

Classification et prévisions du risque de rachat en Assurance-Vie

Bordeaux - 1/09/2010

Journées Modélisation Aléatoire et Statistiques



Plan de la présentation

1	Motivations de modélisation du rachat	03
2	Modèles de segmentation, prévisions	06
3	Application numérique	09
4	Conclusions	17
5	Annexes: crises de corrélation	18
6	Bibliographie	23



Design de nouveaux produits :

- Hypothèse sur le taux de rachat moyen ;
- Impact direct sur la gestion actif/passif et + particulièrement sur l'Embedded Value (E.E.V.) de la compagnie d'assurance.



Compréhension des comportements :

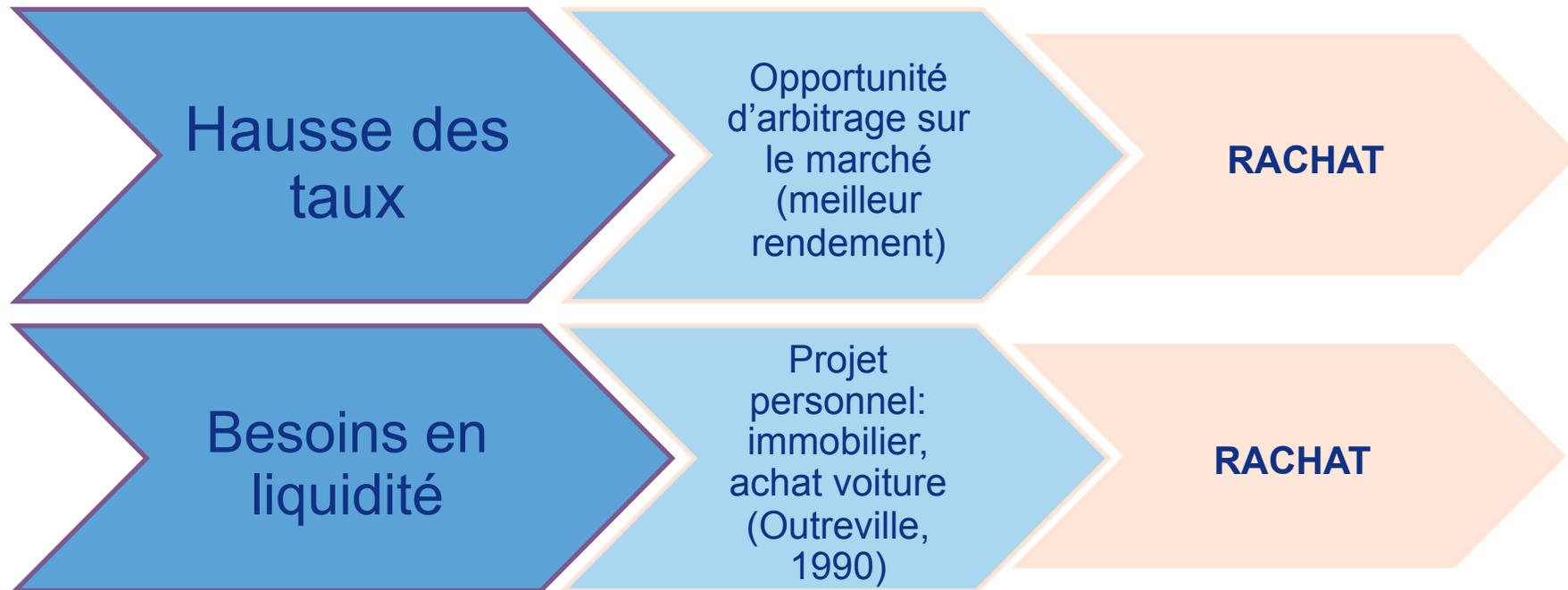
- Quels sont les facteurs discriminants du rachat ?
- La compréhension du risque comportemental est essentielle pour adapter les caractéristiques produits.



Segmentation du risque :

- Pour les nouvelles affaires ;
- Meilleure évaluation du risque lors de la souscription et adaptation éventuelle des pénalités (baisse ou hausse).

Intuitions : 2 approches



Mais pourquoi les rachats sont si difficiles à modéliser? Car ils dépendent de beaucoup de facteurs :

- desirs personnels, caractéristiques de l'assuré et besoins,
- caractéristiques du contrat, temps passé (ancienneté),
- contexte économique et financier,
- aspects socio-culturels.

Principales modélisations des taux de rachat

■ 4 différentes approches pour modéliser les rachats :

- *Approche financière* [4], [5], [14] : valorisation option de rachat, pas comportemental.
- *Approche statistique* [10] : **collective**. Données empiriques permettent de calibrer la fonction de rachat suivante :

$$r_d = r_0 * [1 - a * \ln(d + 1) * (\ln(d + 1) - b)]$$

- *Approche économique* [10] : **individuelle**. Microéconomie, théorie de l'utilité espérée, mais est-ce vraiment le raisonnement de l'assuré (rationalité remise en cause)?

- *Approche probabiliste / économétrique* : **individuelle, notre choix**. Modèles de segmentation, pour définir des classes de risques sur une population donnée (GLM et autres).

➡ DECLENCHEURS + Classification et prévisions des rachats

1^{ère} partie

Classification et prévisions du risque de rachat *(situation éco classique)*

Article associé :

« Surrender triggers in Life Insurance : classification and predictions »

Co-auteurs : V. Maume-Deschamps & S. Loisel, ISFA, Université Lyon 1

CART ([1]), process steps ([6], [7]) partitionnement récursif

1. Gestion données

- Supprimer données aberrantes ;
- Catégorisation éventuelle de variables continues ;
- Problème de données « unbalanced ».

