

Rapport de stage de 5^{ème} année cycle ingénieur
Présenté par

LAPORTE ETIENNE

Filière MAM
Promotion 19
Année 2012-2013

Ingénieur quantitatif

Crédit Foncier de France



CRÉDIT FONCIER

Tuteur de l'entreprise : GAROUI Djaffar

Tuteur de l'école : OUDIN DARDUN Fabienne

Mars-août 2013

POLYTECH Lyon - UNIVERSITE CLAUDE BERNARD LYON 1
Domaine Scientifique de La Doua – 15, Boulevard Latarjet
69622 VILLEURBANNE CEDEX
Tél. (33) 04.72.43.12.24 - Fax. (33) 04.72.43.12.25
<http://polytech.univ-lyon1.fr/>

Remerciements

Je tiens à profiter de ce rapport de stage pour exprimer mes plus vifs remerciements à toutes les personnes qui m'ont aidé et qui ont contribué à rendre cette expérience intéressante et enrichissante.

Tout d'abord, je suis reconnaissant envers M. Djaffar GAROUI, responsable de la Direction Etudes et Projets de l'Asset Liability Management, qui m'a permis d'intégrer le CREDIT FONCIER. Il a été mon tuteur d'entreprise, je le remercie vivement.

Enfin, mes remerciements vont à toutes les personnes du service qui ont bien voulu répondre avec la meilleure attention à mes nombreuses interrogations pour me permettre d'avancer dans la réalisation de ce projet.

Liste des abréviations

- Asset Liability Management : ALM
- Crédit Foncier de France : CFF
- CRD : Capital Restant dû
- Matrest : Maturité résiduelle du prêt
- Ectaux : Ecart de taux entre celui du prêt et celui moyen du marché
- Tmm : Taux moyen du marché
- Mtpret : Montant du prêt initial
- RA : Remboursement Anticipé
- TV : Taux variables
- TF : Taux fixe
- DA : Date d'arrêté

Sommaire

I Présentation de l'entreprise	5
1.1) Présentation générale	5
1.2) Histoire du Crédit Foncier de France	5
1.3) Les métiers du Crédit Foncier de France	6
1.3.1) Financement spécialisé des particuliers	7
INTRODUCTION.....	8
II ATTENTES ET DEVELOPPMENT	9
2.1 Attentes.....	9
2.2 Développements	11
2.2.1 Prise en main du modèle	11
2.2.2 La recherche des données et mise en forme de la base de données	11
2.2.3 Interfaçage du modèle	14
2.2.4 Exemple de code R par VBA pour le tirage	15
2.2.5 Exemple de génération de script pour le backtesting	17
III. Discussions des résultats	19
3.1 Interprétation des résultats	19
3.1.1 Critère de choix du modèle statistique	19
3.1.2 Comment interpréter un taux de RA ?	21
3.1 Perspective d'évolution	25
Conclusion.....	26
Références et Sources.....	27
Annexes	28
Résumé :	29

I Présentation de l'entreprise

1.1) Présentation générale

Le Crédit Foncier de France est le leader spécialisé du projet immobilier et patrimonial. Il fait bénéficier ses clients d'une capacité d'innovation unique et d'une connaissance particulièrement fine du marché immobilier.

Avec près de 4 000 collaborateurs et un réseau commercial couvrant l'ensemble du territoire, il propose à ses clients particuliers et professionnels des solutions sur mesure de financement, de services et de produits d'épargne, dans le cadre d'une approche patrimoniale.

Il accompagne ses clients dans la réalisation de leur projet en mettant à leur service ses activités d'expertise et de conseil, sa technicité dans des domaines tels que la transaction, la gestion, la valorisation d'actifs immobiliers.

Le Crédit Foncier de France, filiale du Groupe Caisse d'Epargne, un des tous premiers groupes bancaires français, s'inscrit dans une dynamique de développement avec l'intégration d'Entenial et d'A3C.

Le Crédit Foncier de France est :

- Leader du projet immobilier
- Leader du crédit immobilier parmi les établissements spécialisés
- Leader du financement de l'investissement locatif
- Leader de l'expertise immobilière avec Foncier Expertise qui réalise et publie des études de référence sur le marché immobilier.
- Premier émetteur privé après l'Etat et premier émetteur d'obligations foncières avec la Compagnie de Financement Foncier.

1.2) Histoire du Crédit Foncier de France

Le Crédit Foncier de France a été créé en 1852 à l'initiative des pouvoirs publics, Il s'agissait alors d'organiser le crédit hypothécaire en France à l'image des banques hypothécaires allemandes instituées à la fin du XVIII siècle.

Dès l'origine, le statut de la société de crédit foncier de l'établissement s'est caractérisé par un mécanisme de prêts hypothécaires amortissables à long terme adossés à un financement obligatoire.

➤ DE 1852 à 1945

Le Crédit Foncier est créé en 1852 pour financer la modernisation de l'agriculture en octroyant des prêts à long terme (50 ans) remboursables par annuités de 5 %. Face à la faible demande des agriculteurs, il s'oriente alors vers le financement des grands travaux immobiliers urbains et étrangers. C'est l'époque du Baron Haussmann et de Napoléon III. La banque finance très largement les grandes percées du Baron Haussmann à Paris puis les écoles de Jules Ferry. Le Crédit Foncier participe particulièrement à l'aménagement de la rue

de Rivoli, du grand hôtel du Louvre, du quartier de l'Opéra et de la plaine Monceau. Après des opérations malheureuses à l'étranger, notamment en Egypte, et la crise de l'immobilier parisien, il investit en province vers 1880 puis dans la construction de pavillons en région parisienne. La période de l'entre deux guerres est marquée par les crises financières.

➤ Depuis la libération

Après la seconde guerre mondiale, le Crédit Foncier participe activement au boom de la construction de logements bon marché. Il est notamment chargé de la gestion du Fonds National d'Amélioration de l'Habitat créé en 1945. La loi du 21 juillet 1950 missionne la banque pour accorder des prêts spéciaux aux particuliers bénéficiant de primes à la construction. Enfin, en 1972, l'Etat lui délègue le soin d'instruire et de gérer les dossiers de prêts immobiliers conventionnés. Aujourd'hui encore, les prêts aux communes et au logement constituent toujours les principaux secteurs d'activité du Crédit Foncier. Depuis la guerre, il a permis à 15 millions de Français d'acquérir leur logement. Son rôle était clair : aider l'Etat à financer le logement en distribuant des prêts bonifiés aux classes moyennes.

➤ Depuis 1999

La loi du 25 juin 1999 a profondément réformé le régime des sociétés de crédit foncier institué en 1852 : le Crédit Foncier a adopté le régime des sociétés à directoire et conseil de surveillance et a créé une filiale de refinancement, la Compagnie de Financement Foncier, première société de crédit foncier au regard de la loi. La seconde étape essentielle a été l'adossment au groupe des Caisses d'Epargne devenu actionnaire majoritaire. Cet adossment a permis au Crédit Foncier, par ailleurs profondément restructuré, de rejoindre un groupe bancaire important, dynamique et de taille européenne, au sein duquel il a l'ambition d'exploiter au mieux ses atouts.

➤ Le 1^{er} juin 2005, Entenial fusionne avec le Crédit Foncier

Entenial est né en 2000 de la fusion de deux établissements complémentaires : le Comptoir des Entrepreneurs, créé en 1848 afin de relancer la construction immobilière, et la Banque La Hénin née en 1852. La mise en commun et la complémentarité des savoir-faire de ces deux entreprises ont permis à la banque Entenial d'acquérir une expertise complète dans les métiers du crédit et de l'épargne, des services bancaires aux professionnels de l'immobilier, pour accompagner ses clients, particuliers comme entreprises, dans tous leurs projets immobiliers Et patrimoniaux. Aujourd'hui, avec l'acquisition d'Entenial, le Crédit Foncier a vocation à construire le numéro 1 français des financements et services immobiliers au sein du Groupe Caisse d'Epargne.

