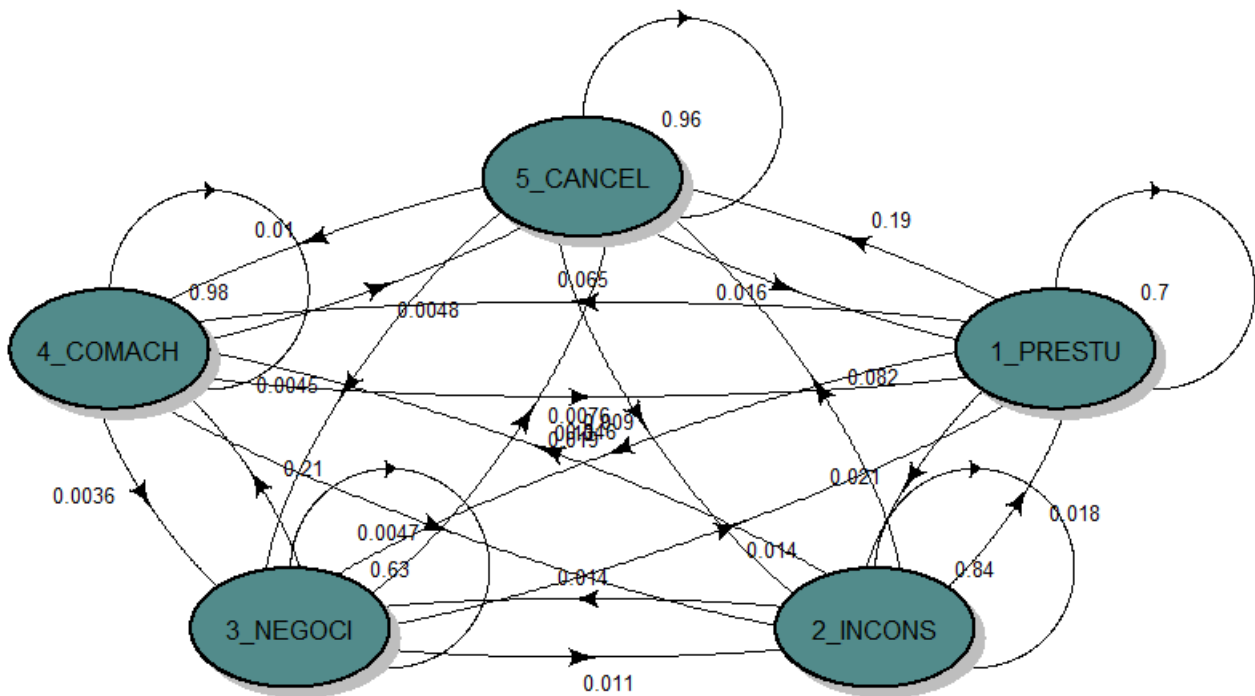


FIGURE 2.3 – Diagramme des transitions annuelles avec 5 états

Annual Transition Probability Diagram for Commutations 4 states

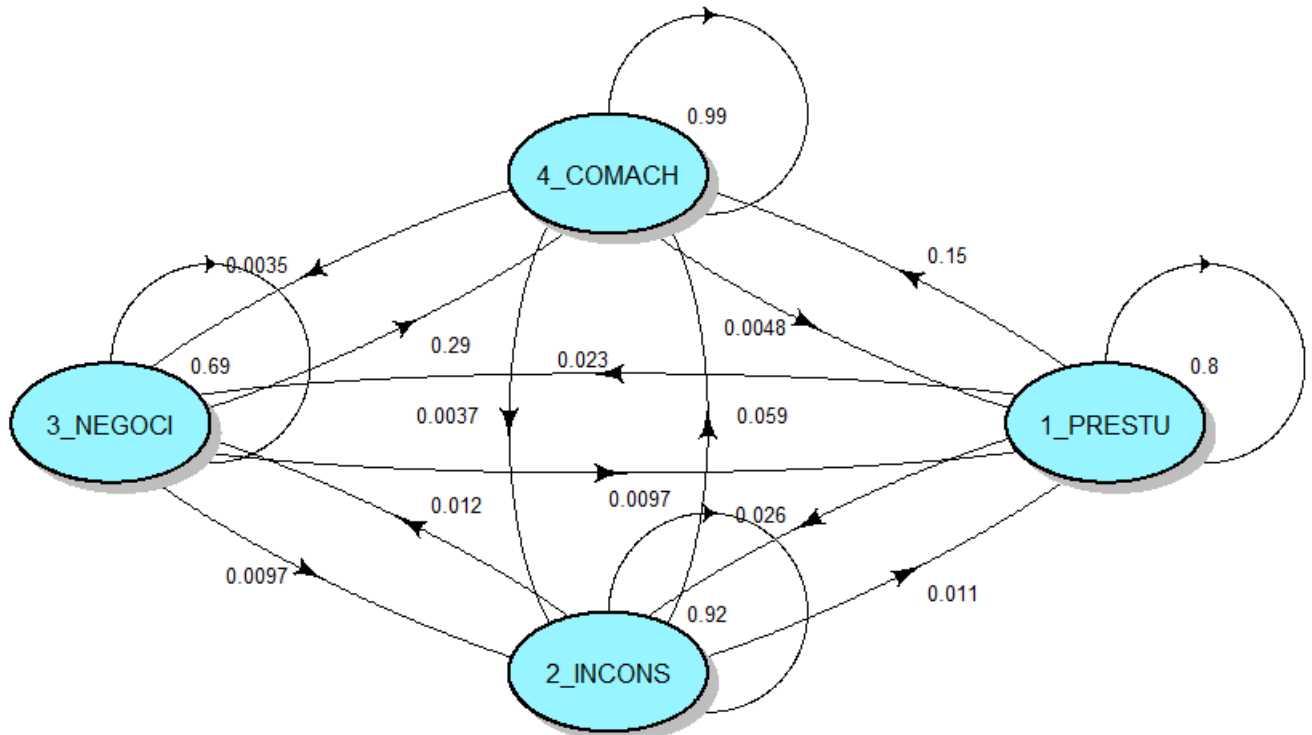


FIGURE 2.4 – Diagramme des transitions annuelles avec 4 états

### 2.3.3 Interprétation des matrices de transition

La matrice de transition estimée confirme quelques résultats *a priori* connus vu la nature même du processus de commutation. Par exemple, les valeurs du triangle inférieur de la matrice sont très proches de zéro. Ceci s’explique par le fait que, un projet de commutation qui a atteint un stade donné, ne devrait plus revenir au stade antérieur. Par ailleurs, les valeurs situées sur la diagonales pour les statuts (4) et (5) sont très proches de 1. En effet, ces deux statuts sont absorbants. C’est-à-dire que, une fois qu’un de ces deux statuts est atteint par un projet de commutation donné, le projet reste à ce stade indéfiniment : lorsqu’un accord de commutation est signé et les transferts de fonds sont effectués, la commutation s’achève (statut (4)), et il n’est plus possible de revenir sur l’accord de commutation. Le même raisonnement peut être appliqué à l’annulation (statut (5)). Dans la pratique, de rares cas se présentent où un projet annulé pourrait être relancé.

Par ailleurs, l’estimation révèle des résultats intéressants pour la suite de l’étude. Les matrices de transition estimées retracent statistiquement les passages d’un statut à un autre. Par exemple, durant un semestre, il n’y a que 18% ( $1 - 0.815$ ) des projets initiés qui changent de statut, les 82% restants demeurent en phase d’initiation (statut (1)). En raisonnant trimestriellement, la proportion des projets initiés et qui changent de statut est de 11% ( $1 - 0.890$ ).

Plus le pas d’analyse est long, moins il y a de projets de commutation qui se maintiennent dans un statut donné. Sauf pour les deux états absorbants (commuté et annulé) qui, de par leur nature, ne permettent pas d’évolution d’un projet de commutation.

Enfin, toutes les matrices affichent un taux plus élevé de projets en statut initial ③ (Negotiation in course) qui passent vers le statut commuté, et ce, en comparaison avec les projets avec statut de départ ① (Pre-Study) ou ② (In Consultation). Le fait d'entrer en négociation avec la cédante permet clairement d'accélérer le processus de commutation et même d'aboutir à un accord de commutation.

## 2.4 État stationnaire

**Définition 2.3.** [2] (Loi de probabilité invariante). On appelle mesure invariante (ou stationnaire ou d'équilibre) de la chaîne de Markov homogène  $(X_n)$  de matrice de transition  $P$  un vecteur  $\mu$  de  $\mathbb{R}_+^d$  (où  $d = \text{Card}(E)$ ) tel que :

$$\mu = \mu P$$

Cela correspond au vecteur propre (à gauche) de la matrice  $P$  pour la valeur propre 1. Si de plus  $\sum_{x \in E} \mu_x = 1$ , alors  $\mu$  est une mesure de probabilité invariante.

**Exemple 2.3.** (Bon/mauvais conducteur). Suite de l'exemple 2.2, cherchons la loi invariante vérifiant donc


$$\begin{pmatrix} \mu_B & \mu_M \end{pmatrix} = \begin{pmatrix} \mu_B & \mu_M \end{pmatrix} \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix} \iff \begin{cases} \mu_B = \alpha \mu_B + (1 - \beta) \mu_M \\ \mu_M = (1 - \alpha) \mu_B + \beta \mu_M \end{cases}$$

Les équations sont liées donc il suffit de garder la première et d'ajouter la condition  $\mu_B + \mu_M = 1$ . On trouve :

$$\mu_B = \frac{1 - \beta}{2 - \beta - \alpha}, \mu_M = \frac{1 - \alpha}{2 - \beta - \alpha}$$

Si on fixe les probabilités  $\alpha = 1/5$  et  $\beta = 2/5$ , on obtient  $(\mu_B, \mu_M) = (0.57142; 0.42857)$ .

### 2.4.1 Estimation de l'état stationnaire

La fonction **SteadyStates** de  permet de calculer l'état stationnaire de la chaîne de Markov. C'est-à-dire la situation dans laquelle chaque projet reste au statut qu'il aura atteint et ne change plus de statut à toutes les périodes suivantes. L'état stationnaire donne la répartition des projets par statut à une période assez lointaine pour permettre une stabilité de la répartition.

Les résultats de l'estimation montrent que, quel que soit le pas de temps retenu pour les transitions, l'état stationnaire est sensiblement le même. En effet, l'état stationnaire estimé donne approximativement, dans tous les cas, la répartition suivante (en incluant les projets annulés) :

$$\text{État stationnaire avec transition annuelle} = (0.040 \quad 0.040 \quad 0.013 \quad 0.509 \quad 0.398)$$

- La moitié (soit 50% ou 51%) des projets de commutation sont achevés (statut ④).
- 2/5 (soit 39% ou 40%) des projets de commutation sont annulés (statut ⑤).

- Les 10% restants des projets sont dans des phases intermédiaires, répartis entre projets initiés (statut ① : 4%), projets en cours de réconciliation (statut ② : 4%) et négociation en cours (statut ③ : 2%).

Selon le *lemme de Kac* (voir [2]), l'inverse de la mesure (état) stationnaire donne le temps moyen pour accéder à l'état correspondant à cette mesure. En appliquant ce lemme, le temps moyen pour atteindre le statut « commuté » est d'environ 2 ans (1/0.509) alors que le temps moyen pour le statut « annulé » est d'environ 2.5 ans (1/0.398).

Pour ce qui est de l'analyse de l'état stationnaire lorsque les projets annulés sont exclus, nous obtenons la répartition suivante :

$$\text{État stationnaire avec transition annuelle} = (0.026 \quad 0.051 \quad 0.014 \quad 0.909)$$

- 9/10 (soit 90% ou 91%) des projets de commutation sont achevés (statut ④).
- 3% des projets de commutation reste au stade d'initiation (statut ①).
- 5% des projets de commutation reste au stade d'étude et consultation (statut ②).
- 2% des projets de commutation reste au stade de négociation avec la cédante (statut ③).

Les graphes en FIGURE 2.5 illustrent la répartition des projets de commutation à l'état stationnaire (c'est-à-dire à un horizon de temps très lointain). A gauche, nous avons inclus les projets de commutation annulés ; alors que, la figure à droite exclut les projets annulés.

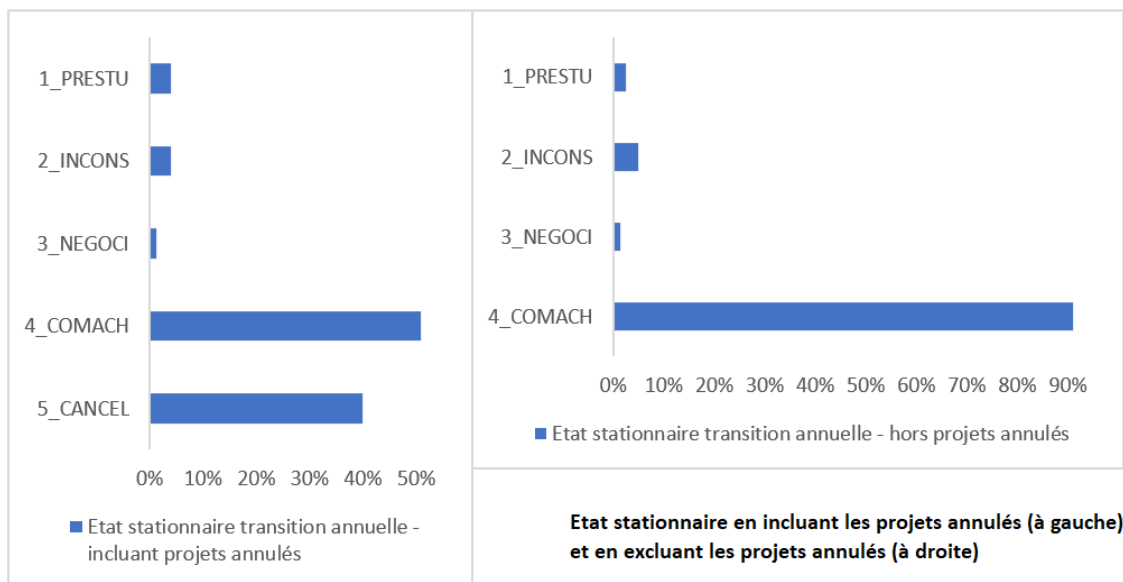



FIGURE 2.5 – État stationnaire

Comme indiqué sur les graphes, la FIGURE 2.5 est obtenue à partir de l'état stationnaire des transitions annuelles. Cependant, quel que soit le pas de temps choisi, nous obtenons sensiblement le même résultat. Pour voir ce qui en est des autres pas de temps, l'ensemble des résultats sous  de l'estimation de l'état stationnaire sont présentés en Annexe A.3.5.

### 2.4.2 Interprétation des résultats de l'estimation de l'état stationnaire

Le but final d'un projet de commutation étant d'aboutir à un accord de commutation, la majorité des projets lancés devraient aboutir à un moment donné. Cependant, nous constatons qu'il n'y a que la moitié des projets qui s'achèvent avec succès, 2/5 projets risquent de ne jamais aboutir. Les 10% restant sont à des stades intermédiaires. La question qui peut se poser, est-il possible d'améliorer le taux de réussite des projets de commutation et, par conséquent, diminuer le taux de projets annulés ?

Par ailleurs, l'analyse en excluant les projets annulés montrent qu'à un certain moment, la plupart (90%) des projets de commutation aboutissent. Certes, seulement 10% des projets de commutation restent à des stades intermédiaires. Mais est-il possible d'améliorer encore plus cette répartition et faire en sorte que très peu de projets restent dans des états intermédiaires (tendre vers un taux de commutations achevées proche de 95% par exemple) ?

Pour répondre à ces questions, la prise en compte des caractéristiques des projets, contrats et cédantes peut révéler des conclusions intéressantes. Cette étude approfondie tenant compte de ces caractéristiques fera l'objet du chapitre 3.

## 2.5 Conclusion

Suite aux lectures et interprétations faites de la matrice de transition ainsi que de l'état stationnaire, il est intéressant de se demander dans quelle mesure peut-on améliorer le passage des projets de commutation vers le statut commuté plutôt que vers le statut annulé ? Cela permettra d'augmenter la proportion de projets effectivement commutés et réduire la proportion des projets annulés et de ceux qui restent longtemps en suspens.

Par ailleurs, étant donné que ces conclusions sont très globales, il convient de segmenter l'analyse et de voir s'il y a des caractéristiques de projets qui influent sur les transitions d'un statut vers un autre.

Le chapitre 3 permet d'approfondir l'étude et d'effectuer une segmentation des données pour tirer des conclusions sur des catégories de projets plus restreintes.

Enfin, concernant le pas de temps à utiliser, vu que les matrices de transition montrent que c'est le pas annuel qui affichent les plus grandes probabilités de transition entre statuts différents (la probabilité de rester dans le même statut est très proche de 1 quand le laps de temps est de l'ordre de quelques mois car il ne se passe pas grand chose dans le projet de commutation). Par conséquent, nous retiendrons pour les analyses réalisées dans la suite de ce document le pas de temps annuel.



## Chapitre 3

# Modélisation logistique et segmentation des commutations

La régression logistique est devenue universelle, puisqu'elle peut traiter des variables à prédire à 2 valeurs (sans faire d'hypothèses aussi restrictives que l'analyse discriminante), à plus de 2 valeurs *ordonnées* mais aussi à plus de 2 valeurs *nominales*. Un des avantages de cette régression est que les variables explicatives peuvent être quantitatives ou qualitatives. Les résultats sont très explicites, surtout dans la version *logit* avec des odds-ratios faciles à interpréter si les modalités de référence sont choisies de façon pertinente.


La régression logistique est aussi une des méthodes de classement les plus fiables, et plusieurs indicateurs statistiques permettent de contrôler facilement cette fiabilité. L'analyse discriminante linéaire peut encore opposer une résistance dans le cas de variables explicatives continues multinomiales, et homoscédastiques, puisqu'elle fournit alors une solution directe dont la régression logistique ne fournit qu'une approximation. Mais la généralité, l'interprétabilité et la robustesse de la régression logistique sont trois atouts majeurs de cette technique fondamentale. [7]

Dans ce chapitre, la modélisation logistique sera utilisée afin d'expliquer la variable d'intérêt la plus pertinente possible pour le modèle par les caractéristiques des projets de commutation, des cédantes et des contrats. La variable d'intérêt pourra être le statut actuel ou la transition d'un statut vers un autre selon l'approche choisie.

Deux approches seront ainsi abordées. La première est une approche statique signifiant qu'il ne sera pas considéré de dynamique particulière dans le modèle. L'observation, à un instant  $t$ , des valeurs des variables explicatives permettra de prédire la valeur prise par la variable d'intérêt. La deuxième approche est dite dynamique car elle prendra en compte l'évolution du statut d'un projet de commutation d'une période à une autre.

La première section de ce chapitre détaillera les aspects théoriques utilisés dans la suite de ce chapitre. Après les rappels de base sur le modèle logistique, les différents tests statistiques seront exposés ainsi que les éléments d'interprétation des résultats de ces tests.

La deuxième section dressera un récapitulatif des variables d'intérêt et des variables explicatives qui seront prises en compte dans les différents modèles et précisera notamment le choix des modalités de référence de chacune des variables.

La mise en oeuvre sous  ainsi que certains éléments pratiques d'interprétation des résultats seront présentés dans la troisième section.

La quatrième section traitera de la modélisation statique. La variable d'intérêt choisie sera le dernier statut connu de chaque projet de commutation. Tandis que la cinquième section sera consacrée au modèle logistique dynamique. La variable d'intérêt, permettant une interprétation plus pertinente, sera l'évolution d'un statut vers un autre. Nous verrons que dans les deux cas, statique ou dynamique, des regroupements de statuts permettent de rendre le modèle plus robuste car le nombre d'observations par statut est parfois trop faible.

Nous terminons ce chapitre par une section qui fait la synthèse des critères de segmentation au vu des résultats des régressions précédentes. Une nouvelle estimation des matrices de transition du modèle de Markov sera proposée avec segmentation. Les états stationnaires associés à ces matrices segmentées seront aussi évalués.

### 3.1 Le modèle de régression logistique : Aspects théorique

Cette section est basée sur les ouvrages de Tufféry S. *Data Mining et statistique décisionnelle, L'intelligence des données*, Éditions Technip, 2010 [7], Hosmer D.W. and Lemeshow S. *Applied Logistic Regression*, Wiley Series in Probability and Mathematical Statistics, 2000 [3] et de Bardos M. *Analyse Discriminante - Application au risque et scoring financier*, Dunod, 2001 [4].

#### 3.1.1 Présentation du modèle

La régression logistique est une méthode statistique qui permet de modéliser la relation entre une variable dépendante et plusieurs variables explicatives. Elle est utilisée lorsque la variable dépendante ou variable endogène est qualitative (nominale ou ordinale). Les variables explicatives ou variables exogènes peuvent être qualitatives ou quantitatives.

Rappelons que dans le cas d'une variable endogène non quantitative, le modèle de régression linéaire n'est plus adapté. Notons  $Y$  la variable dépendante pouvant prendre  $K$  modalités  $y_k$ ,  $k \in \llbracket 1 ; K \rrbracket$ . Soient  $(x_i)_{1 \leq i \leq p}$  les  $p$  variables explicatives avec  $x'_i = (x_{i,1}, x_{i,2}, \dots, x_{i,K})$  le vecteur des variables explicatives pour chaque modalité.

On dispose de  $n$  observations.

Le problème de régression est de modéliser l'espérance de  $Y$  conditionnellement à  $X$  au moyen d'une équation linéaire.

Ainsi, comme on a :

$$\mathbb{E}(Y|X = x) = \mathbb{P}(Y = y_k|X = x)$$

On peut alors écrire,

$$\mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

avec  $\beta_i, i \in \{0, \dots, p\}$ , les paramètres ou coefficients à estimer des variables explicatives.

Cependant, si les variables explicatives sont continues, le terme de droite n'est pas borné or

$$\mathbb{P}(Y = y_k|X = x) \in [0; 1]$$

Il faut donc le transformer en utilisant une des fonctions de lien suivantes :

- La fonction *Logit* :  $g(\pi) = \ln\left(\frac{\pi}{1 - \pi}\right)$
- La fonction *log-log* ou *Gumbel* :  $g(\pi) = \log(-\log(1 - \pi))$
- La fonction *Probit* :  $g(\pi) = \phi^{-1}(\pi)$  avec  $\phi(r) = \mathbb{P}(\mathcal{N}(0; 1) \leq r)$

Ainsi, en notant  $\pi_k(x) = \mathbb{P}(Y = y_k|X = x)$ , on effectue la transformation suivante :

$$g(\pi_k(x)) = \beta_{0,k} + \beta_{1,k} x_{1,k} + \dots + \beta_{p,k} x_{p,k}$$

Si la variable est polytomique, le modèle est multinomial. Nous étudierons ici le modèle multinomial car certaines variables présentent plus de deux modalités. On parle dans ce cas de régression logistique généralisée.

On cherche à modéliser une probabilité dont la courbe représentative prend la forme d'un *S*.

La fonction de répartition d'une loi logistique de moyenne  $\mu$  et de variance  $\sigma^2$  est de la forme :

$$F(x) = \frac{\exp\left(\frac{\pi}{3} \frac{x - \mu}{\sigma}\right)}{1 + \exp\left(\frac{\pi}{3} \frac{x - \mu}{\sigma}\right)}$$

La fonction de répartition d'une loi logistique est de la forme illustrée par la FIGURE 3.1.

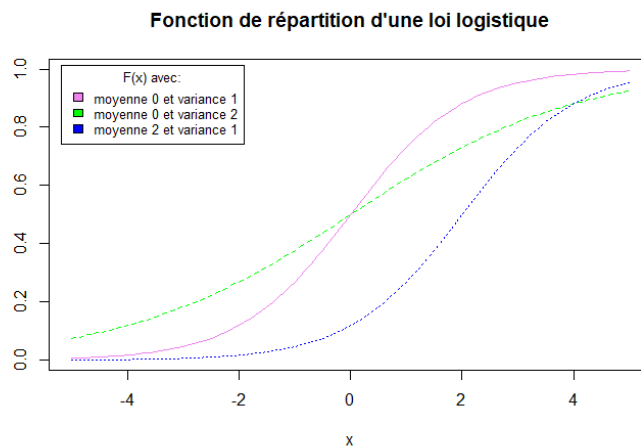


FIGURE 3.1 – Fonction de répartition d'une loi logistique

Le choix de la fonction de lien ne modifie pas les conclusions finales, on choisit donc ici d'utiliser la fonction de lien *Logit*, pour la simplicité de sa mise en œuvre mais aussi pour faciliter l'interprétation du modèle.

On obtient donc :

$$\text{Logit} [\pi_k(x)] = \ln \left( \frac{\pi_k(x)}{1 - \pi_k(x)} \right) = \beta_{0,k} + \beta_{1,k}x_{1,k} + \cdots + \beta_{p,k}x_{p,k}$$

On obtient alors la probabilité  $\pi_k(x)$  en inversant la fonction de lien  $g$ , ce qui donne :

$$\pi_k(x) = \frac{\exp(\beta_{0,k} + \beta_{1,k}x_{1,k} + \cdots + \beta_{p,k}x_{p,k})}{1 + \exp(\beta_{0,k} + \beta_{1,k}x_{1,k} + \cdots + \beta_{p,k}x_{p,k})}$$

Afin de déterminer la combinaison linéaire représentative du *Logit*, on doit d'abord déterminer les variables explicatives.

Il existe trois méthodes afin de sélectionner les variables explicatives :

- La méthode **pas à pas ascendante** : aucune variable n'est sélectionnée au départ puis on ajoute une à une celles qui contribuent le plus au modèle ;
- La méthode **pas à pas descendante** : toutes les variables sont sélectionnées dans un premier temps ensuite on rejette une à une celles qui contribuent le moins au modèle ;
- La méthode **pas à pas mixte** : elle est similaire à la méthode ascendante mais on peut retrancher une variable à chaque étape si son pouvoir discriminant est contenu dans une combinaison de nouvelles variables.

Parmi ces trois méthodes, il est recommandé d'utiliser la méthode descendante. En effet, cette méthode permet de tester à chaque étape toutes les combinaisons possibles de variables explicatives et ainsi obtenir un meilleur modèle.

### 3.1.2 Les estimateurs dans les cas binaire et multinomial

Notre modèle s'écrit :

$$\pi_k(x) = \frac{\exp(\beta_{0,k} + \beta_{1,k}x_{1,k} + \cdots + \beta_{p,k}x_{p,k})}{1 + \exp(\beta_{0,k} + \beta_{1,k}x_{1,k} + \cdots + \beta_{p,k}x_{p,k})}$$

Afin d'estimer les paramètres  $\beta_{i,k}$ ,  $i \in \llbracket 0; p \rrbracket$ ,  $k \in \llbracket 1; K \rrbracket$ , on utilise la méthode du maximum de vraisemblance.

La vraisemblance est la probabilité d'obtenir les données observées exprimées en fonction des coefficients  $\beta_{i,k}$ ,  $k \in \llbracket 1; K \rrbracket$ . On note  $\mathcal{L}(\beta_{0,k}, \beta_{1,k}, \dots, \beta_{p,k})$  la vraisemblance associée à la modalité  $k$  de la variable dépendante.

On cherche les coefficients pour chaque modalité en maximisant la vraisemblance et ajustant le mieux possible les données observées. Cette maximisation ne pouvant se faire de façon explicite, on a recours à une méthode de maximisation numérique.

Notons que si le nombre de modalités de la variable à expliquer est supérieur à deux, le modèle binaire devra être adapté, on parle alors de régression logistique polytomique. Cette dernière peut être ordinaire si les modalités peuvent suivre un ordre donné (par exemple niveau de satisfaction dans une enquête sur des consommateurs d'un produit donné). Dans le cas contraire, elle est dite multinomiale. Et c'est ce dernier modèle qui nous intéresse.

Avant de présenter le modèle multinomial, nous présenterons d'abord le modèle ordinal.

### Régression logistique ordinaire

On suppose dans ce cas, que la variable à prédire  $Y$  prend  $m$  valeurs ordonnées, que l'on note  $1, 2, \dots, m$ .

Le modèle le plus simple et le plus utilisé est le *modèle à pentes égales (ou parallèles)*. On suppose que le *logit* des probabilités cumulatives s'écrit :

$$\text{logit}[\mathbb{P}(Y \leq r | X = x)] = \alpha_r + \sum_i \beta_i x_i, \text{ pour } 1 \leq r \leq m$$

Seule la constante dépend de  $r$ .

Les probabilités s'écrivent :

$$\mathbb{P}(Y \leq r | X = x) = \frac{\exp\left[\alpha_r + \sum_k \beta_k x_k\right]}{1 + \exp\left[\alpha_r + \sum_k \beta_k x_k\right]}, \text{ pour } 1 \leq r \leq m$$

On parle de *proportional odds model* car les odds-ratios pour un  $r$  fixé sont tous proportionnels entre eux, le rapport ne dépendant pas de  $r$  :

$$\frac{\mathbb{P}(Y \leq r | X = x) / \mathbb{P}(Y > r | X = x)}{\mathbb{P}(Y \leq r | X = x') / \mathbb{P}(Y > r | X = x')} = \frac{\exp\left[\alpha_r + \sum_i \beta_i x_i\right]}{\exp\left[\alpha_r + \sum_i \beta_i x'_i\right]} = \exp\left[\sum_i \beta_i (x_i - x'_i)\right]$$

Ce modèle est souvent préféré à des modèles qui deviennent vite trop complexes avec l'accroissement du nombre de variables explicatives ou du nombre de modalités de la variable à expliquer. Dans les applications courantes, cette dernière peut représenter un niveau de gravité d'une maladie ou un niveau de satisfaction d'un client.

Dans la régression logistique ordinaire, en notant  $\mu = \mathbb{P}(Y \leq r | X = x)$ , on peut utiliser les fonctions de lien suivantes :

- **logit**, qui a l'avantage d'avoir la facilité d'interprétation des coefficients
- **probit**, dont les coefficients sont moins facilement interprétables que ceux du logit
- **log-log**, donnée par  $\log(-\log(1 - \mu))$  et utilisée quand les valeurs élevées de la cible sont plus probables (exemple : valeurs 3 à 5 sur 5 dans une enquête de satisfaction)

- **cauchit**, donnée par  $\tan[\pi \times (\mu - 0.5)]$  et utilisée quand les valeurs extrêmes de la cible sont plus probables (exemple : valeur 5 sur 5 dans une enquête de satisfaction).

Comme les valeurs des modalités de la variable à expliquer dans notre étude ne sont pas ordonnées, le modèle ordinal n'est pas vraiment adapté, à moins qu'on restreigne l'étude aux quatre premiers statuts de commutation qui représentent une certaine hiérarchie. Le modèle le plus adapté en tenant compte de tous les statuts de commutation possibles, est plutôt le modèle multinomial nominal.

### Régression logistique multinomiale

On suppose ici que la variable à prédire  $Y$  prend  $m$  valeurs non ordonnées. Dans ce cas, le logit de  $\mathbb{P}(Y = 1|X = x)$  du cas binaire peut être remplacé par le *logit généralisé*  $\log\left(\frac{\mathbb{P}(Y = j|X = x)}{\mathbb{P}(Y = m|X = x)}\right)$  calculé par rapport à une modalité de référence (ici  $Y = m$ ), et c'est cette expression, définie pour  $1 \leq j \leq m$ , qui est écrite comme combinaison linéaire des variables explicatives, dont les coefficients  $(\beta_{ij})_i$  varient pour chaque  $j$ .

Les probabilités s'écrivent :

$$\mathbb{P}(Y = j|X = x) = \frac{\exp\left[\alpha_j + \sum_k \beta_{jk}x_k\right]}{1 + \sum_{i=1}^{m-1} \exp\left[\alpha_i + \sum_k \beta_{ik}x_k\right]}, \quad j = 1, \dots, m-1$$

$$\mathbb{P}(Y = m|X = x) = \frac{1}{1 + \sum_{i=1}^{m-1} \exp\left[\alpha_i + \sum_k \beta_{ik}x_k\right]}$$

Ici, il n'est pas possible de supposer les coefficients  $\beta_{jk}$  indépendants de  $j$ .

### 3.1.3 Tests statistiques

Une fois les paramètres estimés, on effectue des tests afin de vérifier, par exemple, l'adéquation du modèle ainsi que la significativité des coefficients.

Un test statistique comporte deux hypothèses : une hypothèse dite nulle et une hypothèse dite alternative. L'objectif est de trancher pour l'une ou l'autre des deux hypothèses au vu des observations dont on dispose. Plus précisément, un test statistique porte sur le rejet ou non de l'hypothèse nulle. A cet effet, une statistique de test est alors calculée et comparée à une valeur critique théorique.

Les tests statistiques se lisent de la manière suivante : Si la probabilité que la valeur calculée de la statistique du test dépasse la valeur critique est inférieure au niveau de risque maximum toléré choisi  $\alpha$  (en général on prend  $\alpha = 5\%$  mais on peut aller jusqu'à 1% voire moins), on rejette l'hypothèse nulle du test. Cette probabilité est appelée *p-value*.

## Validation du modèle

Les tests suivants permettent de vérifier l'adéquation du modèle.

### *La déviance*

Les hypothèses de test sont les suivantes :

$$\begin{cases} H_0 : \text{ le modèle décrit bien le lien entre les variables} \\ H_1 : \text{ le modèle ne décrit pas bien le lien entre les variables} \end{cases}$$

L'idée ici est de comparer le modèle proposé avec un modèle qui ne spécifie pas de lien entre les variables.

*Sous le modèle proposé :*

$$Y = y_k | x_{1,k}, \dots, x_{p,k} \sim \mathcal{B}(\pi_k(x))$$

$$\text{avec } \pi_k(x) = \frac{\exp(\beta_{0,k} + \beta_{1,k}x_{1,k} + \dots + \beta_{p,k}x_{p,k})}{1 + \exp(\beta_{0,k} + \beta_{1,k}x_{1,k} + \dots + \beta_{p,k}x_{p,k})}$$

Sous ce modèle, on obtient donc :

$$\hat{\pi}_k(x) = \frac{\exp(\hat{\beta}_{0,k} + \hat{\beta}_{1,k}x_{1,k} + \dots + \hat{\beta}_{p,k}x_{p,k})}{1 + \exp(\hat{\beta}_{0,k} + \hat{\beta}_{1,k}x_{1,k} + \dots + \hat{\beta}_{p,k}x_{p,k})}$$

*Sous le modèle général :*

$$Y = y_k | x_{1,k}, \dots, x_{p,k} \sim \mathcal{B}(\pi_k(x))$$

Mais sans aucune spécification particulière sur la probabilité. Un tel modèle est appelé modèle saturé, on a alors :

$$\hat{\pi}_k(x) = y$$

Notons  $\mathcal{L}(\beta_k)$  la vraisemblance du modèle réduit à  $k$  variables et  $\mathcal{L}(\beta_n)$  la vraisemblance du modèle saturé *i.e.* la vraisemblance maximale.

La déviance est la statistique du rapport de vraisemblance entre le modèle proposé et le modèle saturé.

$$D = -2 [\log(\mathcal{L}(\beta_k)) - \log(\mathcal{L}(\beta_n))]$$

Ainsi, sous  $H_0$ , on a  $D \sim \chi_{n-(p+1)}^2$ .

Le but de la régression logistique étant de maximiser la vraisemblance, ceci équivaut à minimiser la déviance.

Cependant, la déviance dépend de la façon dont les données sont agencées. Il est donc difficile de se prononcer sur la validité de la déviance dans le cas de tests d'ajustement.

On peut alors utiliser l'outil suivant :

***Le test du Khi-deux de Pearson***

On suppose que  $Y_i \sim \mathcal{B}(m_i, \pi_i)$  où  $m_i$  vaut 1 si les données ne sont pas groupées,  $m_i$  représente le nombre d'observations dans le groupe  $i$ .

La statistique du *khi<sup>2</sup> de Pearson* s'écrit :

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - m_i \hat{\pi}_i)^2}{m_i \hat{\pi}_i (1 - \hat{\pi}_i)}$$

Sous  $H_0$ , on a :

$$\chi^2 \sim \chi_{n-(p+1)}^2$$

Pour un modèle correct, la Déviance ainsi que le *Khi-deux de Pearson* divisés par leur degré de liberté devraient être approximativement égaux à 1.

***Le test d'Hosmer-Lermershow***

Ce test est recommandé lorsque les variables explicatives sont continues. En effet, dans ce cas, les effectifs estimés étant trop faibles pour chaque sous-population, les conditions d'utilisation du  $\chi^2$  sont rarement réalisées. C'est un test qui consiste à découper les observations en  $g$  groupes ordonnés selon des probabilités croissantes. Le  $\chi^2$  du tableau  $g \times k$  des fréquences pour l'événement modélisé est calculé et comparé à la loi du  $\chi^2$  à  $(g - k)$  degrés de liberté.

Si le  $\chi^2$  est trop grand, le modèle ne s'ajuste pas bien aux données car les fréquences observées et attendues sont significativement différentes.

***La statistique du  $R^2$  ou coefficient de détermination***

Il s'agit d'un indicateur de la qualité d'ajustement du modèle. Il mesure le rapport entre la variance expliquée par le modèle et la variance totale. Il donne le pourcentage de la variance totale de la variable dépendante expliquée par le modèle.