1.3) Les métiers du Crédit Foncier de France

Le Crédit Foncier couvre tous les secteurs de l'immobilier grâce à l'exercice de quatre métiers principaux :

- Le financement spécialisé des particuliers ;
- Le financement des entreprises et des investisseurs
- Les services immobiliers ;
- Le refinancement sécurisé

1.3.1) Financement spécialisé des particuliers

Une politique particulièrement innovante en matière de crédits, de services, de placements et de produits immobiliers packagés.

Le Crédit Foncier apporte une solution à la fois globale et sur-mesure à chaque besoin de financement immobilier et d'investissement patrimonial. Il propose des solutions de crédits innovantes dans le neuf et l'ancien, ainsi que dans l'investissement locatif. Il offre un choix diversifié de produits et services :

- Prêts à taux fixe, prêts à taux révisable, prêts in fine ou amortissables et prêts spécifiques,
- Prêts réglementés : prêts à 0%, prêts à l'accession sociale - PAS, prêts locatifs sociaux - PLS, prêts locatifs intermédiaires - PLI,
- Assurances liées au crédit (décès invalidité, perte d'emploi, garantie revente) et au bien financé (construction, multirisque habitation),
- Produits financiers (assurance-vie, PERP, OPCVM...).

1.3.2) Financement des entreprises et investisseurs

Une offre qui répond à l'ensemble des besoins des clients : entreprises, investisseurs, fonds d'investissement et promoteurs immobiliers.

Le Crédit Foncier développe une gamme complète de produits et services dédiés à l'immobilier professionnel : financement de la promotion immobilière, financements structurés ou non des actifs immobiliers des entreprises, investisseurs et opérateurs sociaux, financement de projets, services bancaires aux administrateurs de biens et agents immobiliers. Il figure parmi les entreprises leader sur ce marché, Encours de 13.3 Milliards d'euros au 31.12.204.

1.3.3) Services Immobiliers :

Un pôle qui fédère des savoir-faire placés au cœur de l'activité immobilière Le pôle "Services Immobiliers", en plein développement, propose directement ou par l'intermédiaire de ses filiales, une offre structurée dans quatre domaines d'intervention : Expertise des biens et analyse des marchés immobiliers, Commercialisation d'immobilier résidentiel, Conseil et valorisation d'actifs, Commercialisation d'immobilier d'entreprise, Gestion locative institutionnelle. Il gère en outre les participations immobilières du Crédit Foncier.

1.3.4) Refinancement sécurisé :

La Compagnie de Financement Foncier et Vauban Mobilisations Garanties, deux filiales du Crédit Foncier de France, assurent le refinancement du Groupe à des conditions de ressources particulièrement avantageuses. Pour la seconde année consécutive, la Compagnie de Financement Foncier est le premier émetteur après l'Etat et premier émetteur d'obligations foncières. 11,2 milliards d'euros d'émission au 31/12/2003, dont 9,6 milliards d'euros d'émission d'obligations foncières pour la Compagnie de Financement Foncier.

INTRODUCTION

Dans le cadre de mes études en Mathématiques Appliquées et Modélisation j'ai été amené à effectuer un stage de fin d'étude de six mois.

J'ai choisi de postuler dans le monde bancaire car ce sont des entreprises qui m'intéressent pour mon futur parcours professionnel.

Ma candidature au groupe BPCE a été transmise dans leur filiale , le CREDIT FONCIER . L'offre me correspondait car elle était dans la continuité de mon projet professionnel et me permettait d'utiliser les compétences que j'avais précédemment acquises.

Lors de mes précédents stages, en tant que développeur Excel\VBA dans un premier temps aux aciéries AUBERT&DUVAL puis à la BANQUE de FRANCE (succursale de Lyon) j'ai eu le temps de bien maîtriser l'outil Excel. J'ai pu ainsi élaborer des applicatifs dans leurs globalités aidant par exemple le Pôle Groupe de la banque de France à l'analyse financière. J'ai ainsi utilisé une combinaison de codes VBA et SQL de manière à récupérer les données sur les disques communs de la Banque de France.

Mon expérience au service CREDIT FONCIER, m'a permis de prendre en main le périmètre de l'ALM, des dimensions d'analyse et des indicateurs financiers d'une base de données. J'ai pu également optimiser, réaliser et sécuriser un modèle de remboursements anticipés des prêts aux particuliers en implémentant le modèle sous R puis sous Rexcel pour une automatisation et interfaçage du modèle.

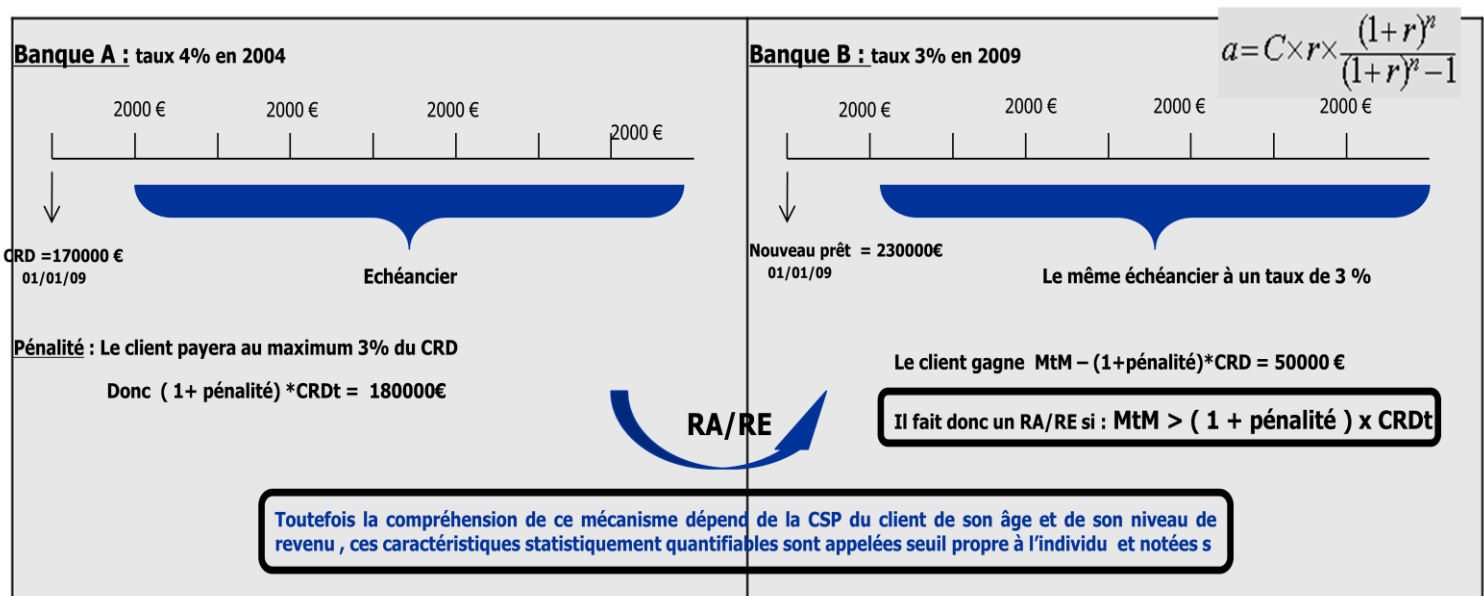
II ATTENTES ET DEVELOPPMENT

2.1 Attentes

Dans un contexte de crise financière, la Direction Asset Liability Management (ALM - Gestion de Bilan) joue un rôle clef en matière de maîtrise des risques financiers. L'ALM (qui fait partie du pôle finance voir Annexe 1 : Organigramme) a pour missions principales de gérer le risque de taux d'intérêt global (sensibilité du résultat de l'établissement aux variations de taux) ainsi que le risque de liquidité à moyen-long terme.

La mission principale de ce stage est de calibrer le modèle actuel de remboursement anticipé des prêts.