On définit le coefficient de détermination ainsi :

$$R^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(Y - \bar{Y})'(Y - \bar{Y})}$$

$$\text{avec } \hat{\varepsilon} = Y - \hat{Y} \text{ résidus estimés et } \bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$$

Plus le  $R^2$  est proche de 1, meilleur est l'ajustement du modèle. Cependant, il ne diminuera pas lorsque l'on ajoute d'autres variables indépendantes. On introduit alors le  $R^2$  ajusté :

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k}$$

On a alors  $\bar{R}^2 \leq R^2$ .



## Comparaison de deux modèles

### *La différence des déviances*

Elle permet de tester s'il convient d'ajouter  $q$  variables explicatives à un modèle qui en contient déjà  $k$ . Il est facile de montrer que :

$$D(\beta_k) - D(\beta_{k+q}) = -2 [\log(\mathcal{L}(\beta_k)) - \log(\mathcal{L}(\beta_{k+q}))]$$

Sous l'hypothèse  $H_0 : \beta_{k+1} = \dots = \beta_{k+q} = 0$ ,

$$D(\beta_k) - D(\beta_{k+q}) \sim \chi^2(q)$$

Sous le seuil critique de la valeur du  $\chi^2$ , on rejette les  $q$  nouvelles variables.

Cette méthode est souvent utilisée en régression pas à pas.

Afin de comparer deux modèles, on dispose des critères AIC et BIC décrits dans les paragraphes suivants.

### *Le critère d'information d'Akaike (AIC)*

L'*AIC* (de l'anglais *Akaike Information Criterion*) est une mesure de la qualité d'un modèle statistique. Il a été proposé par le statisticien japonais *Hirotsugu Akaike* en 1973.

$$AIC = -2 \log(\mathcal{L}(\beta_k)) + 2(k + 1)$$

$k$  = nombre de degrés de liberté = nombre de paramètres à estimer.

### *Le critère d'information bayésien de Schwarz (BIC)*

Le *BIC* (de l'anglais *Bayesian Information Criterion*) est une autre mesure proposée par *Gideon Schwarz* en 1978. Il est dérivé du critère *AIC*. En plus du nombre de paramètres estimés ( $k$ ), ce critère tient compte de la taille de l'échantillon ( $n$ ).

$$BIC = -2 \log(\mathcal{L}(\beta_k)) + (k + 1) \log(n)$$

Les deux critères *AIC* et *BIC* doivent être au niveau le plus bas possible pour que le modèle soit de meilleure qualité.

## Tests de significativité des coefficients

### *Le test de la déviance*

Il permet de tester l'hypothèse nulle suivante :

$$H_0 : \text{tous les coefficients sont nuls}$$

En effet, il suffit de comparer les modèles ne comportant aucune variable significative et le modèle avec les  $k$  variables significatives. Sous  $H_0$ ,

$$D(\beta_0) - D(\beta_k) \sim \chi^2(k)$$

Si la différence des déviances dépasse le seuil critique du  $\chi^2$ , alors on rejette  $H_0$ .

### ***Le test de Wald***

Ce test permet de vérifier aussi bien la significativité globale des coefficients que la significativité d'un seul coefficient.

#### *Significativité d'un seul coefficient*

L'hypothèse à tester dans ce cas est :

$$H_0 : \beta_i = 0$$

La statistique s'écrit alors :

$$W = \left( \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \right)^2$$

Sous  $H_0$ ,  $W \sim \chi^2(1)$

Ce test compare le modèle sans la variable  $x_i$  avec le modèle comprenant toutes les variables. La règle de décision est la suivante :

Si  $W < \chi^2_{1-\alpha}(1)$ , la contrainte n'est pas rejetée au seuil  $\alpha$ . Si la *p-value* est inférieure au seuil  $\alpha$ , on rejette  $H_0$  au seuil  $\alpha$  et la contrainte est alors rejetée.

#### *Significativité globale*

L'hypothèse nulle testée ici est la nullité de tous les coefficients.

Sous  $H_0$ ,

$$W_j = \hat{\beta}'_j \hat{\Sigma}_j^{-1} \hat{\beta}_j \sim \chi^2(K-1)$$

Avec  $\hat{\beta}'_j = (\hat{\beta}_{j,1}, \dots, \hat{\beta}_{j,K-1})$  et  $\hat{\Sigma}_j$  est la matrice de dimensions  $(K-1) \times (K-1)$  correspondant à la partie de la matrice de variance-covariance relative aux coefficients de la variable  $x_j$ .

Cette méthode est également utilisée dans la régression pas à pas.

Si les tests sont significatifs pour une variable ou un groupe de variables explicatives, le coefficient associé à chaque variable est non nul. Les variables peuvent alors être prises en compte dans le modèle.

Pour des variables qualitatives à plus de deux modalités, la significativité du résultat du test dépend du choix de la modalité de référence. Autrement dit, le choix de la modalité de référence influe sur la significativité des coefficients.

Une fois le modèle final atteint, il faut ensuite tester les interactions entre les variables.

Il est possible que les variables soient significatives mais que le modèle n'explique pas bien la variable dépendante. Ceci s'explique par le fait que des facteurs n'ont pas été recherchés ou inclus dans l'analyse.


### 3.1.4 Odds-ratio

L'*odds-ratio* ( $OR$ ), également appelé *rapport des cotes*, d'un régresseur  $x_i$  mesure l'évolution du rapport des probabilités d'apparition de l'événement  $Y = 1$  contre  $Y = 0$  lorsque  $x_i$  passe de  $x$  à  $x + 1$  dans le cas où  $K = 2$ . L'*odds-ratio* est défini par :

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

L'*odds-ratio* est à différencier du *Risque Relatif* ( $RR$ ). En effet, le  $RR$  est défini par :

$$RR = \frac{\pi(x+1)}{\pi(x)}$$

Les odds-ratios nous seront d'un grand intérêt pour l'analyse des résultats des différentes régressions logistiques réalisées dans ce chapitre. Nous introduirons des éléments pratiques d'interprétation sous  dans la section 3.3 et au fur et mesure de la présentation des modèles dans les sections suivantes.

## 3.2 Variable d'intérêt et variables explicatives du modèle

La variable d'intérêt ou variable expliquée sera une variable polytomique basée sur les statuts de projets de commutation. Le caractère polytomique de la variable d'intérêt signifie, par opposition à une variable binaire, qu'elle peut prendre plus de deux modalités. Selon les modèles qui seront présentés dans ce travail, la variable d'intérêt pourra être prise en compte avec modalités à l'état brut ou avec retraitement ou regroupement de modalités. Pour chaque modèle nous préciserons notre choix et sa justification.

Comme la situation la moins avantageuse, pour un projet de commutation donné, est qu'il soit annulé, la modalité de référence qui sera considérée sera le statut « annulé » ou le passage vers le statut « annulé », et ce, afin de permettre une comparabilité entre les modèles qui seront estimés. Ce choix n'est pas aberrant quant à l'effectif représentant cette modalité (près de 30% des observations, voir B.1). Cependant, nous verrons que certains modèles excluent trop d'observations avec modalité « annulé », auquel cas nous adapterons nos modèles et leur interprétation en choisissant une autre modalité de référence.

Quant aux variables explicatives, elles sont communes à l'ensemble des modèles. Cependant, le modèle final retenu peut éliminer certaines variables qui ne sont pas significatives, qui contiennent trop de valeurs manquantes ou qui réduisent la robustesse du modèle.

Le choix de la modalité de référence pour les variables explicatives se fait de façon à ce qu'un certain ordre ou chronologie apparaisse lorsqu'il s'agit de variables temporelles ou avec possibilité d'ordre mais tout en respectant le fait d'avoir assez d'observations dans ces modalités. Notons que des regroupements de modalités sont réalisés afin de garantir des catégories assez représentées.

Notons que le niveau de granularité des observations qui seront en entrée des différents modèles de régression logistique sera la combinaison du projet de commutation et de la cédante. Ainsi, les caractéristiques inhérentes aux projets de commutation ainsi que celles relatives aux cédantes seront prises en compte.

Par conséquent, certaines variables explicatives sont directement liées aux projets de commutation comme par exemple la variable *Initiat* qui indique la partie qui est à l'origine du lancement du projet de commutation. D'autres sont une agrégation de données à des niveau plus fins (notamment niveau contrat).

Par exemple, La variable *PnP* (Proportionnel ou Non Proportionnel) prend les modalités « P » et « NP ». Quand un projet a l'attribut « P », cela signifie qu'il contient un plus grand nombre de contrats proportionnels que de contrats non proportionnels et inversement, si l'attribut est « NP ». Notons que les facultatives sont incluses dans les contrats non proportionnels. Nous retenons comme modalité de référence « NP » car elle est plus fréquente.



Des statistiques sur les effectifs des modalités des variables qualitatives (variables d'intérêt ou variables explicatives) sont présentées en ANNEXE B.1.

La TABLE 3.1 récapitule l'ensemble des variables explicatives utilisées dans les différentes modélisations logistiques envisagées dans la suite de ce chapitre.

<i>Variable</i>	<i>Description</i>	<i>Modalités</i> (en gras la <b>référence</b> )
<i>PnP</i>	Non Proportionnel ou Proportionnel	<b>NP</b> , P
<i>UWY</i>	Année de souscription (Underwriting Year)	<b>UWY90</b> , UWY80, UWY70, UWY60
<i>Ptf</i>	Portefeuille	<b>AXA</b> , Fund
<i>Initiat</i>	Partie à l'initiative du projet de commutation	<b>AXA LM</b> , Client, Third Party
<i>Vol</i>	Volatilité (inclut de l'APH ou non)	<b>Non Volatility</b> , Volatility Confirmed, Volatility to be confirmed
<i>Ccy</i>	Monnaie (Currency)	<b>USD</b> , EUR, Oth (autre)
<i>Rating_Mdf</i>	Rating de la cédante	<b>High</b> (AAA, AA, A), Low (BBB, BB, B, CCC, CC, C, D), None (pas de rating)
<i>Country_Mdf</i>	Regroupement de pays par continents ou par importance en réassurance	<b>US</b> (États-Unis), <b>EU</b> (Europe), <b>AS</b> (Asie), <b>OT</b> (Autre)
<i>Size_Mdf</i>	Taille du projet de commutation en termes de réserves OS	<b>Low</b> (moins de 10kEUR), <b>Mid</b> (entre 10kEUR et 1mEUR), <b>High</b> (plus de 1mEUR)
<i>Nb_Contrat</i>	Nombre de contrats contenus dans le projet de commutation	Nombre de contrats par projet
<i>ASB</i>	Nombre de contrats avec risque amiante (Asbestos)	Nombre de contrats par projet
<i>POL</i>	Nombre de contrats avec risque pollution	Nombre de contrats par projet
<i>HH</i>	Nombre de contrats avec risque Health Hazard	Nombre de contrats par projet
<i>abPaid_ASB_m</i>	Montant de payés cumulés sur l'amiante	Variable montant en millions d'EUR en valeur absolue
<i>abPaid_POL_m</i>	Montant de payés cumulés sur la pollution	Variable montant en millions d'EUR en valeur absolue
<i>abPaid_HH_m</i>	Montant de payés cumulés sur le Health Hazard	Variable montant en millions d'EUR en valeur absolue
<i>abOS_ASB_m</i>	Montant de d'OS (réserves dossier/dossier) sur l'amiante	Variable montant en millions d'EUR en valeur absolue
<i>abOS_POL_m</i>	Montant de d'OS (réserves dossier/dossier) sur la pollution	Variable montant en millions d'EUR en valeur absolue
<i>abOS_HH_m</i>	Montant de d'OS (réserves dossier/dossier) sur le Health Hazard	Variable montant en millions d'EUR en valeur absolue
<i>abOS_m</i>	Montant de d'OS total (réserves dossier/dossier)	Variable montant en millions d'EUR en valeur absolue
<i>abIBNR_m</i>	Montant d'IBNR total	Variable montant en millions d'EUR en valeur absolue
<i>abDEPOT_m</i>	Montant des dépôts	Variable montant en millions d'EUR en valeur absolue
<i>abBALANCES_m</i>	Montant des balances	Variable montant en millions d'EUR en valeur absolue
<i>abPRICE_m</i>	Prix de la commutation	Variable montant en millions d'EUR en valeur absolue

TABLE 3.1 – Description des variables utilisées dans le modèle

### 3.3 Mise en oeuvre sous et éléments d'interprétation

Sous , l'estimation d'un modèle logistique peut être réalisée en utilisant la fonction **multinom** du package **nnet**. La fonction **odds.ratio** du package **questionr** quant à elle permet d'afficher les odds-ratios, leur intervalle de confiance à 95% ainsi que la *p-value* du test de *Wald* de significativité. Les odds-ratios sont les exponentiels des coefficients estimés par le modèle multinomial. On peut obtenir sous  les odds-ratios de la fonction **odds.ratio** (sans les intervalles de confiance et les *p-values*) en appliquant l'exponentiel aux coefficients du résultat de la fonction **multinom** (voir ANNEXE B.2.1).

Par ailleurs, la fonction **step** du package **stats** permet de sélectionner le meilleur modèle par une procédure pas à pas descendante (stepwise descending) basée sur la minimisation du critère *AIC* décrit plus haut en section 3.1.3 (rubrique comparaison de deux modèles). La fonction procède aux différentes étapes de sélection avant de renvoyer le modèle final. Par exemple, en ANNEXE B.2.1, l'*AIC* du modèle de départ a été amélioré pour passer de 8156.153 à 7969.442. Le nombre de variables explicatives au départ était de 24. Le modèle final n'en contient que 17.

Nous ne commenterons que le modèle retenu à la fin de la procédure de sélection pas à pas descendante, étant donné que c'est le modèle le plus robuste. L'interprétation se fera au moyen des odds-ratios. Les estimations chiffrées détaillées seront présentées en annexe.

Afin de faciliter la lecture et l'interprétation des résultats, deux fonctions supplémentaires peuvent être utilisées :


- La fonction **ggcoef** du package **GGally** permet de tracer les odds-ratios sur un graphe, et ce, pour chaque modalité de chaque variable explicative. La lecture des odds-ratios est ainsi facilitée, mais il ne faudra pas oublier de prendre en compte les modalités de référence dans l'interprétation. Lorsque les odds-ratios d'une modalité de variable explicative donnée se trouvent tous proche de 1, on pourra conclure à l'absence d'effet. Sinon, la fonction suivante nous aidera à aller plus loin dans l'analyse.
- La fonction **forest\_model** du package **forestmodel** quant à elle affiche les modalités de référence et des intervalles de confiance pour les odds-ratios. Elle affiche aussi les *p-values*. En indiquant la référence, on pourra lire plus clairement ces odds-ratios et l'influence des modalités associées sur la segmentation des projets de commutation. Voir par exemple FIGURE 3.3.

Nous optons pour l'utilisation de la fonction **forest\_model** car elle affiche la référence et permet une interprétation plus complète des odds-ratios.

Comme l'interprétation se fera par le moyen de odds-ratio, lorsque celui-ci concerne une variable explicative quantitative, il représente l'effet de l'augmentation d'une unité de cette variable sur le classement d'un projet donné parmi les modalités de la variable d'intérêt.

Qu'il s'agisse de variable quantitative ou qualitative, lorsque l'odds-ratio est proche de ou égal à 1, on conclura à l'absence d'effet sur la segmentation des projets de commutation.

Lorsqu'il s'agit d'une variable qualitative, la lecture se fait modalité par modalité. Par exemple, en ANNEXE B.2.2, l'odds-ratio de la modalité « UWY70 » prend la valeur 10.0161 pour la référence 3. Cela signifie que, par rapport à un projet avec des contrats souscrits dans les années 1990 (modalité de référence de la variable explicative *UWY*), un projet avec des contrats souscrits dans les années 1970 a 10 fois plus de chance d'être en statut ③ (négociation en cours avec la cédante) que d'être en statut ⑤ (annulé).

Cependant, au lieu de se focaliser sur les sorties  en annexe comme on a fait dans le paragraphe précédent, la FIGURE 3.2 permet une lecture directe sur le diagramme. Comme on peut le constater, la ligne correspondant à l'odds-ratio 10.02 est en deuxième position parmi les lignes de la modalité « UWY70 », cela signifie qu'elle concerne le statut ③ de la variable d'intérêt. En effet, il y a au total 4 lignes pour la modalité « UWY70 » ordonnées de façon décroissante en commençant par 4 (statut commuté). La modalité de référence pour la variable explicative *UWY* est indiquée par la mention **Reference** à droite du diagramme (« UWY90 »). La modalité de référence de la variable d'intérêt est le statut « annulé ».

Dans cette section, nous allons d'abord tenter d'estimer un modèle contenant la totalité des variables évoquées dans la section précédente. Ensuite, nous allons faire une présélection des variables à utiliser dans les modèles de l'étude qui seront présentés dans les sections suivantes.

L'estimation du modèle statique avec l'ensemble des statuts et la totalité des variables explicatives permet d'obtenir les odds-ratios de la FIGURE B.1 que nous préférons mettre en ANNEXE B.2.2. Ce diagramme étant presque illisible à cause de valeurs aberrantes, l'objectif est simplement de montrer qu'il est nécessaire de faire une sélection de variables afin de permettre une meilleure interprétabilité.

En effet, comme on peut le constater, la variable montant d'OS Amiante « abOS\_ASB\_m » affiche une valeur d'odds-ratio aberrante et très volatile. Vu l'échelle du diagramme, la valeur est très élevée et l'intervalle de confiance associé est très large. Ceci empêche une lecture claire du modèle et rend l'interprétation presque impossible car le diagramme de la FIGURE B.1 en ANNEXE B.2.2 devient illisible à cause de cette variable à odds-ratio aberrant.

Le modèle n'étant pas optimal, il convient de sélectionner les variables explicatives significatives et qui optimisent la qualité du modèle. Nous procédons d'abord à une sélection pas à pas descendante pour éliminer toutes les variables qui affectent la robustesse du modèle. Nous éliminons aussi les variables qui affichent des odds-ratios très élevés et volatiles. Ainsi, nous obtenons un modèle réduit mais plus stable avec des diagrammes plus lisibles et plus facile à interpréter (voir section suivante pour l'interprétation).

Le diagramme des odds-ratio du modèle réduit (après sélection des variables explicatives) s'obtient avec la fonction `forest_model`. Ce diagramme est illustré par la FIGURE 3.2 (voir page suivante). Comparé au diagramme de la FIGURE B.1 en ANNEXE B.2.2, le modèle réduit dont les résultats sont illustrés par la FIGURE 3.2 permet une meilleure lisibilité car les valeurs des odds-ratios sont relativement moins élevées et globalement moins volatiles.

Dans la suite du chapitre, nous n'affichons que le modèle réduit retenu après sélection finale des variables. Les odds-ratios des coefficients dont la *p-value* > 1% ne seront pas commentés.

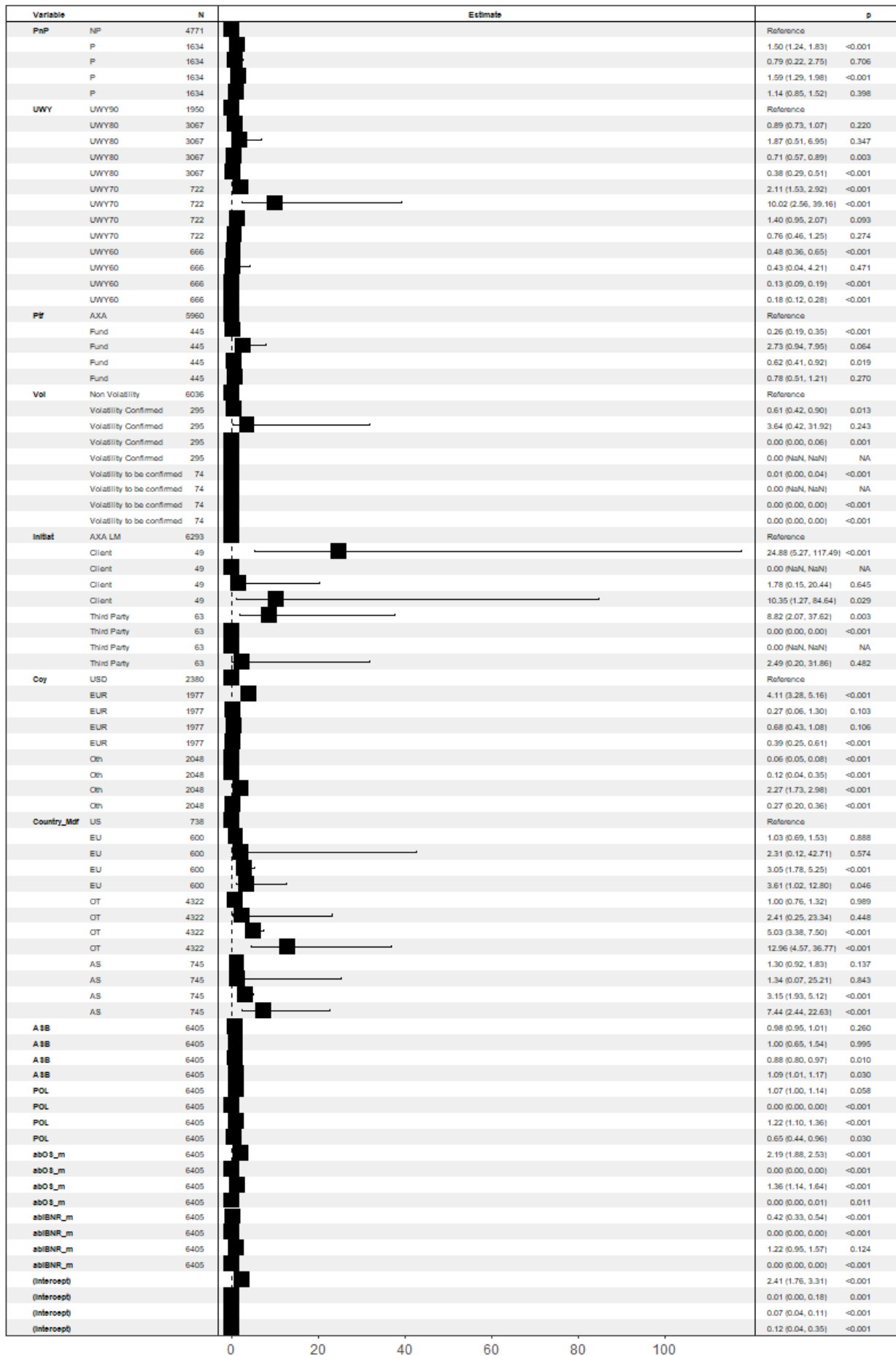


FIGURE 3.2 – Diagramme des odds-ratios du modèle statique après sélection des variables



### 3.4 Modèle multinomial statique

Dans cette section, nous analysons la façon dont sont répartis les projets de commutation, à un instant donné, par statut selon les différentes caractéristiques des projets.

Une première approche sera de faire une régression logistique en considérant comme variable d'intérêt l'ensemble des statuts de ① à ⑤ caractérisant les projets de commutation à un instant donné. Nous choisissons comme date d'observation le 30 juin 2019. La variable d'intérêt est donc le dernier statut connu à fin juin 2019.

Une deuxième approche consiste à regrouper les statuts intermédiaires (①, ②, ③) en un statut unique qu'on appellera « InP » (pour *In Progress*), les statuts restants ④ et ⑤ seront renommés : « Com » (pour *Commutated*) et « Can » (pour *Cancelled*).

Rappelons que dans les deux cas, les individus (lignes) du modèle sont formés ou identifiés par la combinaison des projets de commutation avec les cédantes afin de tenir compte de leurs caractéristiques.

#### 3.4.1 Estimation du modèle statique avec tous les statuts sans regroupement

Dans cette sous-section, nous utilisons comme variable d'intérêt le dernier statut connu à fin juin 2019. Tous les statuts de ① à ⑤ sont pris en compte sans regroupement.

La modalité de référence utilisée pour la variable d'intérêt est le statut ⑤ (annulé). L'interprétation des résultats étant faite à l'aide des odds-ratios, l'influence d'une variable explicative sur le classement d'un projet donné sera mesurée relativement à cette référence ⑤ donc par rapport aux projets annulés. Chaque variable explicative aura aussi sa modalité de référence. Pour interpréter les odds-ratio, nous distinguons les variables quantitatives et les variables qualitatives.

L'estimation de ce modèle a déjà été réalisée précédemment et le diagramme correspondant se trouve sur la FIGURE 3.2.

En comparaison avec les projets avec une majorité de contrats non proportionnels, les projets avec des contrats proportionnels ont 1.5 fois plus de chance d'être commutés « statut ④ » que d'être annulés « statut de référence ⑤ ». La *p-value* associée à cette modalité est  $< 0.001$  et donc le coefficient est très significatif. Dans la suite de ce chapitre, nous ne commenterons que les modalités avec des coefficients significatifs.

Par exemple, la seule modalité significative restante de la variable *PnP* est associée à l'odds-ratio 1.59. L'interprétation est la même que précédemment avec 1.59 fois plus de chance d'être au statut ② que d'être au statut ⑤.

Pour l'année de souscription *UWY*, l'odds-ratio le plus élevé est celui associé à la modalité « *UWY70* ». Ainsi, comme déjà mentionné dans la section précédente, les projets avec souscriptions dans années 1970 ont 10 fois plus de chance d'être au statut ③ que d'être au statut annulé. Pour cette même variable, l'odds-ratio le plus faible et significatif est associé aux années 1960 et au statut ② (réconciliation des chiffres). L'inverse de sa valeur,  $1/0.13 = 7.7$  signale que les projets avec souscriptions dans les années 1960 ont 7.7 fois plus de chance d'être au statut annulé que

d'appartenir au statut (2). Notons ainsi l'inversion des références lorsque l'interprétation concerne l'inverse de l'odds-ratio, très utile notamment lorsque celui-ci est inférieur à 1.

Au seuil de 1%, le seul odds-ratio significatif pour la variable *Ptf* est associé au statut (4) et sa valeur est 0.26. Ainsi, on peut dire que les projets du fonds d'investissement ont environ 4 fois (1/0.26) plus de risques d'être annulés que d'être commutés. Du point de vue du business model d'AXA LM, cela pourrait s'expliquer par le fait que les prix de commutation proposés pour les portefeuilles du fonds sont moins intéressants pour les cédantes.

Pour la volatilité, les odds-ratios sont très proches de 0 et ne permettent pas une analyse cohérente des résultats. Mais la variable *Initiat* présentent deux odds-ratios significatifs qui indiquent que lorsque la commutation est initiée par la cédante (Client) ou une tierce partie (Third Party), le projet a tendance à être en statut commuté plus qu'en statut annulé.

Enfin, comparé à la monnaie USD, les projets en EUR ont 4 fois plus de chance d'être commutés que d'être annulés.

Nous n'allons pas interpréter les odds-ratios un par un mais nous verrons les changements d'interprétation d'un modèle à un autre. Ainsi, nous optons pour une approche comparative entre les résultats des modèles qui seront estimés dans ce chapitre. Les tendances qui seront confirmées par tous les modèles ou la plupart d'entre eux seront retenus notamment pour élaborer une synthèse de la segmentation des projets de commutation.

### 3.4.2 Estimation du modèle statique avec statuts regroupés

Dans cette sous-section, au lieu d'utiliser l'ensemble des valeurs de la variable d'intérêt utilisée dans le modèle précédent, nous considérons un regroupement en 3 modalités : « InP » (In Progress), « Com » (Commutated) et « Can » (Cancelled). Il s'agit toujours du dernier statut connu à fin juin 2019. La modalité de référence est le statut « Can », Cela nous permettra de faire des comparaisons relativement au même statut « annulé » utilisé dans le modèle statique sans regroupement et donc de savoir si c'est cohérent avec les conclusions de la sous-section précédente.

De même que pour l'estimation précédente, nous avons procédé à une sélection de modèle par une méthode pas à pas descendante. Les résultats détaillés de l'estimation du modèle retenu par la fonction **step** est en ANNEXE B.2.2.

La FIGURE 3.3 illustre les odds-ratios de ce modèle. Rappelons que l'ordre des lignes de ce diagramme est très important pour l'interprétation. La référence de la variable d'intérêt étant « Can », il reste deux modalités. Cela explique pourquoi chaque modalité de chaque variable explicative qualitative soit affichée deux fois (sauf la référence indiquée par la mention **Reference** à droite du diagramme). Notons que la première ligne correspond à la modalité « Com » et la deuxième ligne correspond à la modalité « InP ».

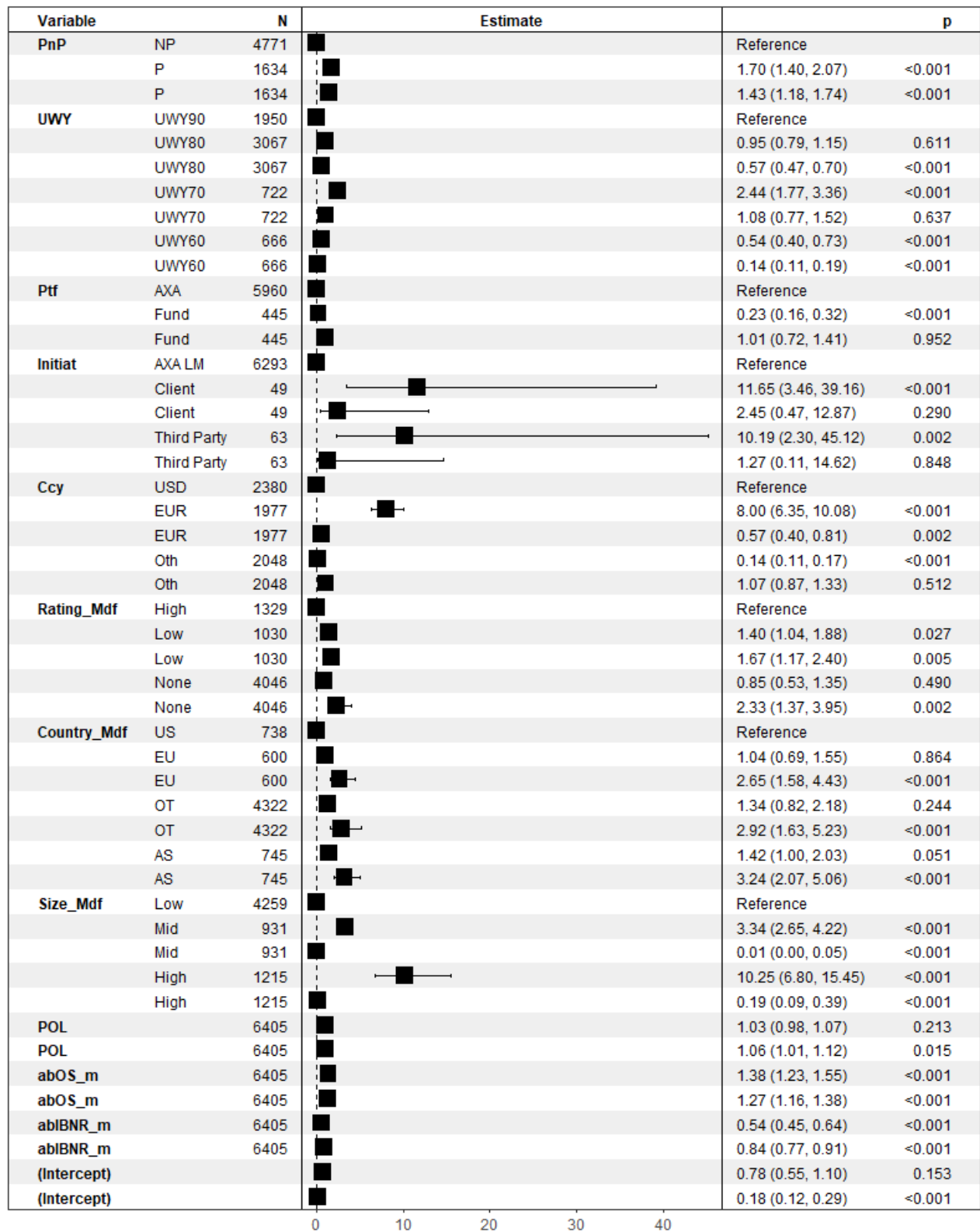


FIGURE 3.3 – Tableau des odds-ratios du modèle statique avec statuts regroupés

L'interprétation est plus aisée avec ce regroupement de statuts. En effet, comme mentionné dans le paragraphe précédent, il n'y a que 2 odds-ratios par modalité de chaque variable explicative au lieu de 4 odds-ratios. Pour la variable *Ptf*, par exemple, nous remarquons que les projets avec des contrats appartenant aux portefeuilles des fonds d'investissement ont  $1/0.23 = 4.34$  fois plus de risque d'être annulés que les portefeuilles AXA.

Les odds-ratios de la variable *Initiat*, indiquent que les projets initiés par le client (la cédante) ou une tierce partie (Third Party), un courtier par exemple, ont respectivement 11.65 et 10.19 fois plus de chance d'être commutés, en comparaison avec les projets initiés par AXA LM.

Ces conclusions sont cohérentes avec celles de la sous-section précédente. C'est le cas aussi pour les autres variables communes interprétées dans les sections précédentes. En effet, la volatilité n'est plus sélectionnée dans ce modèle. Alors qu'une nouvelle variable entre dans l'interprétation de ce modèle. Il s'agit de la variable *Size\_Mdf* pour laquelle les plus hauts odds-ratios indiquent que, plus le volume de la commutation est élevé, plus le projet a de chance d'être commuté que d'être annulé.

L'interprétation des OS et des IBNR montrent aussi une cohérence avec le modèle précédent. En effet, dans les deux modèles, on remarque que plus les OS augmentent plus le projet a de chance d'être en statut commuté que d'être en statut annulé. Cependant, pour les IBNR, c'est le constat inverse des OS.

Enfin, les autres variables explicatives *ASB*, *POL*, *Country\_Mdf*, *Rating\_Mdf* (vue dans le modèle précédent mais absente dans ce modèle) ne présentent pas d'intérêt dans l'analyse vu la non significativité des coefficients ou alors vu que les conclusions concernant ces variables diffèrent entre ce modèle et le modèle précédent. Ou encore parce que l'odds-ratio associé est très proche de 1 donc sans effet sur la variable d'intérêt.

### 3.5 Modèle multinomial dynamique

Dans cette section, nous allons modéliser les transitions vers un statut particulier en  $t = 1$  sachant un état initial donné à l'étape précédente ( $t = 0$ ). Cette modélisation se fera en fonction des variables exogènes observées caractérisant les projets de commutation.

La principale différence entre le modèle qu'on verra dans cette section et le modèle présenté dans la section précédente réside dans l'introduction d'une dynamique au modèle logistique. Il s'agit donc de se rapprocher de l'analyse par chaînes de Markov vue au chapitre 2. L'objectif est d'obtenir un modèle plus complet, et ce, en tentant d'expliquer et de prédire la dynamique d'un projet de commutation par les caractéristiques des variables explicatives.

Comme pour le dernier modèle statique estimé, nous allons regrouper les modalités. Il paraît naturel de ne prendre en compte en  $t = 0$  que les projets de commutation qui sont en cours (*In Progress*) pour savoir comment ils se comportent en  $t = 1$  selon leur caractéristiques. En effet, une fois qu'un projet est commuté ou annulé à une étape donnée, il n'est plus intéressant d'analyser son évolution à l'étape suivante car il sera toujours commuté ou annulé.

Ainsi, en restreignant l'analyse aux projets avec statut en cours « InP » en  $t = 0$ , le statut possible en  $t = 1$  pourra être annulé « Can », commuté « Com » ou toujours en cours « InP ». Cependant, c'est la combinaison de la valeur au départ (à l'étape  $t = 0$ ) et la valeur à la fin (en  $t = 1$ ) qui est plus intéressante pour l'analyse dans ce modèle car cette combinaison rend compte de la dynamique du projet de commutation. Par conséquent, 3 modalités seront traités « A\_A », « A\_B » et « A\_C ». Les lettres A, B et C correspondent respectivement aux statuts regroupés précédents « InP » (In Progress), « Com » (Commutated) et « Can », on utilise une seule lettre pour faciliter la lecture des transitions.

Étant donné que la modalité avec le plus grand effectif est « A\_A », nous la retenons comme modalité de référence. Ainsi, l'interprétation doit être adaptée dans ce modèle, car les comparaisons se feront par rapport au fait qu'il n'y ait pas de changement de statut.

L'objectif de ce modèle est d'apporter une indication sur le comportement des projets de commutation selon leurs caractéristiques, entre  $t = 0$  et  $t = 1$ . Ainsi, le choix du pas du temps est crucial pour ce modèle. Nous avons jugé au chapitre précédent qu'il est plus intéressant pour l'analyse de retenir comme pas de temps la période d'une année. Cependant, le fait de se restreindre uniquement au cas des projets en cours (InP) au départ, réduit considérablement le nombre d'observations par an. Ainsi, nous avons opté pour une démarche consistant à ramener tous les projets de commutation à une date  $t = 0$  identique correspondant à la date du statut initial observé. Cette hypothèse étant forte, nous intégrons dans le modèle une variable explicative supplémentaire *Period0* indiquant la période à laquelle le projet était à son statut de départ. Nous regroupons les valeurs de cette variable en 3 modalités : « Avant06 » (pour les projets initiés avant 2006), « Ent0612 » (pour les projets initiés entre 2006 et 2012) et « Apres12 » (pour les projets initiés après 2012).