Exemple : En 2004 un client a contracté un prêt de 300 000 euros à annuités constantes auprès d'une banque A à un taux fixe de 4% de maturité de 15 ans pour l'achat d'une maison. En 2009 un prêt de maturité 10 ans peut être contracté sur le marché à un taux fixe de 3%, le même client s'approche de la banque B et souhaite rembourser par anticipation en procédant à un rachat de crédit par cette dernière.



La rationalité des clients dépend donc de leur sensibilité aux fluctuations des taux d'intérêt. Cette hypothèse suppose que les RA dépendent du taux d'intérêt et que le client exerce un RA si la valeur actuelle de son prêt au taux du marché est suffisamment supérieure au capital restant dû. Ce qui permet d'écrire à un instant t :

$$\text{RA si } \text{MtM} > (1 + \text{pénalité}) \times \text{CRDt} + s.$$

MtM : valeur du marché du prêt.

CRDt : capital restant dû.

s : coût d'opportunité (seuil de gain).

Le modèle actuel qui modélise ce phénomène est un modèle logistique :
 Soit π la probabilité d'avoir un remboursement anticipé (RA) en fonction des variables explicatives.

$$\pi = P(Y = RA/X)$$

Soit X_1, \dots, X_J les variables explicatives et a_0, \dots, a_J leurs coefficients.

$$\ln \left[\frac{\pi(\omega)}{1 - \pi(\omega)} \right] = a_0 + a_1 X_1 + \dots + a_J X_J$$

Où $\frac{\pi}{1 - \pi} = \frac{P(Y = RA/X)}{P(Y = nonRA/X)}$ exprime un rapport de chance.

On pose $C(X) = a_0 + a_1 X_1 + \dots + a_J X_J$

On obtient alors un score π d'avoir un remboursement anticipé

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

Le modèle actuel est expliqué par deux variables :

$RA \sim tmm + ec_taux$, car actuellement les résultats de la régression logistique sont exploités par un logiciel de sortie (Bancware) qui ne permet d'avoir que deux variables explicatives en entrée.

Les deux variables sont décomposées en variable catégorielles que l'on explicitera ultérieurement.

Ce modèle a été calibré sur des données de décembre 2007 à décembre 2010 sous Matlab.
 Le Crédit Foncier souhaite calibrer leurs modèles avec les données plus récentes et explorer la possibilité d'avoir un modèle plus complet.

2.2 Développements

2.2.1 Prise en main du modèle

La calibration de l'ancien modèle a été réalisée sous Matlab. Ma première mission a donc été de me plonger dans le code pour comprendre chaque étape du projet.

Après avoir compris le mécanisme du code et en attendant la nouvelle base de donnée, j'ai commencé à réaliser le même code sous R qui est un logiciel gratuit et plus adapté que Matlab aux traitements statistiques.

Par la suite j'ai automatisé la procédure en interfaçant R avec une interface VBA, ce qui permet d'avoir une architecture propre et de visualiser chaque étape du traitement sur les tableurs Excel.

2.2.2 La recherche des données et mise en forme de la base de données

Cette partie est une des parties les plus importantes dans un projet : récupérer les données. Il m'a fallu beaucoup de temps pour les récupérer et les croiser avec les différents services afin de valider la base de données pour la régression.

La base actuelle qui a servi à la calibration de 2007 à 2010 est la suivante :

Idcr00	Identifiant de crédit
Matrest	maturité résiduelle en années
Age	âge du prêt en années
Durée	durée initiale du prêt en années
Montant	montant initial du prêt en euros
tmm	un taux trimestriel publié par la banque de France, qui résulte de la moyenne arithmétique simple des TEG des prêts immobiliers à taux fixe et à taux révisable observés sur le marché.
Ecart taux	L'écart entre le taux du prêt et le taux moyen du marché
CRD	capital restant dû en euros
Mensu	mensualité du prêt en euros
RA	1 si l'option de RA est exercée, 0 sinon

On ne trouve pas actuellement cette base propre dans les fichiers du CFF, cette base est une sous base calculée à partir d'un fichier appelé « radio ».

Voici le fichier radio avec ces données :

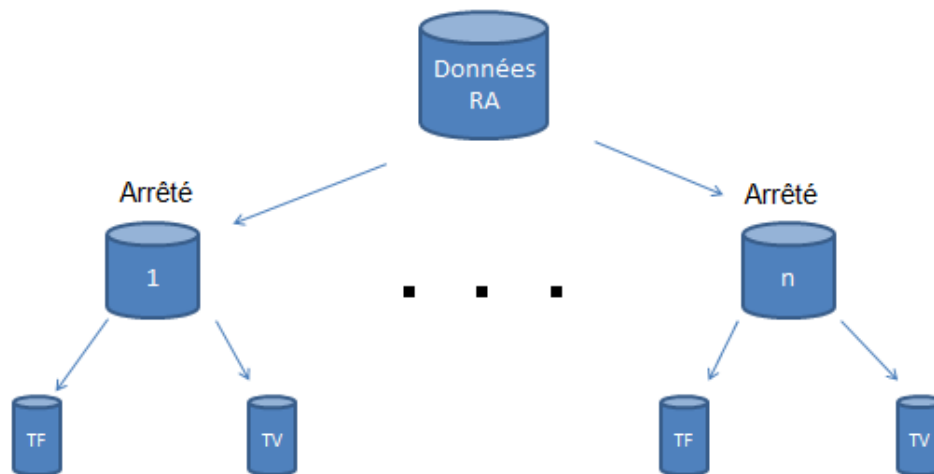
	Type de variable	Variable	Libellé
A l'octroi	Identifiant	nu	Identifiant du crédit
		oper	Identifiant de l'opération (regroupement de crédit)
	Caractéristiques produit	macro1	Macro secteur (macro à vide représentant le particulier)
		catstat1	Catégorie statistique (type de produit)
		dest	Destination du bien (0: accession à la propriété;1: locatif)
	Caractéristiques de l'opération	andp	Année du point de départ informatique (début de l'amortissement)
		anmef	Année de mise en force
		anexth	Année d'extinction du prêt
		duree	Durée du prêt
		mtaut	Montant autorisé du prêt
		quot	Quotité de financement (somme des prêts/coût de l'opération)
	Caractéristiques client	csp1	CSP de l'emprunteur principal
	Caractéristiques de la garantie	depgage	Département du gage
		famg	Famille de garantie
		gage	Valeur du gage
A chaque trimestre	Caractéristiques de gestion (*)	xcrd(i)	Capital restant dû (au trimestre i)
		xsaj(i)	Solde à jour (compte de l'arriéré) (au trimestre i)
		xverst(i)	Montant des versements du client (au trimestre i)
		xech(i)	Montant de l'échéance (au trimestre i)
		xitx(i)	Taux d'intérêt du prêt (au trimestre i)

Calcul et source des données :

Idcr00	idcr00
Matrest	anexth - i
Age	i - anmef
Durée	duree/12
Montant	mtaut
tmm	taux moyen du marché BDF : http://www.banque-france.fr/fileadmin/statistiques/fr/base/html/tmf_mens_france_fr_taux_usure.html
Ecart taux	xitx - tmm
CRD	xcrd
Mensu	xech
RA	$xcrd=0 \text{ et } xsaj>0 \text{ OU } xverst>0.10*mtaut$

Une fois que la base a été créée, il a fallu scinder cette base en différentes images du stock pour chaque dates d'arrêtées. On parle d'images de stock, car pour chaque dates d'arrêtées (Jan-Mars Mars-Juin Juin-Sept Sept-Déc) nous avons une image du prêt avec ses caractéristiques.

Par la suite nous faisons la différence entre les prêts à taux fixe (TF) et les prêts à taux variables (TV).