Nous présenterons dans la suite de cette section le modèle estimé avec transitions sur une année ainsi que les conclusions qu'on peut tirer de l'analyse des odds-ratios. Comme pour les modèles statiques, notons encore une fois que notre analyse portera uniquement sur le modèle retenu après la sélection pas à pas descendante du meilleur modèle (au vu du critère AIC).

Variable		N	Estimate	p
PnP	NP	3213	Reference	
	P	930	1.40 (1.12, 1.74)	0.003
	P	930	0.88 (0.59, 1.32)	0.542
UWY	UWY90	1362	Reference	
	UWY80	1809	1.28 (1.06, 1.56)	0.012
	UWY80	1809	1.36 (0.97, 1.91)	0.072
	UWY70	399	1.91 (1.42, 2.57)	<0.001
	UWY70	399	1.59 (0.92, 2.76)	0.096
	UWY60	573	1.08 (0.74, 1.57)	0.702
	UWY60	573	1.14 (0.62, 2.09)	0.683
Ptf	AXA	3805	Reference	
	Fund	338	0.81 (0.53, 1.24)	0.329
	Fund	338	0.85 (0.52, 1.36)	0.491
Initiat	AXA LM	4072	Reference	
	Client	30	1.32 (0.52, 3.34)	0.557
	Client	30	1.52 (0.43, 5.39)	0.515
	Third Party	41	2.05 (0.99, 4.26)	0.054
Vol	Third Party	41	0.00 (NaN, NaN)	NA
	Non Volatility	3794	Reference	
	Volatility Confirmed	280	0.79 (0.51, 1.22)	0.286
	Volatility Confirmed	280	0.78 (0.36, 1.67)	0.523
	Volatility to be confirmed	69	0.00 (0.00, 0.00)	<0.001
Ccy	Volatility to be confirmed	69	4.26 (2.22, 8.20)	<0.001
	USD	2029	Reference	
	EUR	824	2.98 (2.30, 3.87)	<0.001
	EUR	824	0.85 (0.54, 1.34)	0.485
	Oth	1290	0.20 (0.14, 0.29)	<0.001
Country_Mdf	Oth	1290	0.01 (0.01, 0.03)	<0.001
	US	408	Reference	
	EU	289	0.74 (0.48, 1.13)	0.164
	EU	289	4.79 (1.64, 13.96)	0.004
	OT	2988	1.33 (1.03, 1.73)	0.030
Size_Mdf	OT	2988	3.21 (1.28, 8.01)	0.013
	AS	458	1.02 (0.72, 1.45)	0.899
	AS	458	1.69 (0.60, 4.76)	0.319
	Low	2393	Reference	
	Mid	615	2.58 (1.98, 3.36)	<0.001
Period0	Mid	615	1.06 (0.71, 1.57)	0.785
	High	1135	0.80 (0.60, 1.08)	0.151
	High	1135	0.00 (NaN, NaN)	NA
	Avant06	962	Reference	
	Ent0612	1408	3.52 (2.62, 4.74)	<0.001
ASB	Ent0612	1408	14.12 (6.32, 31.55)	<0.001
	Apres12	1773	0.85 (0.63, 1.13)	0.252
	Apres12	1773	24.15 (11.43, 51.02)	<0.001
	ASB	4143	0.96 (0.92, 1.01)	0.112
	ASB	4143	1.02 (0.95, 1.10)	0.498
	POL	4143	1.11 (0.98, 1.26)	0.112
	POL	4143	0.96 (0.87, 1.06)	0.420
	HH	4143	0.95 (0.85, 1.06)	0.352
	HH	4143	1.08 (1.01, 1.16)	0.035
	abOS_ASB_m	4143	0.00 (0.00, 0.00)	<0.001
	abOS_ASB_m	4143	1.03 (0.06, 16.69)	0.981
	abIBNR_m	4143	1.39 (1.30, 1.48)	<0.001
	abIBNR_m	4143	0.01 (0.00, 0.32)	0.008
(Intercept)			0.12 (0.08, 0.17)	<0.001
(Intercept)			0.01 (0.00, 0.02)	<0.001

FIGURE 3.4 – Tableau des odds-ratios du modèle dynamique avec la variable Period0

Le modèle estimé après sélection des variables explicatives par méthode pas à pas descendante montre que la variable *Period0* est très significative.

En effet, la FIGURE 3.4 montre que l'odds-ratio associé à la modalité « Ent0612 » est très significatif et sa valeur élevée de 14.12 signifie que les projets initiés entre 2006 et 2012 ont 14 fois plus de risque d'être annulés un an plus tard que de rester en statut en cours, et ce, en comparaison aux projets initiés avant 2006. La valeur de l'odds-ratio de la modalité « Apres2012 » est encore plus élevée (24.15) ce qui signifie que le phénomène s'accroît avec le temps : les projets initiés plus récemment présentent plus de risque d'être annulés que les plus anciens projets (initiés avant 2006).

Ce constat peut être expliqué par le fait que les projets avec des cédantes qui acceptent facilement de commuter ont déjà été réalisés dès le départ. Les anciens portefeuilles encore tenus actuellement ont atteint une certaine maturité en termes de possibilité de commutation.

Les valeurs des odds-ratios de *Period0* étant très élevées, la FIGURE 3.4 ne permet pas de distinguer facilement les effets des autres variables. Ainsi, nous retenons cette variable (*Period0*) comme variable de segmentation mais nous jugeons utile d'estimer un autre modèle qui l'exclut afin de faciliter la lecture (voir La FIGURE 3.5). Nous commenterons dans la suite de cette section les résultats de ce modèle.

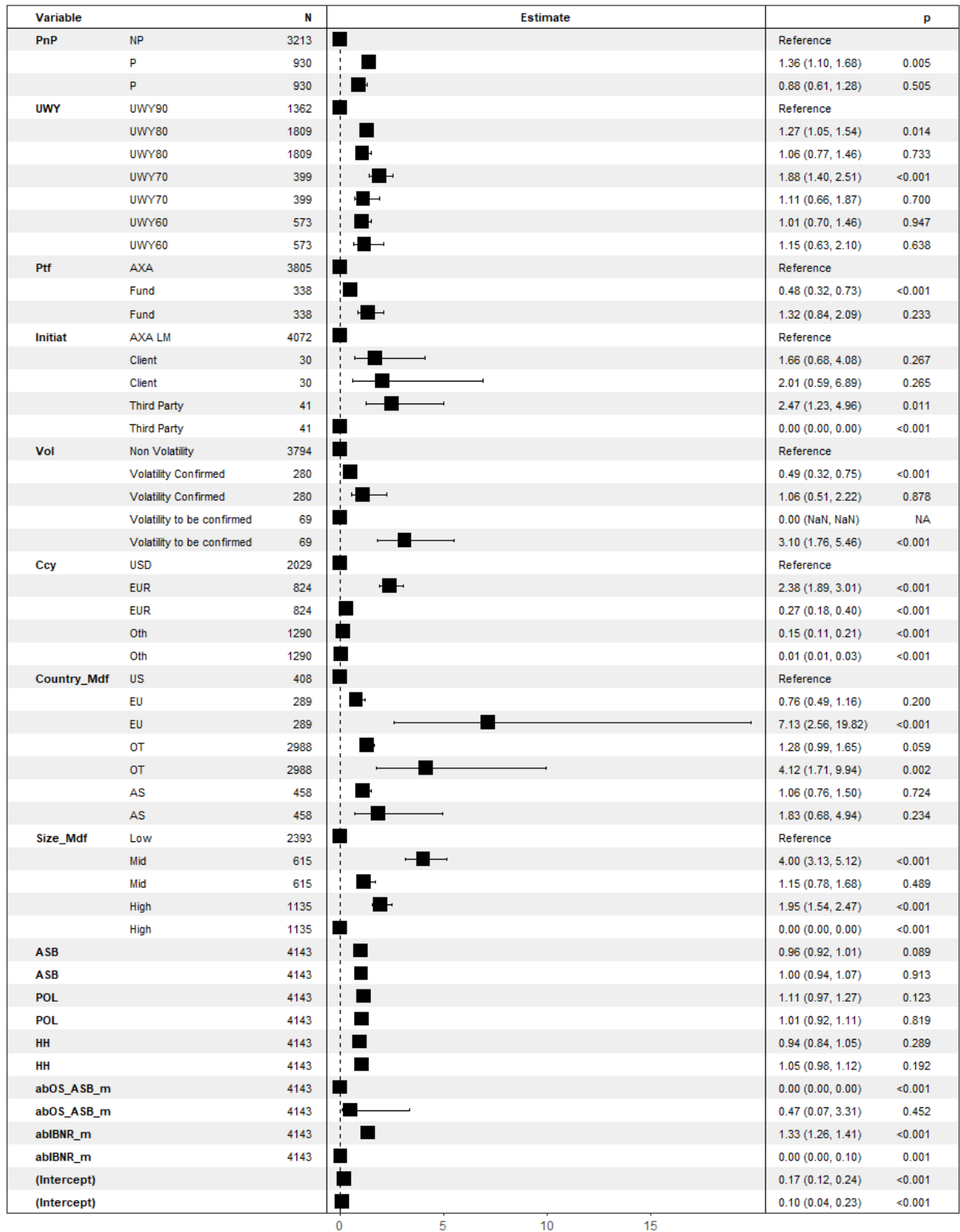


FIGURE 3.5 – Tableau des odds-ratios du modèle dynamique



L'interprétation du modèle de la FIGURE 3.5 se fait par rapport à la modalité de référence « A\_A » car elle est la plus représentative.

Le constat global sur ce modèle dynamique est que les variables, *PnP*, *UWY*, *Ptf*, *Initiat*, *Ccy*, *Size\_Mdf* influencent la transition vers un statut donné depuis le statut « InP » de façon similaire à ce qui a été constaté pour les modèles précédents.

En effet, nous constatons, par exemple, que la modalité « UWY70 » prend la valeur 1.88 avec un coefficient significatif. Cela confirme que les projets avec souscriptions dans les années 1970 ont plus de chance d'être commutés que les autres projets.

## 3.6 Synthèse des critères de segmentation et estimation des matrices de transition avec segmentation

Dans cette section nous dressons un bilan des critères à retenir pour la segmentation. Ensuite, nous présenterons les résultats de l'estimation des matrices de transition du modèle de Markov en appliquant les critères de segmentation retenus.

### 3.6.1 Synthèse des critères de segmentation

Au vu des résultats issus des modèles statique et dynamique présentés dans les sections précédentes, les critères à prendre en compte sont les suivants :

- PnP : les contrats proportionnels ont plus de chance d'être commutés
- UWY : les contrats souscrits dans les années 1970 ont plus de chance d'être commutés
- Ptf : les contrats des portefeuilles AXA ont plus de chance d'être commutés
- Initiat : cette variable n'étant pas très bien représentée en termes d'effectifs de chaque modalité, les résultats sont à prendre avec précaution, mais globalement, les projets initiés par la cédante ou par une tierce partie (courtier) ont plus de chance d'être commutés
- Ccy : les contrats libellés en EUR ont plus de chance d'être commutés
- Period0 : les projets de commutation initiés plus récemment ont plus de risque d'être annulés que les anciens projets

Les variables restantes ne seront pas prises en compte dans la segmentation soit parce que les différents modèles estimés montrent qu'elles n'ont pas d'effet sur la variable d'intérêt considérée (coefficient non significatif ou odds-ratio proche de 1) ou alors parce que les résultats les concernant diffèrent d'un modèle à un autre.

### 3.6.2 Estimation des matrices de transition segmentées du modèle de Markov

Cette sous-section présente les matrices de transition annuelle basées sur la segmentation réalisée précédemment. Nous appliquerons variable par variable les critères de segmentation. Ces résultats sont à comparer avec la matrice de transition annuelle avec prise en compte des 5 statuts vue en sous-section 2.3.2.

#### Matrice segmentée avec des contrats proportionnels

$$P_5^A = \begin{pmatrix} 0.742 & 0.031 & 0.0217 & 0.156 & 0.046 \\ 0.020 & 0.833 & 0.0218 & 0.099 & 0.024 \\ 0.013 & 0.004 & 0.6130 & 0.352 & 0.017 \\ 0.005 & 0.002 & 0.0021 & 0.989 & 0.000 \\ 0.023 & 0.016 & 0.0103 & 0.043 & 0.905 \end{pmatrix}$$

Nous remarquons que les probabilités de transition vers le statut « commuté » (colonne 4) sont nettement meilleures en comparaison avec celles de la matrice évaluée au globale (sans segmentation) en sous-section 2.3.2.

Par exemple, pour la matrice segmentée par le critère proportionnel, les projets proportionnels en statut ① au départ ont une probabilité de 15.6% d'être en statut ④ (commuté) un an plus tard. Alors que la matrice non segmentée présentait une probabilité de 6.5% pour la même transition.

Par ailleurs, les annulations de projets sont beaucoup moins importantes comparées à la matrice non segmentée. Par exemple, les projets proportionnels en statut ① au départ ont une probabilité de 4.6% d'être annulés un an plus tard, cette même probabilité était de 19.4% dans le modèle non segmenté.

Enfin, l'estimation de l'état stationnaire montre une nette amélioration des projets commutés. En effet, 92% de projets sont commutés à l'état stationnaire et seulement 3% sont annulés comme le montre la répartition suivante à l'état stationnaire avec transition annuelle :

$$\text{État stationnaire avec transition annuelle} = (0.023 \quad 0.023 \quad 0.008 \quad 0.919 \quad 0.026)$$

Dans la suite de cette sous-section, nous remarquons que les matrices de transition avec segmentation, quel que soit le critère, améliorent considérablement le passage vers le statut commuté.

### Matrice segmentée avec souscription dans les années 1970

$$P_5^A = \begin{pmatrix} 0.653 & 0.030 & 0.040 & 0.234 & 0.040 \\ 0.011 & 0.887 & 0.028 & 0.062 & 0.011 \\ 0.000 & 0.000 & 0.700 & 0.300 & 0.000 \\ 0.005 & 0.003 & 0.001 & 0.988 & 0.000 \\ 0.017 & 0.014 & 0.005 & 0.052 & 0.909 \end{pmatrix}$$

$$\text{État stationnaire avec transition annuelle} = (0.018 \quad 0.033 \quad 0.012 \quad 0.916 \quad 0.022)$$

### Matrice segmentée avec portefeuilles AXA

$$P_5^A = \begin{pmatrix} 0.696 & 0.029 & 0.033 & 0.181 & 0.058 \\ 0.016 & 0.855 & 0.022 & 0.085 & 0.019 \\ 0.006 & 0.008 & 0.572 & 0.393 & 0.019 \\ 0.006 & 0.003 & 0.002 & 0.987 & 0.000 \\ 0.025 & 0.016 & 0.010 & 0.041 & 0.906 \end{pmatrix}$$

$$\text{État stationnaire avec transition annuelle} = (0.022 \quad 0.028 \quad 0.009 \quad 0.910 \quad 0.030)$$

**Matrice segmentée avec initiative de la cédante**

$$P_5^A = \begin{pmatrix} 0.730 & 0.000 & 0.076 & 0.153 & 0.038 \\ 0.037 & 0.740 & 0.000 & 0.185 & 0.037 \\ 0.047 & 0.000 & 0.619 & 0.333 & 0.000 \\ 0.004 & 0.011 & 0.004 & 0.978 & 0.000 \\ 0.018 & 0.009 & 0.013 & 0.036 & 0.922 \end{pmatrix}$$

$$\text{État stationnaire avec transition annuelle} = (0.026 \quad 0.041 \quad 0.017 \quad 0.882 \quad 0.033)$$

À l'état stationnaire, ce modèle présente la plus faible proportion (88.2%) de projets commutés relativement aux répartitions des états stationnaires des autres critères de segmentation. Ce critère n'est peut être pas le plus prioritaire dans la segmentation, d'autant plus qu'il n'existe pas de marge de manoeuvre pour choisir l'initiative d'un projet de commutation.

**Matrice segmentée avec monnaie EUR**

$$P_5^A = \begin{pmatrix} 0.700 & 0.025 & 0.023 & 0.197 & 0.053 \\ 0.028 & 0.750 & 0.032 & 0.151 & 0.036 \\ 0.008 & 0.004 & 0.540 & 0.430 & 0.016 \\ 0.004 & 0.001 & 0.001 & 0.991 & 0.000 \\ 0.027 & 0.013 & 0.009 & 0.051 & 0.897 \end{pmatrix}$$

$$\text{État stationnaire avec transition annuelle} = (0.017 \quad 0.009 \quad 0.006 \quad 0.950 \quad 0.017)$$

À l'état stationnaire, ce modèle présente la proportion la plus forte (95%) de projets commutés relativement aux répartitions des états stationnaires des autres critères de segmentation. Il est intéressant de mettre ce critère en priorité dans la segmentation.

**Matrice segmentée avec ancienne période d'initiation de projets**

$$P_5^A = \begin{pmatrix} 0.825 & 0.020 & 0.012 & 0.120 & 0.021 \\ 0.007 & 0.907 & 0.016 & 0.060 & 0.007 \\ 0.003 & 0.000 & 0.777 & 0.219 & 0.000 \\ 0.003 & 0.001 & 0.000 & 0.994 & 0.000 \\ 0.034 & 0.015 & 0.007 & 0.062 & 0.880 \end{pmatrix}$$

$$\text{État stationnaire avec transition annuelle} = (0.021 \quad 0.021 \quad 0.005 \quad 0.941 \quad 0.0120)$$

**Matrice segmentée avec combinaison années 1970 et proportionnel**

Comme le montre la matrice ci-dessous, la combinaison de deux critères améliore de façon plus importante la probabilité de transition vers le statut commuté en comparaison avec des segmentations individuelles.

$$P_5^A = \begin{pmatrix} 0.678 & 0.035 & 0.017 & 0.250 & 0.017 \\ 0.022 & 0.750 & 0.045 & 0.181 & 0.000 \\ 0.000 & 0.000 & 0.666 & 0.333 & 0.000 \\ 0.007 & 0.003 & 0.000 & 0.988 & 0.000 \\ 0.011 & 0.005 & 0.000 & 0.059 & 0.922 \end{pmatrix}$$

État stationnaire avec transition annuelle =  $(0.023 \quad 0.015 \quad 0.005 \quad 0.944 \quad 0.013)$

La FIGURE 3.6 est une synthèse des effets de la segmentation sur la probabilité d’être commuté ou annulé pour un projet de commutation à l’état stationnaire.