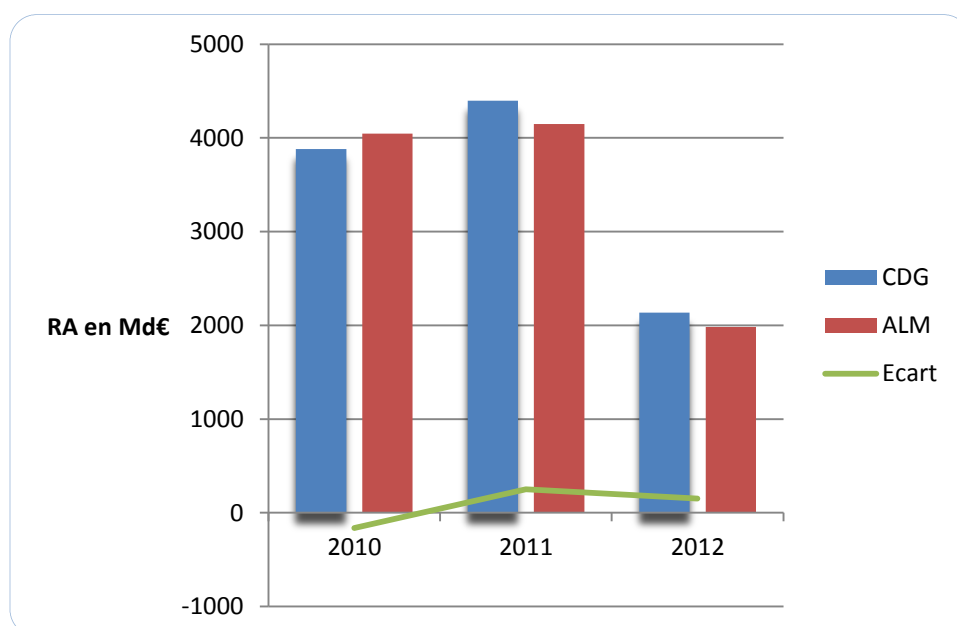


Des procédures stockées ont été implémentées sous SQL pour obtenir ces bases.

Une fois que nous avons obtenu ces bases, nous les avons croisées avec des données du service de contrôle de gestion.

Nous avons observé beaucoup de dissimilitudes au niveau des CRD en cours entre les deux bases. En effet la base que j'ai construite identifie les RA sur une variation de stock, donc si la règle de gestion n'est pas parfaite pour les identifier la base est erronée.

Après réflexion nous avons préféré partir sur des bases qui sont propres à ALM qui identifient les RA et observé et par jointure permet d'obtenir les données correspondantes aux prêts à une DA donnée. Nous avons ensuite comparé les bases ALM et CDG.



On constate sur les derniers trimestres 2012 un écart moyen de 40 Million € (6% du stock) qui a été expliqué en partie par les RA partielle, non modélisable coté ALM mais que globalement les bases sont identiques.

2.2.3 Interfaçage du modèle

L'interface du modèle de remboursement anticipé est la suivante :



Cette interface permet de piloter R à partir d'un classeur Excel et du code VBA.

On peut notamment :

- Charger un jeu de données (respectant le format de base, et les noms de colonnes) .
- Réaliser un tirage sur le jeu de données.
- Obtenir un résumé du jeu de données (Moyenne, Min, Max, quartiles, corrélation) de chaque variables.
- Réaliser une sélection de modèle automatique (méthode stepwise).
- Choisir son propre modèle avec des variables quantitatives ou qualitatives.
- Choisir les classes de décomposition de la variable.
- Lancer la régression avec le modèle choisis.
- Lancer le backtest du modèle choisi sur des données qui n'ont pas servi à la modélisation.

Pré-Conditionnement :

- Réaliser la séparation 80% - 20% pour l'échantillon de modélisation et l'échantillon de backtesting :
 - Pour une DA, on réalise un tirage uniforme de 80% du jeu de données.
 - Ensuite l'ensemble des fichiers de 80% des données d'une DA sont alors fusionnés pour avoir le jeu de donnée en entrée de l'application RExcel
 - Le reste des 20% est utilisé pour le backtesting et chaque fichiers doivent être placés dans le dossier Backtest\TF ou Backtest\TV suivant la nature du taux.
- La mise en forme du fichier avec les en têtes doit être la même et dans les mêmes positions.

Soit :

Idcr00	Matrest	âge	durée	ectaux	tmm	CRD	Mtprêt	Mensualité	RA
12751199	24.75	0.25	25	0.83	4.57	187594	210000	1090	0
12751399	9.92	5.08	15	-0.17	4.72	99368	120000	921	0
12751399	11.92	3.08	15	-0.89	5.44	111832	120000	921	0
12753299	16.25	3.75	20	0.64	5.26	148610	170000	1179	0
12758299	2.92	4.08	7	-2.87	5.87	5926	13400	177	0
12758299	3.17	3.83	7	-2.85	5.85	6411	13400	177	0

2.2.4 Exemple de code R par VBA pour le tirage

Dans cette section, je vais développer les codes qui sont sous les boutons poussoirs :

- Tirage
- Tirage Strat

Pour résumer le code est le suivant :

```
Dim chemin_jeu As String
Dim texte_a_envoyer As String
Dim NomRep As String

'Récupère le chemin du classeur où est le code
NomRep = ThisWorkbook.Path

'Permet de lancer R
RInterface.StartRServer
RInterface.RRun "rm(list = ls())"

'Récupère le chemin de la textbox et le met au format de lecture pour R
chemin_jeu = Replace(textbox_jeux_de_donnees.Value, "\", "/")

'Mise à forme de la commande à envoyer
txt_jeu = "X1=read.table(" & """" & chemin_jeu & """" & ",header=TRUE, sep=""\t"", na.strings=""NA"", dec="".", strip.white=TRUE)"

'Affiche les commandes générées dans R
Debug.Print "La commande pour le jeu de données est :"
Debug.Print txt_jeu

'Lancer la commande dans R
RInterface.RRun txt_jeu

'Choix 1 : Tirage classique
texte_a_envoyer = NomRep & "\tirage.r"
RInterface.RunRFile (texte_a_envoyer)
MsgBox "Tirage OK"

'Choix 2 : Tirage stratifié
texte_a_envoyer = NomRep & "\tirage_strat.r"
RInterface.RunRFile (texte_a_envoyer)
MsgBox "Tirage OK"
```

Tout l'art de RExcel et de pouvoir stocker dans une variable texte les commandes qui vont être lancées dans R.

En effet ici, les variables auront pour commande :

- `txt_jeu :`
`X1=read.table("C:/Users/U33012/Desktop/Test RA/Test/RA_2013/Stock ALM/TDB/Jeux de données/Stock_RA_TF_Model_avec_nom.txt",header=TRUE, sep="\t", na.strings="NA", dec=".", strip.white=TRUE)`
- `texte_a_envoyer :` `"C:\Users\U33012\Desktop\Test RA\Test\RA_2013\Stock ALM\TDB\tirage.r"`

Et le code R pour le tirage.r :

```

1  #Librairies
2  library(tis)
3  library(MASS)
4  require(MASS)
5
6  #On récupère la partie des RA
7  X_RA=subset(X1, X1[, "RA"]==1)
8
9  #On récupère la partie des non RA
10 t=tapply(X1[, "RA"], X1[, "Idcr00"], max)
11 resultat=as.numeric(names(t[t==0]))
12 X_nonRA=X1[X1$Idcr00 %in% resultat,]
13
14 #Tirage des dossiers à identifiant Idcr00 unique
15 coll=X_nonRA[, "Idcr00"]
16 t=tapply(1:length(coll), coll,
17
18     echant<-function(x)
19     {
20         if(length(x)==1){return(x)}
21         return(sample(x, 1))
22     }
23 )
24
25 #On récupère tout les dossiers non RA à identifiants uniques
26 tirage_nonRA=X_nonRA[t,]
27
28
29 #Mise en forme du jeu de données non RA + RA
30 tirage=rbind(tirage_nonRA, X_RA)
31 data_modelisation=tirage
32
33 #Transforme la variable RA à expliquer en facteur : 0 ou 1
34 data_modelisation$RA=as.factor(data_modelisation$RA)
35

```

Et pour tirage_start.r :

C'est le même code sauf que l'on remplace les lignes de 28 à 34 par :

```

#Réalise un tirage parmi les dossiers non RA unique de x valeurs de dossiers. x étant la nombre de RA observées dans l'échantillon
tirage_autant_RA_non_RA=tirage_nonRA[sample(1:nrow(tirage_nonRA), nrow(X_RA)),]
tirage=rbind(tirage_autant_RA_non_RA, X_RA)
data_modelisation=tirage
data_modelisation$RA=as.factor(data_modelisation$RA)

```

Cette ligne rajoutée ici permet de faire un tirage de x lignes (ici le nombre de RA observée dans le jeux de données) parmi les dossiers non RA à identifiant unique (en effet un dossier

étant suivi sur plusieurs dates d'arrêtees, il y a donc dans le jeu de données plusieurs fois ce numéro de dossier pour chaque nouvelle DA).