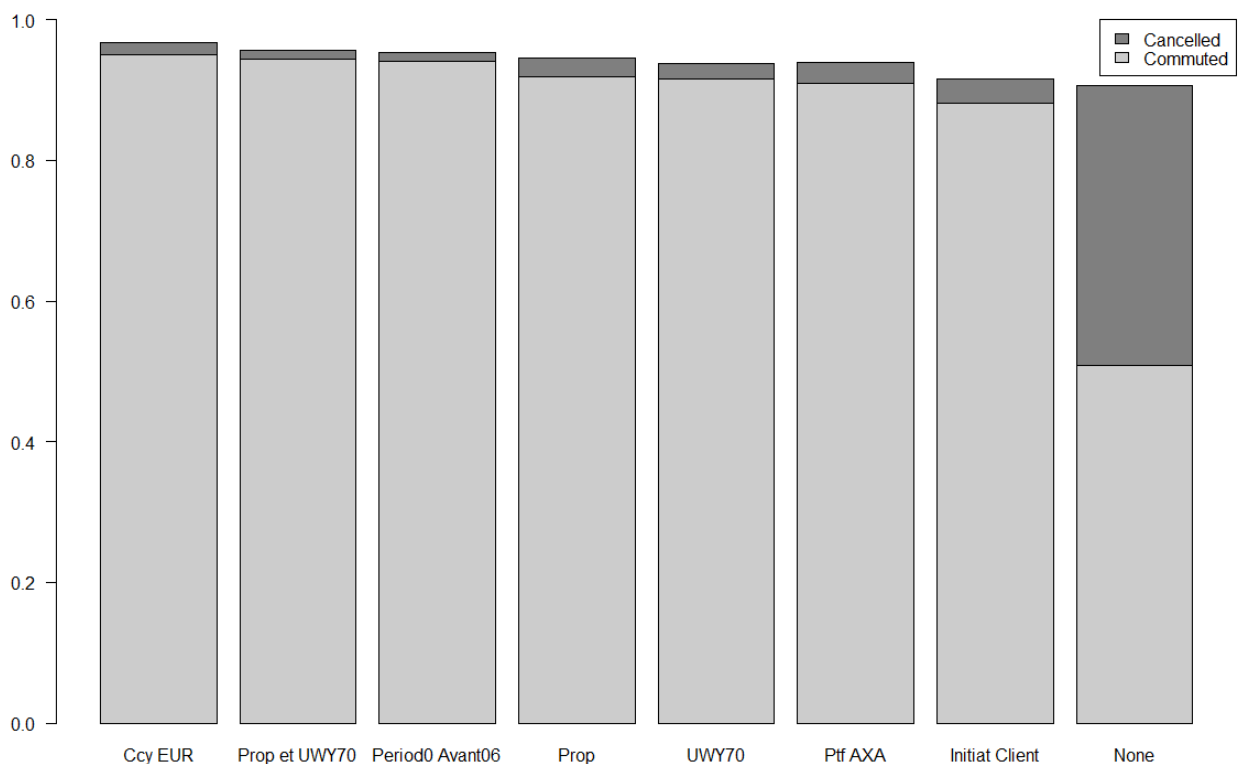


FIGURE 3.6 – Synthèse de l’effet de la segmentation sur l’état stationnaire

### 3.7 Conclusion

Les régressions logistiques réalisées dans ce chapitre nous ont permis de trouver des critères de segmentation pour les projets de commutation. Des tests statistiques ont été réalisés sur les modèles et les variables explicatives, en termes de qualité de modèle (critère AIC), significativité des coefficients et analyse des effets par odds-ratios.

En plus de ces tests statistiques, la fiabilité des résultats a été confirmée en appliquant ces critères aux matrices de transition du modèle de Markov vu au chapitre précédent. En effet, les probabilités de transition vers le statut « commuté » se sont nettement améliorées en appliquant la segmentation. En outre, l'état stationnaire est meilleur qu'avant la segmentation. Et à titre de comparaison, en appliquant le *lemme de Kac*, nous obtenons un temps moyen proche de 1 an (1/0.944) pour atteindre le statut « commuté » en segmentant avec la combinaison années de souscription 1970 et contrats proportionnels alors que ce temps moyen était de 2 ans sans segmentation (voir section 2.4.1).

En conclusion de ce chapitre, nous soulignons l'importance de segmenter selon les différents critères résultant des régressions logistiques réalisées, à savoir : les années de souscription, les types de contrats (proportionnels ou non proportionnels), l'initiateur du projet de commutation, les types de portefeuilles (AXA ou fonds d'investissement), par monnaie et par date d'initiation du projet de commutation.

Cependant, nous n'avons pas de marge de manoeuvre sur certains de ces critères. Par exemple, l'initiateur du projet de commutation ne peut pas être choisi car l'initiative de commuter peut venir de n'importe quelle partie (AXA LM, la cédante, un courtier ou autre tierce partie).

## Conclusion générale

La base de données *Comtool* (base de données des commutations au sein d'AXA LM) n'a jamais été exploitée auparavant dans sa globalité pour réaliser des études statistiques et/ou actuarielles. Ce travail a permis de tirer profit du potentiel lié à l'utilisation de cette base de données. Cependant, certains axes d'amélioration doivent être considérés notamment en lien avec la qualité des données. Malgré le manque de fiabilité de certaines informations, les résultats qui découlent des études réalisées apportent une valeur ajoutée.

Parmi les études réalisées dans ce document, la modélisation par chaînes de Markov révèle, de façon globale, la marge de manoeuvre pour améliorer le processus de commutation existant. En effet, les probabilités de transition vers l'état « commuté » ne sont pas très satisfaisantes et l'état stationnaire montre que 2/5 des projets de commutation sont annulés, et seulement la moitié des projets de commutation aboutissent à la signature d'un accord de commutation. Elle révèle aussi un constat important à prendre en compte par les équipes en charge des commutations, à savoir qu'une fois qu'un projet de commutation entre dans la phase de négociation avec la cédante, la probabilité d'arriver à un accord de commutation est très élevée comparée à d'autres statuts.

Les résultats du modèle de Markov étant globaux, nous avons réalisé des régressions logistiques afin de tenir compte des caractéristiques des contrats, des cédantes et des projets de commutation pour permettre de détecter les influences que peuvent avoir ces caractéristiques sur le processus de commutation. Ainsi, une segmentation des projets a été établie selon différents critères. Les matrices de transition estimées après segmentation montrent une nette amélioration des probabilités de transition vers le statut « commuté » en comparaison avec les matrices sans segmentation. En analysant l'état stationnaire, nous obtenons la même tendance vers une amélioration des commutations abouties après segmentation.

Les différents critères de segmentation qui ont été établis auront un double objectif : d'une part, cela constitue un outil d'aide à la décision pour les équipes chargées des commutations afin de mieux cibler leurs objectifs de commutation sur les portefeuilles existants ; d'autre part, ces critères peuvent être utilisés pour évaluer la qualité des portefeuilles à acquérir.

Ces critères ont été déterminés au vu des observations collectées par l'outil *Comtool*. Comme indiqué dans le premier chapitre sur la qualité des données, certaines informations concernant les cédantes et les contrats sont manquantes. Les croisements entre les différentes informations par contrats, cédantes n'étaient donc pas toujours possibles. Ainsi, les résultats obtenus concernent une partie des projets de commutation et non pas la totalité. Cependant, nous jugeons que cela reste un échantillon assez représentatif car nous avons essayé de regrouper dans des catégories équilibrées en termes d'effectifs. De plus, le seuil de significativité retenu pour les analyses des coefficients du modèle logistique était de 1% afin d'augmenter la fiabilité des conclusions.

Une fois que la fiabilité de la base de données *Comtool* sera améliorée avec l'ajout et/ou la correction des informations contenues aujourd'hui dans cette base de données, il convient de refaire l'analyse sur l'ensemble des projets.

En termes de modèle, nous avons simplifié la modélisation par chaîne de Markov, et ce, en supposant l'homogénéité du processus de commutation. La prise en compte de la variable *Period0* (voir sous-section 3.5) révèle l'importance du moment d'observation du statut de départ d'un projet

de commutation. Bien que cette non homogénéité soit prise en compte dans le modèle logistique, un axe d'amélioration possible de l'étude serait d'utiliser des chaînes de Markov non homogènes afin de voir l'impact sur les probabilités de transition et sur l'état stationnaire.

Enfin, l'objectif de ce mémoire étant d'établir des segmentations pertinentes, il convient par exemple, pour le modèle logistique, de revoir un modèle uniquement basé sur les variables montants qui pourront révéler éventuellement une influence particulière sur le processus de commutation en cours. D'autres variables peuvent être rajoutées au modèle si la base de données *Comtool* est complétée par d'autres informations pertinentes sur le processus de commutation.



## Annexe A

# Annexe : Résultats et code pour le modèle de Markov

### A.1 Vérification de la propriété de Markov

#### A.1.1 Lecture et concaténation des données pour obtenir une seule chaîne de Markov

```
library(readxl)
codedCommutationTransition_S_hmgn <- read_excel("C:/Users/boufgm01/Desktop/Work/
  Transition_S_homogene.xlsx", sheet = "Semesterly_Transition_hmgn")
str(codedCommutationTransition_S_hmgn)
testS_sequence <- c(t(testS[1, -1]))
for (i in 2:nrow(testS)) {
  testS_sequence <- c(testS_sequence, t(testS[i, -1]))
}
```

Listing A.1 – Exemple de création d’une chaîne de Markov sur des transitions semestrielles

#### A.1.2 Vérification de la propriété de Markov sur l’ensemble des données comme une seule chaîne

```
> verifyMarkovProperty(testS_sequence)
Testing markovianity property on given data sequence
Chi - square statistic is: 2694.599 degrees of freedom are: 729 and corresponding
p-value is: 0
```

Listing A.2 – Test du Khi-deux de la propriété de Markov - Exemple sur des transitions semestrielles


### A.1.3 Vérification de la propriété de Markov sur les données projet par projet

```
for (k in 1:nrow(testS)) {verifyMarkovProperty(c(t(testS[k,-1])))}
Chi - square statistic is: 0.02777778 degrees of freedom are: 27 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.02857143 degrees of freedom are: 27 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.5322581 degrees of freedom are: 27 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 1.029412 degrees of freedom are: 64 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.03030303 degrees of freedom are: 64 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.03125 degrees of freedom are: 8 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.03125 degrees of freedom are: 27 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.03125 degrees of freedom are: 8 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.03225806 degrees of freedom are: 27 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.530303 degrees of freedom are: 27 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.5322581 degrees of freedom are: 27 and
corresponding p-value is: 1
Testing markovianity property on given data sequence
Chi - square statistic is: 0.03125 degrees of freedom are: 27 and
corresponding p-value is: 1
```

Listing A.3 – Propriété de Markov projet par projet

## A.2 Code pour l'estimation de la matrice de transition

### A.2.1 Transitions pour 5 états et 4 pas de temps

Cette annexe présente le code  qui permet de modéliser par chaîne de Markov les états ① PRESTU, ② INCONS, ③ NEGOCI, ④ COMACH et ⑤ CANCEL avec différents pas de temps (mensuel, trimestriel, semestriel et annuel). La méthode utilisée est le maximum de vraisemblance.

```
> # modelisation par cha ne de Markov (methode MLE) 5 etats

> codedCommutationTransition_Mmdf_FittedMLE5 <- markovchainFit(data =
  codedCommutationTransition_Mmdf5[,-1], method = "mle", name = "Coded
  Commutation Monthly Modified Transition MLE 5 states")

> codedCommutationTransition_Qmdf_FittedMLE5 <- markovchainFit(data =
  codedCommutationTransition_Qmdf5[,-1], method = "mle", name = "Coded
  Commutation Quarterly Modified Transition MLE 5 states")

> codedCommutationTransition_Smdf_FittedMLE5 <- markovchainFit(data =
  codedCommutationTransition_Smdf5[,-1], method = "mle", name = "Coded
  Commutation Semesterly Modified Transition MLE 5 states")

> codedCommutationTransition_Amdf_FittedMLE5 <- markovchainFit(data =
  codedCommutationTransition_Amdf5[,-1], method = "mle", name = "Coded
  Commutation Annually Modified Transition MLE 5 states")

> # renommer les noms des etats (donner des noms plus explicites) apr s
  modelisation MLE 5 etats

> names(codedCommutationTransition_Mmdf_FittedMLE5$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH", "5_CANCEL")


> names(codedCommutationTransition_Qmdf_FittedMLE5$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH", "5_CANCEL")

> names(codedCommutationTransition_Smdf_FittedMLE5$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH", "5_CANCEL")

> names(codedCommutationTransition_Amdf_FittedMLE5$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH", "5_CANCEL")
```

Listing A.4 – Estimation des matrices de transition pour 5 états et 4 pas de temps

## A.2.2 Transitions pour 4 états et 4 pas de temps

Cette annexe présente le code  qui permet de modéliser par chaîne de Markov les états ① PRESTU, ② INCONS, ③ NEGOCI et ④ COMACH avec différents pas de temps (mensuel, trimestriel, semestriel et annuel). La méthode utilisée est le maximum de vraisemblance.

```
> # modelisation par cha ne de Markov (methode MLE) 4 etats

> codedCommutationTransition_Mmdf_FittedMLE4 <- markovchainFit(data =
  codedCommutationTransition_Mmdf4[,-1], method = "mle", name = "Coded
  Commutation Monthly Modified Transition MLE 4 states")

> codedCommutationTransition_Qmdf_FittedMLE4 <- markovchainFit(data =
  codedCommutationTransition_Qmdf4[,-1], method = "mle", name = "Coded
  Commutation Quarterly Modified Transition MLE 4 states")

> codedCommutationTransition_Smdf_FittedMLE4 <- markovchainFit(data =
  codedCommutationTransition_Smdf4[,-1], method = "mle", name = "Coded
  Commutation Semesterly Modified Transition MLE 4 states")

> codedCommutationTransition_Amdf_FittedMLE4 <- markovchainFit(data =
  codedCommutationTransition_Amdf4[,-1], method = "mle", name = "Coded
  Commutation Annually Modified Transition MLE 4 states")

> # renommer les noms des etats (donner des noms plus explicites) apr s
  modelisation MLE 4 etats

> names(codedCommutationTransition_Mmdf_FittedMLE4$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH")

> names(codedCommutationTransition_Qmdf_FittedMLE4$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH")

> names(codedCommutationTransition_Smdf_FittedMLE4$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH")

> names(codedCommutationTransition_Amdf_FittedMLE4$estimate) <- c("1_PRESTU", "2_
  INCONS", "3_NEGOCI", "4_COMACH")
```

Listing A.5 – Estimation des matrices de transition pour 4 états et 4 pas de temps

## A.3 Affichage des matrices de transition estimées par

### A.3.1 Estimation de la matrice de transition mensuelle

```

> codedCommutationTransition_Mmdf_FittedMLE5
$estimate
Coded Commutation Monthly Modified Transition MLE 5 states
A 5 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH, 5_CANCEL
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
1_PRESTU  0.9574  0.00948  0.00648  0.00615  0.02048
2_INCONS  0.0028  0.97327  0.00925  0.00492  0.00982
3_NEGOCI  0.0017  0.00308  0.92810  0.04713  0.01999
4_COMACH  0.0015  0.00054  0.00035  0.99743  0.00018
5_CANCEL  0.0026  0.00100  0.00047  0.00031  0.99566

$standardError
      1      2      3      4      5
1 4.1e-03 4.1e-04 3.4e-04 3.3e-04 6.1e-04
2 2.3e-04 4.3e-03 4.2e-04 3.1e-04 4.3e-04
3 3.1e-04 4.1e-04 7.1e-03 1.6e-03 1.0e-03
4 5.0e-05 3.0e-05 2.4e-05 1.3e-03 1.7e-05
5 7.4e-05 4.7e-05 3.2e-05 2.6e-05 1.5e-03

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4      5
1 0.9493 0.00867 0.00581 0.00550 0.01930
2 0.0023 0.96484 0.00843 0.00432 0.00897
3 0.0011 0.00227 0.91410 0.04398 0.01794
4 0.0014 0.00048 0.00030 0.99490 0.00014
5 0.0024 0.00091 0.00041 0.00026 0.99278

$upperEndpointMatrix
      1      2      3      4      5
1 0.9655 0.0103 0.00714 0.00680 0.02167
2 0.0032 0.9817 0.01007 0.00552 0.01066
3 0.0023 0.0039 0.94209 0.05028 0.02205
4 0.0016 0.0006 0.00039 0.99996 0.00021
5 0.0027 0.0011 0.00053 0.00036 0.99854

```

Listing A.6 – Estimation de la matrice de transition mensuelle à 5 états

```

> codedCommutationTransition_Mmdf_FittedMLE4
$estimate
Coded Commutation Monthly Modified Transition MLE 4 states
A 4 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
1_PRESTU  0.9640  0.01211  0.00801  0.0159
2_INCONS  0.0018  0.98498  0.00697  0.0062
3_NEGOCI  0.0012  0.00133  0.93659  0.0609
4_COMACH  0.0011  0.00046  0.00038  0.9981

$standardError
      1      2      3      4
1 7.7e-03 8.6e-04 7.0e-04 0.00099
2 2.5e-04 5.9e-03 4.9e-04 0.00047
3 3.8e-04 4.0e-04 1.1e-02 0.00271
4 4.9e-05 3.2e-05 2.9e-05 0.00150

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4
1 0.94894 0.01042 0.00664 0.0140
2 0.00133 0.97346 0.00600 0.0053
3 0.00046 0.00054 0.91574 0.0556
4 0.00097 0.00040 0.00032 0.9952

$upperEndpointMatrix
      1      2      3      4
1 0.9790 0.01379 0.00938 0.0178
2 0.0023 0.99649 0.00794 0.0071
3 0.0020 0.00211 0.95743 0.0662
4 0.0012 0.00053 0.00043 1.0000

```

Listing A.7 – Estimation de la matrice de transition mensuelle à 4 états

### A.3.2 Estimation de la matrice de transition trimestrielle

```

> codedCommutationTransition_Qmdf_FittedMLE5
$estimate
Coded Commutation Quarterly Modified Transition MLE 5 states
A 5 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH, 5_CANCEL
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
1_PRESTU  0.8902  0.0186  0.0153  0.0181  0.05792
2_INCONS  0.0074  0.9314  0.0166  0.0158  0.02879
3_NEGOCI  0.0046  0.0084  0.8234  0.1094  0.05411
4_COMACH  0.0038  0.0016  0.0011  0.9929  0.00067
5_CANCEL  0.0065  0.0030  0.0015  0.0013  0.98765

$standardError
      1      2      3      4      5
1 0.00693 1.0e-03 9.1e-04 9.9e-04 1.8e-03
2 0.00065 7.3e-03 9.7e-04 9.5e-04 1.3e-03
3 0.00088 1.2e-03 1.2e-02 4.3e-03 3.0e-03
4 0.00014 8.8e-05 7.3e-05 2.2e-03 5.8e-05
5 0.00020 1.4e-04 1.0e-04 9.3e-05 2.5e-03

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4      5
1 0.8766 0.0166 0.01348 0.0161 0.05445
2 0.0061 0.9171 0.01474 0.0139 0.02628
3 0.0029 0.0061 0.80055 0.1010 0.04825
4 0.0035 0.0014 0.00094 0.9886 0.00056
5 0.0061 0.0027 0.00135 0.0012 0.98270

$upperEndpointMatrix
      1      2      3      4      5
1 0.9037 0.0206 0.0170 0.0200 0.06138
2 0.0087 0.9456 0.0186 0.0177 0.03130
3 0.0063 0.0108 0.8463 0.1177 0.05998
4 0.0040 0.0017 0.0012 0.9973 0.00079
5 0.0069 0.0033 0.0017 0.0015 0.99259

```

Listing A.8 – Estimation de la matrice de transition trimestrielle à 5 états

```

> codedCommutationTransition_Qmdf_FittedMLE4
$estimate
Coded Commutation Quarterly Modified Transition MLE 4 states
A 4 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
1_PRESTU  0.9125  0.0242  0.0182  0.045
2_INCONS  0.0045  0.9639  0.0126  0.019
3_NEGOCI  0.0039  0.0036  0.8498  0.143
4_COMACH  0.0025  0.0013  0.0012  0.995

$standardError
      1      2      3      4
1 0.01288 2.1e-03 1.8e-03 0.0029
2 0.00069 1.0e-02 1.1e-03 0.0014
3 0.00119 1.1e-03 1.7e-02 0.0072
4 0.00013 9.2e-05 8.8e-05 0.0026

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4
1 0.8873 0.0201 0.01462 0.039
2 0.0032 0.9442 0.01031 0.016
3 0.0016 0.0014 0.81556 0.129
4 0.0022 0.0011 0.00098 0.990

$upperEndpointMatrix
      1      2      3      4
1 0.9378 0.0283 0.0217 0.051
2 0.0058 0.9836 0.0148 0.022
3 0.0063 0.0058 0.8840 0.157
4 0.0028 0.0014 0.0013 1.000

```

Listing A.9 – Estimation de la matrice de transition trimestrielle à 4 états



### A.3.3 Estimation de la matrice de transition semestrielle

```

> codedCommutationTransition_Smdf_FittedMLE5
$estimate
Coded Commutation Semesterly Modified Transition MLE 5 states
A 5 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH, 5_CANCEL
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
1_PRESTU  0.8150  0.0218  0.0185  0.0341  0.1105
2_INCONS  0.0117  0.8875  0.0167  0.0295  0.0547
3_NEGOCI  0.0082  0.0102  0.7267  0.1644  0.0904
4_COMACH  0.0060  0.0029  0.0022  0.9872  0.0017
5_CANCEL  0.0109  0.0055  0.0033  0.0036  0.9766

$standardError
      1      2      3      4      5
1 0.00943 0.00154 0.00142 0.00193 0.00347
2 0.00116 0.01007 0.00138 0.00184 0.00250
3 0.00164 0.00183 0.01546 0.00735 0.00545
4 0.00025 0.00017 0.00015 0.00314 0.00013
5 0.00037 0.00027 0.00021 0.00022 0.00355

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4      5
1 0.7965 0.0188 0.0157 0.0303 0.1037
2 0.0094 0.8677 0.0140 0.0259 0.0498
3 0.0050 0.0066 0.6964 0.1500 0.0797
4 0.0055 0.0026 0.0019 0.9810 0.0015
5 0.0101 0.0050 0.0029 0.0032 0.9697

$upperEndpointMatrix
      1      2      3      4      5
1 0.8335 0.0248 0.0213 0.0379 0.117
2 0.0139 0.9072 0.0194 0.0331 0.060
3 0.0114 0.0138 0.7570 0.1788 0.101
4 0.0065 0.0032 0.0025 0.9933 0.002
5 0.0116 0.0061 0.0037 0.0041 0.984

```

Listing A.10 – Estimation de la matrice de transition semestrielle à 5 états

```

> codedCommutationTransition_Smdf_FittedMLE4
$estimate
Coded Commutation Semesterly Modified Transition MLE 4 states
A 4 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
1_PRESTU  0.8644  0.0293  0.0257  0.081
2_INCONS  0.0073  0.9430  0.0124  0.037
3_NEGOCI  0.0064  0.0056  0.7721  0.216
4_COMACH  0.0036  0.0023  0.0022  0.992

$standardError
      1      2      3      4
1 0.01780 0.00328 0.00307 0.0054
2 0.00124 0.01405 0.00161 0.0028
3 0.00212 0.00200 0.02334 0.0123
4 0.00022 0.00017 0.00017 0.0036

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4
1 0.8295 0.0229 0.0196 0.070
2 0.0049 0.9155 0.0092 0.032
3 0.0022 0.0017 0.7263 0.192
4 0.0032 0.0019 0.0019 0.985

$upperEndpointMatrix
      1      2      3      4
1 0.8993 0.0357 0.0317 0.091
2 0.0098 0.9706 0.0155 0.043
3 0.0105 0.0096 0.8178 0.240
4 0.0041 0.0026 0.0026 0.999

```

Listing A.11 – Estimation de la matrice de transition semestrielle à 4 états

### A.3.4 Estimation de la matrice de transition annuelle

```

> codedCommutationTransition_Amdf_FittedMLE5
$estimate
Coded Commutation Annually Modified Transition MLE 5 states
A 5 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH, 5_CANCEL
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
1_PRESTU  0.701  0.0210  0.0192  0.065  0.1944
2_INCONS  0.018  0.8405  0.0136  0.046  0.0824
3_NEGOCI  0.014  0.0108  0.6312  0.209  0.1342
4_COMACH  0.009  0.0047  0.0036  0.978  0.0048
5_CANCEL  0.016  0.0076  0.0045  0.010  0.9614

$standardError
      1      2      3      4      5
1 0.01251 0.00216 0.00207 0.0038 0.00659
2 0.00201 0.01393 0.00177 0.0033 0.00436
3 0.00311 0.00271 0.02069 0.0119 0.00954
4 0.00042 0.00030 0.00026 0.0044 0.00031
5 0.00064 0.00044 0.00033 0.0005 0.00489

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4      5
1 0.6762 0.0167 0.0151 0.0573 0.1815
2 0.0136 0.8131 0.0101 0.0396 0.0739
3 0.0081 0.0055 0.5906 0.1861 0.1155
4 0.0082 0.0041 0.0031 0.9694 0.0042
5 0.0151 0.0068 0.0038 0.0091 0.9519

$upperEndpointMatrix
      1      2      3      4      5
1 0.7252 0.0252 0.0233 0.072 0.2073
2 0.0215 0.8678 0.0171 0.052 0.0910
3 0.0203 0.0162 0.6717 0.233 0.1529
4 0.0098 0.0052 0.0041 0.986 0.0054
5 0.0176 0.0085 0.0051 0.011 0.9710

```

Listing A.12 – Estimation de la matrice de transition annuelle à 5 états

```

> codedCommutationTransition_Amdf_FittedMLE4
$estimate
Coded Commutation Annually Modified Transition MLE 4 states
A 4 - dimensional discrete Markov Chain defined by the following states:
1_PRESTU, 2_INCONS, 3_NEGOCI, 4_COMACH
The transition matrix (by rows) is defined as follows:
      1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
1_PRESTU  0.8043  0.0256  0.0228  0.147
2_INCONS  0.0106  0.9181  0.0123  0.059
3_NEGOCI  0.0097  0.0097  0.6911  0.289
4_COMACH  0.0048  0.0037  0.0035  0.988

$standardError
      1      2      3      4
1 0.02393 0.00427 0.0040 0.0102
2 0.00209 0.01939 0.0022 0.0049
3 0.00366 0.00366 0.0309 0.0200
4 0.00036 0.00031 0.0003 0.0051

$confidenceLevel
[1] 0.95

$lowerEndpointMatrix
      1      2      3      4
1 0.7574 0.0173 0.0149 0.127
2 0.0066 0.8801 0.0079 0.049
3 0.0025 0.0025 0.6305 0.250
4 0.0041 0.0031 0.0029 0.978

$upperEndpointMatrix
      1      2      3      4
1 0.8512 0.0340 0.0307 0.167
2 0.0147 0.9561 0.0167 0.069
3 0.0169 0.0169 0.7518 0.329
4 0.0055 0.0043 0.0041 0.998

```

Listing A.13 – Estimation de la matrice de transition annuelle à 4 états

### A.3.5 État stationnaire

#### Cas 5 états

```
> steadyStates(codedCommutationTransition_Mmdf_FittedMLE5$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
[1,]  0.045  0.043  0.015  0.5  0.4
```

Listing A.14 – État stationnaire - Transition mensuelle

```
> steadyStates(codedCommutationTransition_Qmdf_FittedMLE5$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
[1,]  0.044  0.042  0.014  0.5  0.4
```

Listing A.15 – État stationnaire - Transition trimestrielle

```
> steadyStates(codedCommutationTransition_Smdf_FittedMLE5$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
[1,]  0.043  0.042  0.014  0.51  0.39
```

Listing A.16 – État stationnaire - Transition semestrielle

```
> steadyStates(codedCommutationTransition_Amdf_FittedMLE5$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH 5_CANCEL
[1,]  0.04  0.04  0.013  0.51  0.4
```

Listing A.17 – État stationnaire - Transition annuelle

#### Cas 4 états

```
> steadyStates(codedCommutationTransition_Mmdf_FittedMLE4$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
[1,]  0.03  0.053  0.015  0.9
```

Listing A.18 – État stationnaire - Transition mensuelle

```
> steadyStates(codedCommutationTransition_Qmdf_FittedMLE4$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
[1,]  0.029  0.053  0.015  0.9
```

Listing A.19 – État stationnaire - Transition trimestrielle

```
> steadyStates(codedCommutationTransition_Smdf_FittedMLE4$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
[1,]  0.028  0.052  0.015  0.91
```

Listing A.20 – État stationnaire - Transition semestrielle

```
> steadyStates(codedCommutationTransition_Amdf_FittedMLE4$estimate)
  1_PRESTU 2_INCONS 3_NEGOCI 4_COMACH
[1,]  0.026  0.051  0.014  0.91
```


Listing A.21 – État stationnaire - Transition annuelle



## Annexe B

# Annexe : Résultats détaillés du modèle logistique et estimation des matrices de transition avec segmentation

### B.1 Effectifs des modalités des variables qualitatives utilisées dans le modèle logistique

Dans les sorties  ci-dessous, la première ligne de chacun des résultats de la fonction `freq` indique la modalité de référence.

#### B.1.1 Effectifs des modalités de la variable d'intérêt

```
> freq(Data_Logistic_Static_excl_6_7$Last_Status)
  n   % val%
5 1907 29.8 29.8
4 3519 54.9 54.9
3   22  0.3  0.3
2  641 10.0 10.0
1  316  4.9  4.9

> freq(Data_Logistic_Static_excl_6_7$Last_Status_Mdf)
  n   % val%
Can 1907 29.8 29.8
Com 3519 54.9 54.9
InP  979 15.3 15.3
```

Listing B.1 – Fréquence des modalités de la variable d'intérêt

## B.1.2 Effectifs des modalités des variables explicatives

```

> freq(Data_Logistic_Static_excl_6_7$PnP)
      n    % val%
NP 4771 74.5 74.5
P  1634 25.5 25.5

> freq(Data_Logistic_Static_excl_6_7$UWY)
      n    % val%
UWY90 1950 30.4 30.4
UWY80 3067 47.9 47.9
UWY70  722 11.3 11.3
UWY60  666 10.4 10.4

> freq(Data_Logistic_Static_excl_6_7$Ptf)
      n    % val%
AXA  5960 93.1 93.1
Fund  445  6.9  6.9

> freq(Data_Logistic_Static_excl_6_7$Initiat)
      n    % val%
AXA LM    6293 98.3 98.3
Client     49  0.8  0.8
Third Party 63  1.0  1.0

> freq(Data_Logistic_Static_excl_6_7$Vol)
      n    % val%
Non Volatility    6036 94.2 94.2
Volatility Confirmed    295  4.6  4.6
Volatility to be confirmed    74  1.2  1.2

> freq(Data_Logistic_Static_excl_6_7$Ccy)
      n    % val%
USD  2380 37.2 37.2
EUR  1977 30.9 30.9
Oth  2048 32.0 32.0

> freq(Data_Logistic_Static_excl_6_7$Rating_Mdf)
      n    % val%
High  1329 20.7 20.7
Low   1030 16.1 16.1
None  4046 63.2 63.2

> freq(Data_Logistic_Static_excl_6_7$Country_Mdf)
      n    % val%
US   738 11.5 11.5
EU   600  9.4  9.4
OT  4322 67.5 67.5
AS   745 11.6 11.6

```

Listing B.2 – Fréquence des modalités des variables explicatives



## B.2 Résultats et code du modèle logistique

### B.2.1 Choix de modèle par sélection pas à pas descendante

```
# modelisation logistique multinomiale - modele initial

> Static_Model_excl_6_7 <- multinom>Last_Status ~ PnP + UWY + Ptf + Initiat + Vol +
  Ccy + Rating_Mdf + Country_Mdf + Size_Mdf + Nb_Contrat + ASB + POL + HH +
  abPaid_ASB_m + abPaid_POL_m + abPaid_HH_m + abOS_ASB_m + abOS_POL_m + abOS_HH_m
  + abOS_m + abIBNR_m + abDEPOT_m + abBALANCES_m + abPRICE_m, data = Data_
  Logistic_Static_excl_6_7)

> AIC(Static_Model_excl_6_7)

[1] 8156.153
```

Listing B.3 – Exemple de choix de modèle par sélection pas à pas descendante - modèle initial

```
# choix de modele par selection pas a pas descendante

step(Static_Model_excl_6_7) # calcul du meilleur modele
```

	Df	AIC
<none>	108	7969.442
- abOS_POL_m	104	7970.062
- abOS_ASB_m	104	7979.789
- abDEPOT_m	104	7984.636
- POL	104	7985.863
- Country_Mdf	96	7988.608
- PnP	104	7997.317
- Rating_Mdf	100	7999.667
- ASB	104	8002.209
- Initiat	100	8009.888
- abOS_m	104	8036.233
- abIBNR_m	104	8056.130
- Ptf	104	8076.500
- Size_Mdf	100	8157.101
- Vol	100	8202.716
- UWY	96	8209.110
- abBALANCES_m	104	8377.186
- Ccy	100	9827.219

```
Call:
multinom(formula = Last_Status ~ PnP + UWY + Ptf + Initiat +
  Vol + Ccy + Rating_Mdf + Country_Mdf + Size_Mdf + ASB + POL +
  abOS_ASB_m + abOS_POL_m + abOS_m + abIBNR_m + abDEPOT_m +
  abBALANCES_m, data = Data_Logistic_Static_excl_6_7)
```

```

Coefficients:
 (Intercept)      PnPP      UWYUWY80      UWYUWY70      UWYUWY60      PtfFund
4  -0.6222413    0.5654299  -0.02556758   0.7097284  -0.6232960  -1.7094164
3  -12.0354986  -0.2677355   0.72211322   2.3121643  -0.7685528   1.1391678
2  -2.5183871   0.4630426  -0.41133689   0.2377333  -2.1291748  -0.2424484
1  -2.3576375   0.1202433  -0.96432524  -0.2137364  -1.7157910  -0.1686307
  InitiatClient  InitiatThird Party Vol  Confirmed Vol to be confirmed
4      3.170516           2.82747447          -1.186272          -4.714464
3     -1.626628           0.03520732           1.070305          -4.911671
2      0.941127          -24.27635952         -1.059104         -44.520305
1      1.863024           1.12238061          -4.284926          -5.922150
      CcyEUR      CcyOth  Rating_MdfLow  Rating_MdfNone
4   1.7821000  -2.0706782     0.3956113     -0.1469261
3  -1.0443126  -1.5701802     7.6141428     8.7648267
2  -0.9985887   0.4778578     1.0877988     1.4691945
1  -0.7673230  -0.9156816    -0.4031738    -0.1600192
  Country_MdfEU  Country_MdfOT  Country_MdfAS  Size_MdfMid  Size_MdfHigh
4     0.4785204     0.6840529     0.6873663     1.375931     1.580228
3     1.2010610    -0.6083961     0.8324687    -5.616984    -6.807772
2     0.7693423     0.5137211     0.9550247   -60.953779   -1.689513
1     1.3075578     2.6173171     1.9024806    -3.907741    -6.832383
      ASB      POL  abOS_ASB_m  abOS_POL_m  abOS_m
4  -0.008325925   0.1261712  -205.493203  -119.565999  -0.04579051
3  -0.049331441  -2.4044995   2.236748   -7.466569   0.38595929
2  -0.302999590   0.3336156  16.824148  -39.475458   0.08231518
1   0.030315576  -0.5225546  16.749653   1.802919   0.31391587
  abIBNR_m  abDEPOT_m  abBALANCES_m
4  -0.3473398   1.4296668     7.918489
3  -1.2655788   0.3088284     6.995145
2   0.4413888   2.4681538    -44.758918
1  -1.7262710   2.0023896    -2.954286

Residual Deviance: 7753.442
AIC: 7969.442
    
```

Listing B.4 – Procédure de sélection pas à pas descendante

```
# modelisation logistique multinomiale - modele apres selection pas a pas
  descendante

> Static_Model_excl_6_7_Select <- multinom(formula = Last_Status ~ PnP + UWY + Ptf
  + Initiat + Vol + Ccy + Rating_Mdf + Country_Mdf + Size_Mdf + ASB + POL + abOS_
  ASB_m + abOS_POL_m + abOS_m + abIBNR_m + abDEPOT_m + abBALANCES_m, data = Data_
  Logistic_Static_excl_6_7)

# calcul des exponentiels pour permettre une interpr tabilit plus facile (Odd
  Ratios)

> round(exp(coef(Static_Model_excl_6_7_Select)), digits = 4)

(Intercept)    PnP UWYUWY80 UWYUWY70 UWYUWY60 PtfFund
4      0.5367 1.7602   0.9748   2.0334   0.5362  0.1810
3      0.0000 0.7651   2.0588  10.0963   0.4637  3.1242
2      0.0806 1.5889   0.6628   1.2684   0.1189  0.7847
1      0.0946 1.1278   0.3812   0.8076   0.1798  0.8448
  InitiatClient InitiatThird Party Vol Confirmed Vol to be confirmed
4      23.8198                16.9027           0.3054           0.0090
3       0.1966                1.0358           2.9163           0.0074
2       2.5629                0.0000           0.3468           0.0000
1       6.4432                3.0722           0.0138           0.0027
  CcyEUR Ccy0th Rating_MdfLow Rating_MdfNone
4  5.9423 0.1261           1.4853           0.8634
3  0.3519 0.2080       2026.6568           6404.9518
2  0.3684 1.6126           2.9677           4.3457
1  0.4643 0.4002           0.6682           0.8521
  Country_MdfEU Country_MdfOT Country_MdfAS Size_MdfMid Size_MdfHigh
4       1.6137           1.9819           1.9885           3.9588           4.8561
3       3.3236           0.5442           2.2990           0.0036           0.0011
2       2.1583           1.6715           2.5987           0.0000           0.1846
1       3.6971           13.6989           6.7025           0.0201           0.0011
  ASB    POL    abOS_ASB_m    abOS_POL_m    abOS_m
4  0.9917 1.1345 0.000000e+00      0.0000  0.9552
3  0.9519 0.0903 9.362800e+00      0.0006  1.4710
2  0.7386 1.3960 2.025978e+07      0.0000  1.0858
1  1.0308 0.5930 1.880536e+07      6.0673  1.3688
  abIBNR_m    abDEPOT_m    abBALANCES_m
4   0.7066      4.1773      2747.6166
3   0.2821      1.3618      1091.3223
2   1.5549     11.8006           0.0000
1   0.1779      7.4067           0.0521
```