2.2.5 Exemple de génération de script pour le backtesting

L'une des parties qui m'a demandé le plus d'astuces est celle-là.

En effet pour le backtesting du modèle choisi par l'utilisateur, je me demandai :

« Comment lancer un script R dont les données dépendaient de l'utilisateur ? »

En effet, quand j'ai repris le code en première approche :

```
1  taille=12
2  liste_data=vector(mode = "list", length = taille)
3  nb_arrete=10
4  for (n_i in 1:taille)
5  {
6    chemin_actuelle=paste(chemin_backtest,"Back_test_RA_TF_",n_i,".txt",sep="")
7    #Lecture et Backtest sur chaque mois
8    data=read.table(chemin_actuelle,header=FALSE, sep="\t", na.strings="NA", dec=".", strip.white=TRUE)
9
10   #Renome les colonnes
11   names(data)[c(1)] <- c("Idrc00")
12   names(data)[c(2)] <- c("Matrest")
13   names(data)[c(3)] <- c("âge")
14   names(data)[c(4)] <- c("durée")
15   names(data)[c(5)] <- c("Mtprêt")
16   names(data)[c(6)] <- c("Mensualité")
17   names(data)[c(7)] <- c("ectaux")
18   names(data)[c(8)] <- c("tmm")
19   names(data)[c(9)] <- c("CRD")
20   names(data)[c(10)] <- c("RA")
21   data$RA=as.factor(data$RA)
22
23   #Modalité ectaux découpées en classe, 1 si dans intervalle 0 sinon
24   data$sectaux_qual <- cut(data$sectaux,breaks=c(-Inf,0.05,1.23,Inf),include.lowest=TRUE,right=FALSE)
25   #Modalité tmm découpées en classe, 1 si dans intervalle 0 sinon
26   data$tmm_qual <- cut(data$tmm,breaks=c(-Inf,4.72,5.26,5.44,Inf),include.lowest=TRUE,right=FALSE)
27
28   #Enregistre des valeurs de chaque mois
29   new_valeurs=Backtest(modele,data,n_i,chemin_sauvergarde,nb_arrete)
30   resultat=rbind(resultat,new_valeurs)
31
32   #Supprime l'objet pour le réutiliser
33   rm(data)
34 }
35
```

En effet les données « *breaks=c()* » sont figées dans le code, et alors non exploitables pour une utilisation dynamique où les données du *breaks* seront définies au départ par l'utilisateur.

Ces deux lignes permettent ici de calculer les nouvelles variables qualitatives du modèles pour les nouvelles données qui n'ont pas été utilisés pour la modélisation.

L'astuce à laquelle j'ai songé ici est en fait de créer ce fichier à chaque fois que l'application se lance avec les caractéristiques choisies.

```

Sub Mise_en_forme_fichier_backtesting()

'Ajouter la Référence Microsoft Scripting Runtime
Dim oFSO As Scripting.FileSystemObject
Dim oFl As Scripting.File
Dim oTxt As Scripting.TextStream
Dim i As Integer

'Instanciation du FSO
Set oFSO = New Scripting.FileSystemObject
Set oFl = oFSO.GetFile("C:\Users\U33012\Desktop\Test RA\Test\RA_2013\Stock ALM\TDB\backtesting.r")
Set oTxt = oFl.OpenAsTextStream(ForWriting)

oTxt.WriteLine "liste_data=vector(mode = ""list"", length = taille)"
oTxt.WriteLine "source("""C:/Users/U33012/Desktop/Test RA/Test/RA_2013/Stock ALM/TDB/Backtest.r"")"
oTxt.WriteLine "resultat=0"
oTxt.WriteLine "for (n_i in 1:taille)"
oTxt.WriteLine "{"
oTxt.WriteLine "chemin_actuelle=paste(chemin_backtest, ""Arrete_TF_Back_"", n_i, ""_.txt"", sep = """)"
oTxt.WriteLine "#Lecture et Backtest sur chaque mois"
oTxt.WriteLine "data=read.table(chemin_actuelle,header=TRUE, sep=""\t"", na.strings=""NA"", dec="".", strip.white=TRUE)"

oTxt.WriteLine "#Renome les colonnes"
oTxt.WriteLine "names(data)[c(1)] <- c("""Idrc00"")"
oTxt.WriteLine "names(data)[c(2)] <- c("""Matrest"")"
oTxt.WriteLine "names(data)[c(3)] <- c("""âge"")"
oTxt.WriteLine "names(data)[c(4)] <- c("""durée"")"
oTxt.WriteLine "names(data)[c(5)] <- c("""Mtpret"")"
oTxt.WriteLine "names(data)[c(6)] <- c("""Mensualité"")"
oTxt.WriteLine "names(data)[c(7)] <- c("""ectaux"")"
oTxt.WriteLine "names(data)[c(8)] <- c("""tmm"")"
oTxt.WriteLine "names(data)[c(9)] <- c("""CRD"")"
oTxt.WriteLine "names(data)[c(10)] <- c("""RA"")"
oTxt.WriteLine "data$RA=as.factor(data$RA)"
oTxt.WriteLine ""

oTxt.WriteLine "#Calculer les nouvelles variables du modèle"
oTxt.WriteLine txt_ec_backtest
oTxt.WriteLine txt_tmm_backtest
oTxt.WriteLine txt_Mtpret_backtest
oTxt.WriteLine txt_âge_backtest
oTxt.WriteLine txt_Matrest_backtest

oTxt.WriteLine txt_durée_backtest
oTxt.WriteLine txt_CRD_backtest
oTxt.WriteLine txt_Mensualité_backtest

oTxt.WriteLine "#Enregistre des valeurs de chaque mois"
oTxt.WriteLine "new_valeurs = Backtest(modele, data, n_i, chemin_sauvergarde , " & nb_fichier_backtest & ")"
oTxt.WriteLine "resultat = rbind(resultat, new_valeurs$resultat)"

oTxt.WriteLine "#Supprime l'objet pour le réutiliser"
oTxt.WriteLine "echantillon=new_valeurs$donnees"
oTxt.WriteLine "rm(data)"
oTxt.WriteLine "}"

oTxt.Close

End Sub

```

Parties « durs » du code, indépendantes des caractéristiques du modèle

Parties dynamique du code, les variables de type txt sont remplis en amont du code suivant les caractéristiques du modèle

Exemple de comment la variable txt_ec_backtest est incrémenté :

```

If ectaux_qual.Value = True Then
txt_ec = "data_modelisation$" & ectaux_qual.Name & "<-cut(data_modelisation$ectaux," & val_ec_qual & ",include.lowest=TRUE,right=FALSE)"
txt_ec_backtest = "data$" & ectaux_qual.Name & "<-cut(data$ectaux," & val_ec_qual & ",include.lowest=TRUE,right=FALSE)"
Debug.Print txt_ec
RInterface.RRun txt_ec
End If

```

Où *ectaux_qual.Name* est le nom de la textbox ici donc « ectaux », et les valeurs du *breaks* sont récupérées par la variable *val_ec_qual* de la manière suivante :

```

If ectaux_qual.Value = True Then
nb = ectaux_qual_val.Value
val_ec_qual = ""
For i = 1 To (nb + 1)

If i = 1 Then
val_ec_qual = Me.Controls("ectaux" & i).Value
Else
val_ec_qual = val_ec_qual & "," & Me.Controls("ectaux" & i).Value
End If

Next i
val_ec_qual = "breaks=c(" & val_ec_qual & ")"
End If

```

Me.Controls sont les textbox créés dynamiquement lors de la saisie par l'utilisateur d'un nombre de classes par laquelle, il souhaite découper la variable quantitative en qualitative.

III. Discussions des résultats

3.1 Interprétation des résultats

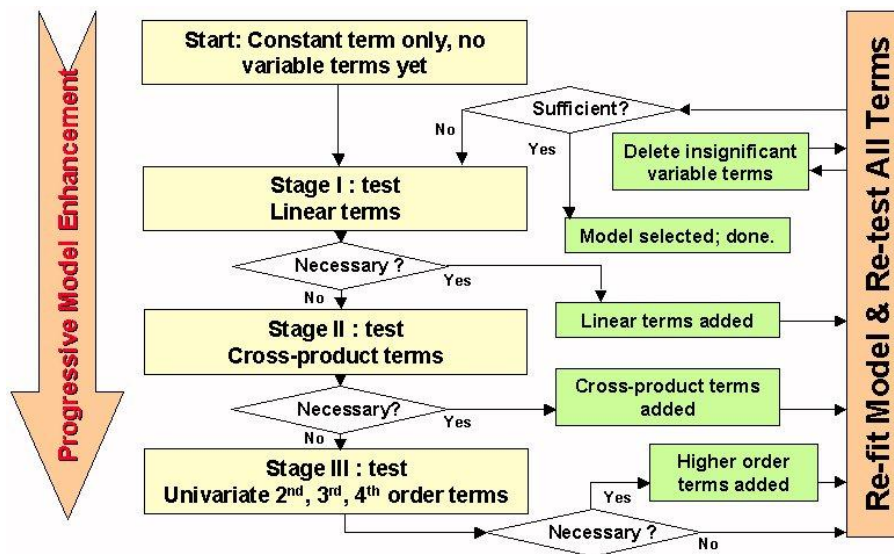
3.1.1 Critère de choix du modèle statistique

- Méthode stepwise

Les méthodes de type pas à pas consistent à considérer d'abord un modèle faisant intervenir un certain nombre de variables explicatives. Puis nous procédons par élimination ou ajout successif de variables suivant des tests de student, du R^2 ou C_p .

Nous parlons de la méthode :

- descendante lorsque nous éliminons des variables
- ascendante lorsque nous ajoutons des variables.
- stepwise est une combinaison de ces deux



- Matrice de confusion

Une matrice de confusion confronte les valeurs observées de la variable dépendante avec celles qui sont prédites. Cette technique permet de comptabiliser le nombre de bons et mauvais classement.

L'avantage de cette matrice est qu'elle permet à la fois de quantifier le taux d'erreur et de rendre compte de la structure de l'erreur.

RA obs/RA pred	RA	non RA
RA	a	b
non RA	c	d

Où :

- a sont les vrais RA
- b sont les faux non RA
- c sont les faux RA
- d sont les vrais non RA

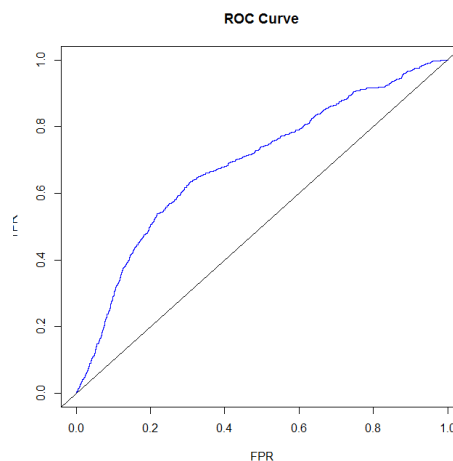
Le taux d'erreur est alors donné par :

$$taux_erreur = \frac{b + c}{a + b + c + d}$$

▪ Courbe ROC et AUC

La courbe ROC met en relation le taux de vraie positifs TVP, vraie RA dans notre cas (la sensibilité, le rappel) et le taux de faux positifs TFP, faux RA (TFP = 1 - Spécificité) dans un graphique nuage de points.

La courbe ROC généralise cette idée en faisant varier sur tout le continuum des valeurs possibles entre 0 et 1. Pour chaque configuration, nous construisons la matrice de confusion et nous calculons TVP et TFP.



Il est possible de caractériser numériquement la courbe ROC en calculant la surface située sous la courbe. C'est le critère AUC. Elle exprime la probabilité de placer un individu positif devant un négatif. Ainsi, dans le cas d'une discrimination parfaite, les positifs sont sûrs d'être placés devant les négatifs, nous avons $AUC = 1$. A contrario, si le classifieur attribue des scores au hasard, il y a autant de chances de placer un positif devant un négatif que l'inverse, la courbe ROC se confond avec la première bissectrice, nous avons $AUC = 0.5$. C'est la situation de référence, notre classifieur doit faire mieux. On propose généralement différents paliers pour donner un ordre d'idées sur la qualité de la discrimination.

Valeur de l'AUC	Commentaire
$AUC = 0.5$	Pas de discrimination.
$0.7 \leq AUC < 0.8$	Discrimination acceptable
$0.8 \leq AUC < 0.9$	Discrimination excellente
$AUC \geq 0.9$	Discrimination exceptionnelle

3.1.2 Comment interpréter un taux de RA ?

La régression logistique affecte donc un score suivant les caractéristiques du prêt. Ce score est la probabilité pour le dossier d'effectuer un remboursement anticipé.

Exemple :

Soit le modèle suivant :

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  9.935e-01  1.006e-01   9.873  < 2e-16 ***
ectaux       1.552e-01  9.869e-03  15.729  < 2e-16 ***
tmm         -2.148e-01  1.984e-02 -10.831  < 2e-16 ***
Mtprêt       2.088e-06  1.475e-07  14.154  < 2e-16 ***
âge          1.989e-01  7.184e-02   2.769  0.00563 **
Matrest      1.334e-01  7.169e-02   1.860  0.06284 .
durée       -1.593e-01  7.153e-02  -2.227  0.02593 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Soit une nouvelle observation :

ldrc00	Matrest	âge	durée	Mtprêt	Mensualité	ectaux	tmm	CRD	RA
127528	15,19	6,99	22,18	178100	1027,59	-0,54	4,29	143000,79	0

Variables	Coefficient	Nouvelle obs	Produit	Score	Modalité
constante	0,99345051	1	0,99345051	0,560316345	1
ectaux	0,15522708	-0,54	-0,08382262	exp(somme)/(1+exp(somme))	
tmm	-0,21483249	4,29	-0,92163138		
Mtprêt	2,0876E-06	178100	0,37179363		
âge	0,19888214	6,99	1,39018618		
Matrest	0,13337252	15,19	2,02592863		
durée	-0,15930834	22,18	-3,53345893		
		Somme	0,24244602		

On remarque ici que le prêt est considéré à tort comme RA.

Comment approcher le score et la modalité pour les RA ?

➤ Approche 1 :

On considère le score sans la modalité.

Le score correspond à un pourcentage à prélever du CRD :

Ici le CRD est de 143000.79€

56% sera considéré comme RA soit 80125,68€

➤ Approche 2 :

On considère la modalité et le score.

Si la modalité est de 1, le score correspond à un pourcentage à prélever du CRD.

Si la modalité est 0 aucune partie du CRD n'est prélevée.

Ici le CRD est de 143000.79€

56% sera considéré comme RA soit 80125,68€

➤ Approche 3 :

On considère la modalité et non le score.

Si la modalité est de 1 alors l'intégralité du CRD sera considérée comme RA.

Si la modalité est 0 aucune partie du CRD n'est prélevée.