Listing B.5 – Odds-ratios du modèle sélectionné

```
# modelisation logistique multinomiale - modele final retenu

> Static_Model_excl_6_7_Select <- multinom(formula = Last_Status ~ PnP + UWY + Ptf
+ Vol + Initiat + Ccy + Country_Mdf + ASB + POL + abOS_m + abIBNR_m, data =
Data_Logistic_Static_excl_6_7)

# calcul des exponentiels pour permettre une interpr tabilit plus facile (Odd
Ratios)

> round(exp(coef(Static_Model_excl_6_7_Select)), digits = 4)

(Intercept)    PnP UWYUWY80 UWYUWY70 UWYUWY60 PtfFund
4      2.4149 1.5040    0.8880    2.1122    0.4814 0.2555
3      0.0138 0.7860    1.8744   10.0161    0.4329 2.7348
2      0.0706 1.5944    0.7116    1.3996    0.1336 0.6188
1      0.1199 1.1353    0.3822    0.7574    0.1789 0.7842
Vol Confirmed Vol to be confirmed InitiatClient InitiatThird Party
4      0.6131                0.0105          24.8844          8.8155
3      3.6416                0.0000          0.0000          0.0000
2      0.0007                0.0000          1.7766          0.0000
1      0.0000                0.0000          10.3523          2.4947
CcyEUR CcyOth Country_MdfEU Country_MdfOT Country_MdfAS
4 4.1122 0.0647    1.0288    0.9980    1.2992
3 0.2705 0.1208    2.3102    2.4104    1.3448
2 0.6824 2.2706    3.0540    5.0342    3.1461
1 0.3863 0.2682    3.6138   12.9637    7.4351
ASB    POL abOS_m abIBNR_m
4 0.9821 1.0677 2.1854 0.4233
3 1.0014 0.0000 0.0000 0.0000
2 0.8821 1.2235 1.3639 1.2196
1 1.0875 0.6497 0.0000 0.0000
```

Listing B.6 – Odds-ratios du modèle final retenu

## B.2.2 Modèle logistique statique

### *Modèle statique - Variable d'intérêt avec tous les statuts de 1 à 5*

```
# modelisation logistique multinomiale - modele statique statuts de 1 a 5

> Static_Model_excl_6_7_Select <- multinom(formula = Last_Status ~ PnP + UWY + Ptf
+ Vol + Initiat + Ccy + Country_Mdf + ASB + POL + abOS_m + abIBNR_m, data =
Data_Logistic_Static_excl_6_7)

# calcul des exponentiels pour permettre une interpr tabilit plus facile (Odd
Ratios)

> round(exp(coef(Static_Model_excl_6_7_Select)), digits = 4)

(Intercept)    PnP UWYUWY80 UWYUWY70 UWYUWY60 PtfFund
4      2.4149 1.5040    0.8880    2.1122    0.4814    0.2555
3      0.0138 0.7860    1.8744   10.0161    0.4329    2.7348
2      0.0706 1.5944    0.7116    1.3996    0.1336    0.6188
1      0.1199 1.1353    0.3822    0.7574    0.1789    0.7842
Vol Confirmed Vol to be confirmed InitiatClient InitiatThird Party
4      0.6131                0.0105          24.8844          8.8155
3      3.6416                0.0000          0.0000          0.0000
2      0.0007                0.0000          1.7766          0.0000
1      0.0000                0.0000          10.3523          2.4947
CcyEUR Ccy0th Country_MdfEU Country_MdfOT Country_MdfAS
4 4.1122 0.0647    1.0288    0.9980    1.2992
3 0.2705 0.1208    2.3102    2.4104    1.3448
2 0.6824 2.2706    3.0540    5.0342    3.1461
1 0.3863 0.2682    3.6138   12.9637    7.4351
ASB    POL abOS_m abIBNR_m
4 0.9821 1.0677 2.1854 0.4233
3 1.0014 0.0000 0.0000 0.0000
2 0.8821 1.2235 1.3639 1.2196
1 1.0875 0.6497 0.0000 0.0000
```

Listing B.7 – Modèle statique avec tous les statuts observés à fin juin 2019

### *Diagramme des odds-ratios du modèle statique avec tous les statuts observés à fin juin 2019*

La fonction `forest_model` permet de tracer le diagramme des odds-ratios, leur intervalle de confiance, les *p-value* ainsi que les références des modalités de chaque variable explicative. Le diagramme issu de ce modèle n'est malheureusement pas assez lisible, c'est la raison pour laquelle nous préférons l'afficher dans cette annexe plutôt que dans le corps du mémoire.



*Modèle statique - Variable d'intérêt avec regroupement de statuts*

```
# modelisation logistique multinomiale - modele statique statuts regroupes

> Static_Model_Regroup_Select <- multinom(formula = Last_Status_Mdf ~ PnP + UWY +
  Ptf + Initiat + Ccy + Rating_Mdf +Country_Mdf + Size_Mdf + POL + abOS_m +
  abIBNR_m, data = Data_Logistic_Static_excl_6_7)

# calcul AIC du modele

> AIC(Static_Model_Regroup_Select)

[1] 7441.104

# calcul des exponentiels pour permettre une interpretabilite plus facile (Odd
  Ratios)

> round(exp(coef(Static_Model_Regroup_Select)), digits = 4)

      (Intercept)   PnP  UWYUWY80  UWYUWY70  UWYUWY60
Com      0.7798 1.7017   0.9514   2.4419   0.5420
InP      0.1838 1.4318   0.5726   1.0845   0.1437
  PtfFund InitiatClient InitiatThird Party
Com  0.2289      11.6452           10.1937
InP  1.0102      2.4484           1.2698
  CcyEUR CcyOth Rating_MdfLow Rating_MdfNone
Com  7.9992 0.1354      1.3979      0.8486
InP  0.5695 1.0737      1.6744      2.3301
  Country_MdfEU Country_MdfOT Country_MdfAS
Com      1.0358      1.3370      1.4227
InP      2.6507      2.9237      3.2391
  Size_MdfMid Size_MdfHigh      POL abOS_m abIBNR_m
Com      3.3438      10.2483 1.0284 1.3767  0.5382
InP      0.0064      0.1865 1.0641 1.2657  0.8394
```

Listing B.8 – Modèle statique avec statuts regroupés observés à fin juin 2019

## B.2.3 Modèle logistique dynamique

### *Modèle dynamique y compris la variable Period0*

```
# modelisation logistique multinomiale - modele dynamique avec Period0

> Dynamic_Model_Final_Select_Period <- multinom(Var_Int_Final ~ PnP + UWY + Ptf +
  Initiat + Vol + Ccy + Country_Mdf + Size_Mdf + Period0 + ASB + POL + HH + abOS_
  ASB_m + abIBNR_m, data = Data_Logistic_Dynamic[Data_Logistic_Dynamic$Var_Int_
  Final %in% c("A_A", "A_B", "A_C"),])

# calcul AIC du modele

> AIC(Dynamic_Model_Final_Select_Period)

[1] 4566.795

# calcul des exponentiels pour permettre une interpretabilite plus facile (Odd
  Ratios)

> round(exp(coef(Dynamic_Model_Final_Select_Period)), digits = 4)

      (Intercept)   PnP UWYUWY80 UWYUWY70 UWYUWY60
A_B      0.1189 1.3995   1.2839   1.9072   1.0761
A_C      0.0064 0.8836   1.3644   1.5944   1.1356
  PtfFund InitiatClient InitiatThird Party
A_B  0.8090      1.3204                2.0519
A_C  0.8457      1.5221                0.0000
  Vol Confirmed Vol to be confirmed CcyEUR CcyOth
A_B      0.7867                0.0000 2.9815 0.2003
A_C      0.7797                4.2631 0.8487 0.0137
  Country_MdfEU Country_MdfOT Country_MdfAS Size_MdfMid Size_MdfHigh
A_B      0.7355      1.3337      1.0229      2.5812      0.8035
A_C      4.7896      3.2074      1.6911      1.0562      0.0000
  Period0Ent0612 Period0Après12   ASB   POL   HH abOS_ASB_m abIBNR_m
A_B      3.5225      0.8455 0.9646 1.1091 0.9507      0.0000   1.3881
A_C     14.1163     24.1466 1.0250 0.9608 1.0807      1.0346   0.0132
```

Listing B.9 – Modèle dynamique avec la variable Period0



*Modèle dynamique sans la variable Period0*

```
# modelisation logistique multinomiale - modele dynamique sans Period0

> Dynamic_Model_Final_Select <- multinom(Var_Int_Final ~ PnP + UWY + Ptf + Initiat
+ Vol + Ccy + Country_Mdf + Size_Mdf + ASB + POL + HH + abOS_ASB_m + abIBNR_m,
data = Data_Logistic_Dynamic[Data_Logistic_Dynamic$Var_Int_Final %in% c("A_A",
"A_B", "A_C"),])

# calcul AIC du modele

> AIC(Dynamic_Model_Final_Select)

[1] 4800.565

# calcul des exponentiels pour permettre une interpretabilite plus facile (Odd
Ratios)

> round(exp(coef(Dynamic_Model_Final_Select)), digits = 4)

      (Intercept)   PnP UWYUWY80 UWYUWY70 UWYUWY60
A_B      0.1696 1.3598   1.2700   1.8774   1.0125
A_C      0.0965 0.8803   1.0579   1.1080   1.1546
      PtfFund InitiatClient InitiatThird Party
A_B  0.4847         1.6633             2.4652
A_C  1.3221         2.0119             0.0000
      Vol Confirmed Vol to be confirmed CcyEUR CcyOth
A_B      0.4927             0.0000 2.3832 0.1507
A_C      1.0598             3.1017 0.2696 0.0140
      Country_MdfEU Country_MdfOT Country_MdfAS Size_MdfMid Size_MdfHigh
A_B      0.7568             1.2802             1.0635         4.0048         1.9472
A_C      7.1292             4.1235             1.8289         1.1452         0.0000
      ASB   POL   HH abOS_ASB_m abIBNR_m
A_B 0.9642 1.1117 0.9418         0.0000         1.3342
A_C 1.0035 1.0107 1.0471         0.4741         0.0026
```

Listing B.10 – Modèle dynamique sans la variable Period0

## B.2.4 Matrices de transition du modèle de Markov avec segmentation

```
# Markov avec segmentation PnP = P

> Markov_Segment_Prop_Amdf5 <- markovchainFit(data = subset(Data_Markov_Seg_Amdf5[
  Data_Markov_Seg_Amdf5$PnP == "P",], select = -c(ID_FILE, PnP, UWY, Ptf, Initiat
  , Ccy, Period0) ), method = "mle", name = "Markov Segmentation PnP = P")
> Markov_Segment_Prop_Amdf5$estimate
Markov Segmentation PnP = P
A 5 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5
1 0.742904841 0.031719533 0.021702838 0.15692821 0.0467445743
2 0.020249221 0.833333333 0.021806854 0.09968847 0.0249221184
3 0.013043478 0.004347826 0.613043478 0.35217391 0.0173913043
4 0.005180546 0.002783577 0.002165004 0.98909766 0.0007732158
5 0.023592715 0.016970199 0.010347682 0.04387417 0.9052152318

# Markov avec segmentation UWY = UWY70

> Markov_Segment_UWY70_Amdf5 <- markovchainFit(data = subset(Data_Markov_Seg_Amdf5[
  Data_Markov_Seg_Amdf5$UWY == "UWY70",], select = -c(ID_FILE, PnP, UWY, Ptf,
  Initiat, Ccy, Period0) ), method = "mle", name = "Markov Segmentation UWY70")
> Markov_Segment_UWY70_Amdf5$estimate
Markov Segmentation UWY70
A 5 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5
1 0.653061224 0.030612245 0.040816327 0.23469388 0.0408163265
2 0.011299435 0.887005650 0.028248588 0.06214689 0.0112994350
3 0.000000000 0.000000000 0.700000000 0.30000000 0.0000000000
4 0.005898789 0.003104626 0.001862776 0.98820242 0.0009313878
5 0.017441860 0.014534884 0.005813953 0.05232558 0.9098837209

# Markov avec segmentation Ptf = AXA

> Markov_Segment_PtfAXA_Amdf5 <- markovchainFit(data = subset(Data_Markov_Seg_Amdf5
  [Data_Markov_Seg_Amdf5$Ptf == "AXA",], select = -c(ID_FILE, PnP, UWY, Ptf,
  Initiat, Ccy, Period0) ), method = "mle", name = "Markov Segmentation Ptf = AXA
  ")
> Markov_Segment_PtfAXA_Amdf5$estimate
Markov Segmentation Ptf = AXA
A 5 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5
1 0.696917808 0.029965753 0.033390411 0.18150685 0.0582191781
2 0.016382700 0.855176933 0.022935780 0.08584535 0.0196592398
3 0.006451613 0.008602151 0.572043011 0.39354839 0.0193548387
4 0.006078577 0.003113417 0.002186805 0.98773165 0.0008895478
5 0.025286909 0.016339234 0.010114764 0.04182066 0.9064384361
```

```

# Markov avec segmentation Initiat = Client

> Markov_Segment_InitClient_Amdf5 <- markovchainFit(data = subset(Data_Markov_Seg_
  Amdf5[Data_Markov_Seg_Amdf5$Initiat == "Client",], select = -c(ID_FILE, PnP,
  UWY, Ptf, Initiat, Ccy, Period0) ), method = "mle", name = "Markov Segmentation
  Initia = Client")
> Markov_Segment_InitClient_Amdf5$estimate
Markov Segmentation Initia = Client
A 5 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5
1 0.730769231 0.000000000 0.076923077 0.15384615 0.03846154
2 0.037037037 0.740740741 0.000000000 0.18518519 0.03703704
3 0.047619048 0.000000000 0.619047619 0.33333333 0.00000000
4 0.004694836 0.011737089 0.004694836 0.97887324 0.00000000
5 0.018181818 0.009090909 0.013636364 0.03636364 0.92272727

# Markov avec segmentation Ccy = EUR

> Markov_Segment_CcyEUR_Amdf5 <- markovchainFit(data = subset(Data_Markov_Seg_Amdf5
  [Data_Markov_Seg_Amdf5$Ccy == "EUR",], select = -c(ID_FILE, PnP, UWY, Ptf,
  Initiat, Ccy, Period0) ), method = "mle", name = "Markov Segmentation Ccy = EUR
  ")
> Markov_Segment_CcyEUR_Amdf5$estimate
Markov Segmentation Ccy = EUR
A 5 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5
1 0.700280112 0.025210084 0.023809524 0.19747899 0.0532212885
2 0.028790787 0.750479846 0.032629559 0.15163148 0.0364683301
3 0.008196721 0.004098361 0.540983607 0.43032787 0.0163934426
4 0.004612454 0.001704602 0.001955279 0.99127645 0.0004512183
5 0.027444254 0.013036021 0.009948542 0.05180103 0.8977701544

# Markov avec segmentation Period0 = Avant06

> Markov_Segment_Avant06_Amdf5 <- markovchainFit(data = subset(Data_Markov_Seg_Amdf
  5[Data_Markov_Seg_Amdf5$Period0 == "Avant06",], select = -c(ID_FILE, PnP, UWY,
  Ptf, Initiat, Ccy, Period0) ), method = "mle", name = "Markov Segmentation
  Period0 = Avant06")
> Markov_Segment_Avant06_Amdf5$estimate
Markov Segmentation Period0 = Avant06
A 5 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5
1 0.825750242 0.020329138 0.0125847047 0.12003872 0.0212971926
2 0.007839721 0.907665505 0.0165505226 0.06010453 0.0078397213
3 0.003134796 0.000000000 0.7774294671 0.21943574 0.0000000000
4 0.003222585 0.001452433 0.0003631082 0.99405410 0.0009077705
5 0.034351145 0.015267176 0.0076335878 0.06202290 0.8807251908

```

```
# Markov avec segmentation Prop et UWY70

> Markov_Segment_PropUWY70_Amdf5 <- markovchainFit(data = subset(Data_Markov_Seg_
  Amdf5[Data_Markov_Seg_Amdf5$PnP == "P" & Data_Markov_Seg_Amdf5$UWY == "UWY70"
  ,], select = -c(ID_FILE, PnP, UWY, Ptf, Initiat, Ccy, Period0) ), method = "mle
  ", name = "Markov Segmentation Prop & UWY70")
> Markov_Segment_PropUWY70_Amdf5$estimate
Markov Segmentation Prop & UWY70
A 5 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5
1 0.678571429 0.035714286 0.0178571429 0.25000000 0.0178571429
2 0.022727273 0.750000000 0.0454545455 0.18181818 0.0000000000
3 0.000000000 0.000000000 0.6666666667 0.33333333 0.0000000000
4 0.007371007 0.003071253 0.0006142506 0.98832924 0.0006142506
5 0.011904762 0.005952381 0.000000000 0.05952381 0.9226190476
```

Listing B.11 – Matrices de transition markovienne avec segmentation

# Bibliographie

- [1] Markov A.A. *Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain*. The Notes of the Imperial Academy of Sciences of St. Petersburg, VIII Series, Physio-Mathematical College XXII, No. 9, 1907.
- [2] Dutang C. *Cours de Modélisation Stochastique (M1 - Actuariat)*. Notes basées sur les cours de Nicolas Klutchnikoff, d'Olivier François et Hervé Guiol, 2012-2013.
- [3] Hosmer D.W. and Lemeshow S. *Applied Logistic Regression*. Wiley Series in Probability and Mathematical Statistics, 2000.
- [4] Bardos M. *Analyse Discriminante - Application au risque et scoring financier*. Dunod, 2001.
- [5] El Karoui N. and Benaïm M. *Promenade Aleatoire - Chaînes de Markov et simulations ; martingales et strategie*. École Polytechnique, 2005.
- [6] Désesquelles P. *Les processus de Markov ; en biologie, sociologie, géologie, chimie, physique et applications industrielles*. Ellipses, Paris, 2016.
- [7] Tufféry S. *Data Mining et statistique décisionnelle, L'intelligence des données*. Éditions Technip, 2010.