Ici le CRD est de 143000.79€

Donc la totalité est considérée comme RA soit 143000.79€

Plus d'exemples :

Idrc00	Matrest	âge	durée	Mtprêt	Mensualité	ectaux	tmm	CRD	RA	DA	prevision	pred.moda	CRD RA	Approche 1	Approche 2	Approche 3
127528	15,19	6,99	22,18	178100	1027,59	-0,54	4,29	143000,79	0	21	0,5603163	1	0	80125,68	80125,68	143000,79
127604	4,27	7,74	12,01	95000	0	-0,89	4,29	39735,43	0	21	0,5811607	1	0	23092,67	23092,67	39735,43
127759	23,45	7,58	31,02	103000	571,08	1,11	4,29	93251,73	0	21	0,538053	1	0	50174,38	50174,38	93251,73
127765	21,36	6,66	28,02	110100	400,1	0,51	4,29	104471,91	0	21	0,522584	1	0	54595,34	54595,34	104471,91
127774	18,27	6,74	25,02	106403	521,32	0,61	4,29	98206,26	0	21	0,5448583	1	0	53508,49	53508,49	98206,26
127856	0,35	7,66	8,01	3700	0	1,46	4,29	237,88	0	21	0,6457095	1	0	153,6014	153,6014	237,88
127865	14,77	7,49	22,27	92600	0	0,41	4,29	75313,36	0	21	0,55984	1	0	42163,43	42163,43	75313,36
127911	9,6	5,9	15,51	149000	0	-0,19	4,29	100332,45	0	21	0,5832867	1	0	58522,58	58522,58	100332,45
128010	15,02	7,32	22,35	144500	0	0,11	4,29	110630,97	0	21	0,5717702	1	0	63255,49	63255,49	110630,97
128056	0,35	5,66	6,01	4428	0	-1,79	4,29	329,53	0	21	0,504529	1	0	166,2574	166,2574	329,53
128057	19,36	5,66	25,02	98000	511,54	0,31	4,29	87643,1	0	21	0,5115955	1	0	44837,82	44837,82	87643,1
128113	22,61	7,41	30,02	204000	1052,87	1,11	4,29	176539,72	0	21	0,5931048	1	0	104706,6	104706,6	176539,72
128131	2,85	7,16	10,01	63766	0	-0,34	4,29	21259,46	0	21	0,5894249	1	0	12530,85	12530,85	21259,46
128233	8,69	7,32	16,01	90500	0	-0,44	4,29	60845,44	0	21	0,563835	1	0	34306,79	34306,79	60845,44
128247	5,35	7,32	12,67	128000	0	-0,39	4,29	61408,25	0	21	0,6057331	1	0	37197,01	37197,01	61408,25
128249	0,68	7,32	8,01	13000	0	1,46	4,29	1487,55	0	21	0,6447497	1	0	959,0975	959,0975	1487,55
128294	13,52	6,49	20,01	72100	339,53	-0,24	4,29	63657,84	0	21	0,522799	1	0	33280,25	33280,25	63657,84
128321	16,03	6,41	22,44	90000	0	0,01	4,29	69066,02	0	21	0,5247718	1	0	36243,9	36243,9	69066,02
128325	19,44	5,58	25,02	117160	581,5	-1,39	4,29	103842,33	0	21	0,454442	0	0	47190,31	0	0
128506	11,77	3,99	15,76	101775	0	-0,09	4,29	81291,9	0	21	0,5306975	1	0	43141,41	43141,41	81291,9
128606	23,03	6,99	30,02	70849	374,41	1,11	4,29	64921,84	0	21	0,5178278	1	0	33618,33	33618,33	64921,84
128620	2,26	6,74	9,01	59550	0	-0,59	4,29	17387,57	0	21	0,5771461	1	0	10035,17	10035,17	17387,57
128656	0,02	5,99	6,01	3915	0	-1,64	4,29	58,89	0	21	0,515482	1	0	30,35674	30,35674	58,89
128682	13,02	6,99	20,01	58003	0	-1,34	4,29	41669,2	0	21	0,4809682	0	0	20041,56	0	0
128716	23,03	5,83	28,86	130300	645,1	0,31	4,29	109819,74	0	21	0,50634	1	0	55606,13	55606,13	109819,74
128836	25,87	4,49	30,36	125679	0	0,31	4,29	120926,28	0	21	0,4722855	0	0	57111,73	0	0
128846	14,44	5,58	20,01	115100	0	0,01	4,29	92557,09	0	21	0,5403001	1	0	50008,6	50008,6	92557,09

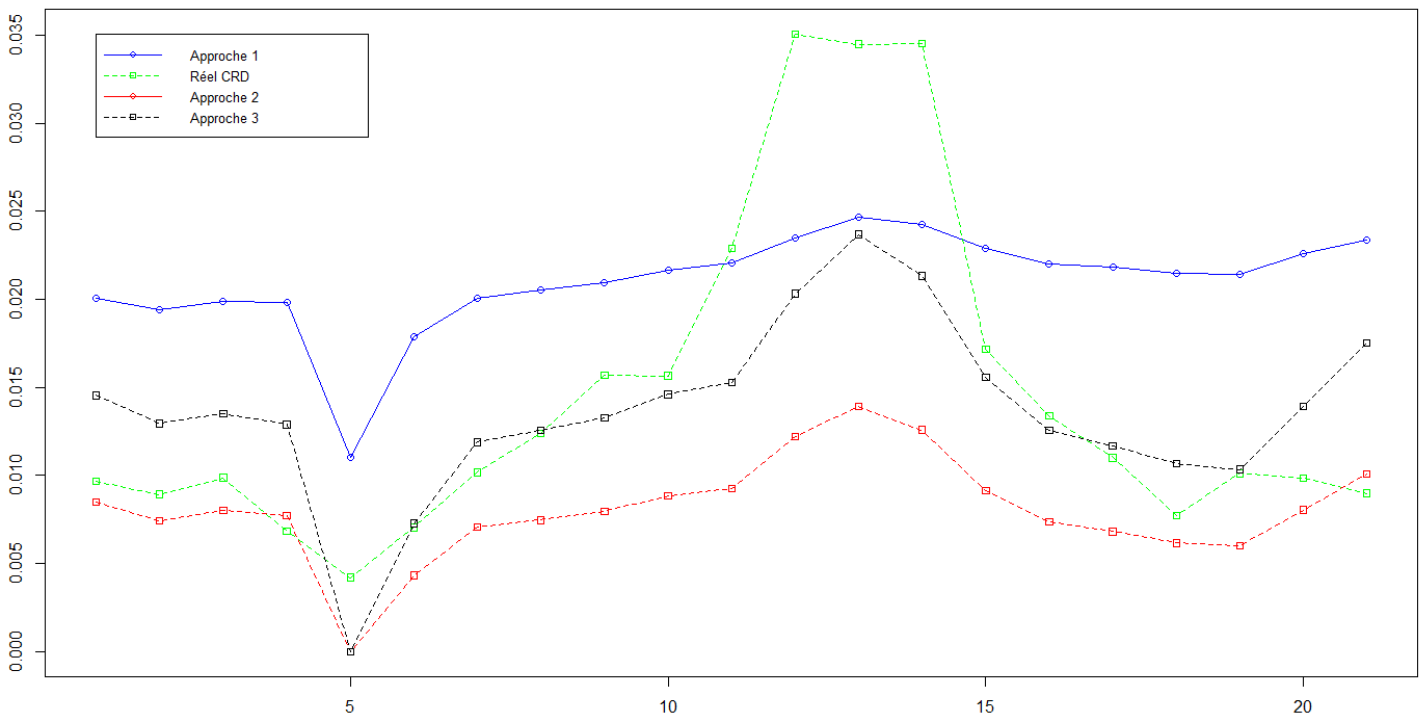
Ainsi pour un stock nous avons :

	Nb DA	SOMME CRD	
	21	5 321 133 354,79	
	Somme Approche 1	Somme Approche 2	Somme Approche 3
	2 612 958 358,41	1 127 757 753,61	1 956 060 682,90
	% RA Approche 1	% RA Approche 2	% RA Approche 3
	49,11%	21,19%	36,76%
Pondération	2,34%	1,01%	1,75%
	CRD RA observée	% RA observée	
	47 807 116,55	0,90%	

Et ainsi pour chaque date d'arrêtée nous pouvons tracer l'évolution des RA :

DA	Réel	Approche 1	Approche 2	Approche 3	AUC	MC
1	0,97%	2,01%	0,85%	1,45%	0,69	23,15
2	0,89%	1,94%	0,74%	1,29%	0,70	20,89
3	0,99%	1,99%	0,80%	1,35%	0,67	21,66
4	0,68%	1,98%	0,77%	1,29%	0,68	19,79
5	0,42%	1,10%	0,00%	0,00%	0,57	0,47
6	0,70%	1,79%	0,43%	0,73%	0,69	12,45
7	1,02%	2,00%	0,71%	1,19%	0,68	18,93
8	1,24%	2,05%	0,75%	1,26%	0,68	19,42
9	1,57%	2,09%	0,80%	1,33%	0,66	19,87
10	1,57%	2,17%	0,88%	1,46%	0,66	20,48
11	2,29%	2,21%	0,92%	1,53%	0,64	21,25
12	3,50%	2,35%	1,22%	2,03%	0,62	25,02
13	3,45%	2,47%	1,39%	2,37%	0,49	58,98
14	3,45%	2,43%	1,26%	2,13%	0,51	54,18
15	1,72%	2,29%	0,91%	1,56%	0,52	45,27
16	1,34%	2,20%	0,74%	1,26%	0,54	39,27
17	1,10%	2,18%	0,68%	1,17%	0,54	37,11
18	0,77%	2,15%	0,62%	1,07%	0,56	34,62
19	1,01%	2,14%	0,60%	1,03%	0,57	33,84
20	0,98%	2,26%	0,80%	1,39%	0,55	38,50
21	0,90%	2,34%	1,01%	1,75%	0,51	46,05

Back test RA TF CRD



Où pour rappel le taux de RA réels est donnée par :

$$Taux_RA_reel = \frac{Somme_CRD_RA}{Somme_CRD}$$

Indictaeur	Approche1	Approche2	Approche3
MAPE (%)	66.3418784	39.3226675	18.9980676
RMSE (%)	0.8205525	0.5588077	0.2531774

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2} \quad MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right| :$$

L'approche 3 est celle qui interprète le mieux les résultats où l'on considère la modalité et non le score.

3.1 Perspective d'évolution

Le modèle étant actuellement contraint à deux variables explicatives, dûes à un second logiciel exploitant les données en sortie.

L'idée serait alors de passer sur un modèle multi variables hors de ce logiciel afin d'avoir plusieurs variables explicatives.

C'est l'idée que j'ai anticipée ici en attendant la nouvelle base de données complète.

Une autre idée a été avancée, celle de mettre en place dans le modèle une variable qui considère l'opération globale du prêt avec toutes ses opérations et dans un second temps, de découper la régression par plusieurs régressions suivant le type de produits de prêts.

Conclusion

Ce stage a perfectionné mes connaissances en R ce qui m'a permis de maîtriser une bonne partie de la logique de ce logiciel.

Des acquis aussi bien au niveau des procédures R ou de lignes de code en R.

Durant ce stage, ces compétences m'ont également permis de réaliser un produit répondant aux besoins de l'entreprise qui souhaitait une automatisation du processus .

Néanmoins le modèle reste à être améliorer.

Enfin, j'espère que par la suite, cette pratique et ces acquis vont me permette de trouver une opportunité pour une nouvelle expérience dans la finance à l'internationale.

Références et Sources

Forums d'entraide :

- Forum des professionnels en informatique
www.developpez.net
 - VBA :
<http://www.developpez.net/forums/f542/logiciels/microsoft-office/excel/>
 - R :
<http://www.developpez.net/forums/f1179/autres-langages/autres-langages/r/>
 - SQL :
<http://www.developpez.net/forums/f37/bases-donnees/langage-sql/>

Cours statistiques :

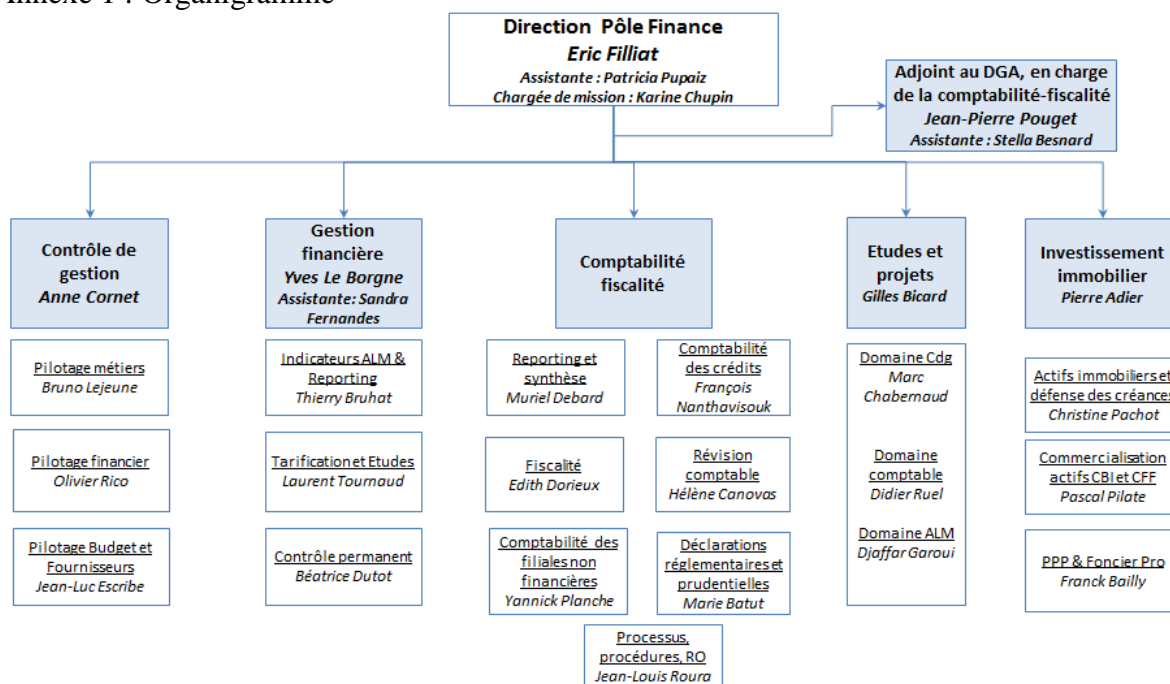
- Ricco Rakotomalala
<http://eric.univ-lyon2.fr/~ricco/cours/index.html>
- Statistique avec R
<http://math.agrocampusouest.fr/infoglueDeliverLive/enseignement/support2cours/livres/statistiques.avec.R>

Cours Rexcel :

- Erich Neuwirth
<http://rcom.univie.ac.at/>
Mail d'entraide : rcom-l@lists.univie.ac.at

Annexes

Annexe 1 : Organigramme



Résumé :

Ce stage a pour but la calibration d'un modèle de remboursement anticipé de prêt. Ce phénomène est modélisé par modèle logistique, dont ce dernier a été dynamisé suivant un interfacage de R et VBA, appelé RExcel.

Cette interface permet de tirer un jeu de données, de choisir un modèle avec variables quantitatives ou qualitatives et de le backtester en aillant des indicateurs de performances.

Mots-clefs : Jeux de données, ALM, Régression logistique, RExcel

Abstract:

The purpose of this internship is to calibrate a model of anticipated repayment.

This phenomenon is modeled by logistic regression, which it has been boosted following a interfacing R & VBA called RExcel.

This interface allows sampling data, choosing a model with quantitative or qualitative variables and to backtest the chosen model and getting performance indicators.

Keywords: Set of data, ALM, Logistic regression, RExcel