2. Construction de l'arbre

- Fonction d'impureté règle de division binaire ;
- Règle d'arrêt de division ;
- Objectif: atteindre une homogénéité maximale.

3. Optimisation, élagage

- Suppression de branches de l'arbre: éviter sur-apprentissage, capturer tendance sans le bruit;
- Introduction d'un coût de complexité / minimisation de l'erreur.

4. Classification

- Définir la fonction de classification;
- Assigner les classes aux observations;
- Calculer l'erreur: resubstitution, échantillon test, validations croisées.

Modèle LOGIT : rappels ([2], [3])

■ Modèle de choix :

- * Fréquemment utilisé en Marketing et Economie (en risque décès chez AGL)
- * Applications nombreuses et nombreuses publications en Santé (prévisions)

■ Hypothèses : indépendance entre assurés, covariables indépendantes.

■ Equation régression logistique:

$$\text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

■ Estimation de la proba. indiv. finale de racheter, interprétation des résultats

$$\hat{p} = \Phi(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k)$$

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

■ Intervalle de confiance du taux de rachat du portefeuille (collectif, risques agrégés). Avec $\hat{p} = N^s/n$, on obtient au seuil de 5% :

$$CI(p) = \left[\hat{p} - 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Scope :

- * Produits « Mixtes »
- * 28506 contrats
- * 15571 rachats
- * Période d'étude:
01/01/1999 - 31/12/2007

Applications numériques CART & LOGIT

Variables explicatives du modèle :

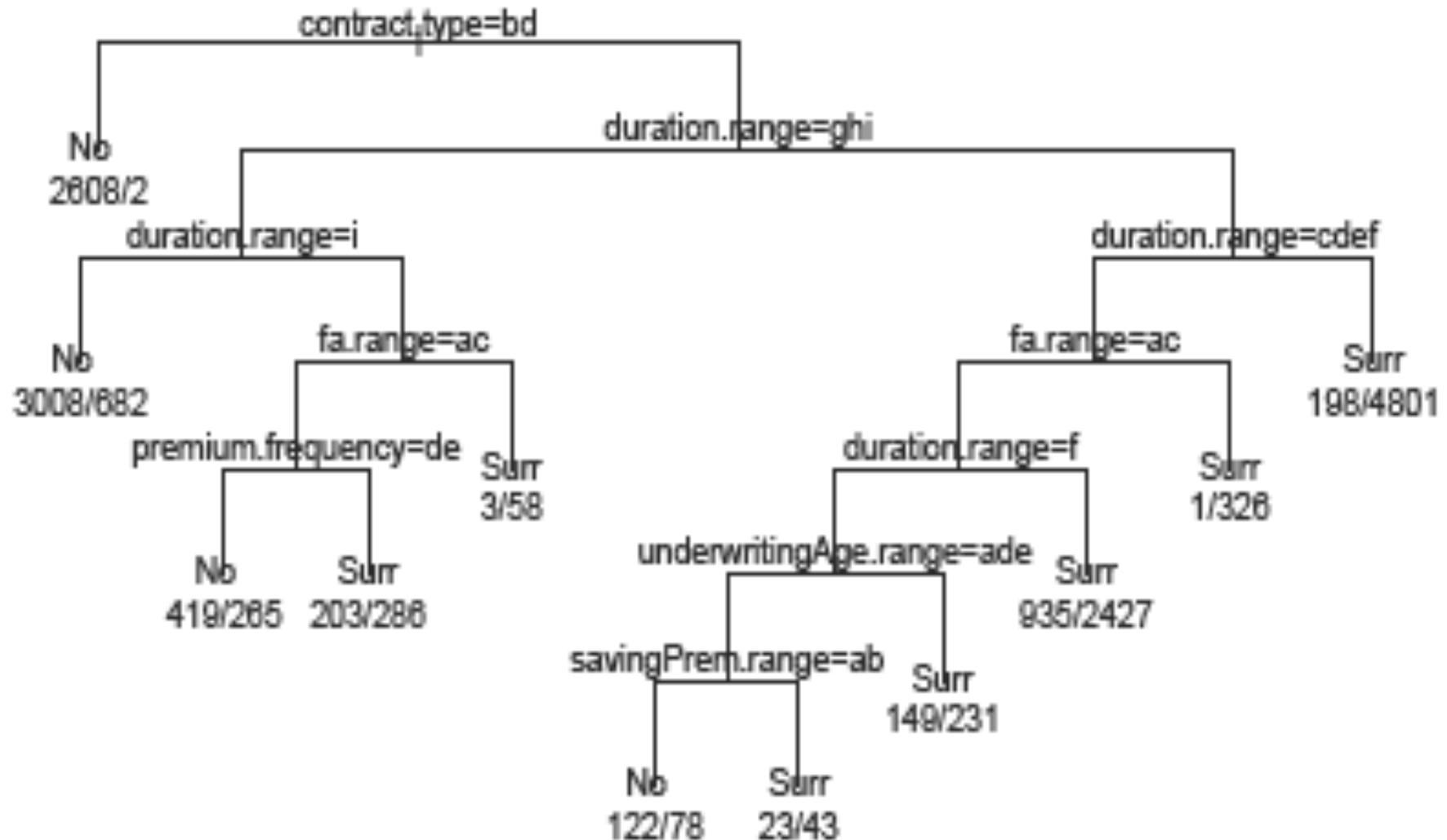
- * sexe
- * âge
- * ancienneté du contrat
- * prime de risque
- * prime d'épargne
- * fréquence de prime
- * type de contrat (participation au bénéf)

Software :

- Open source langage de stat : R
- Packages:
MASS, rpart, RandomForest



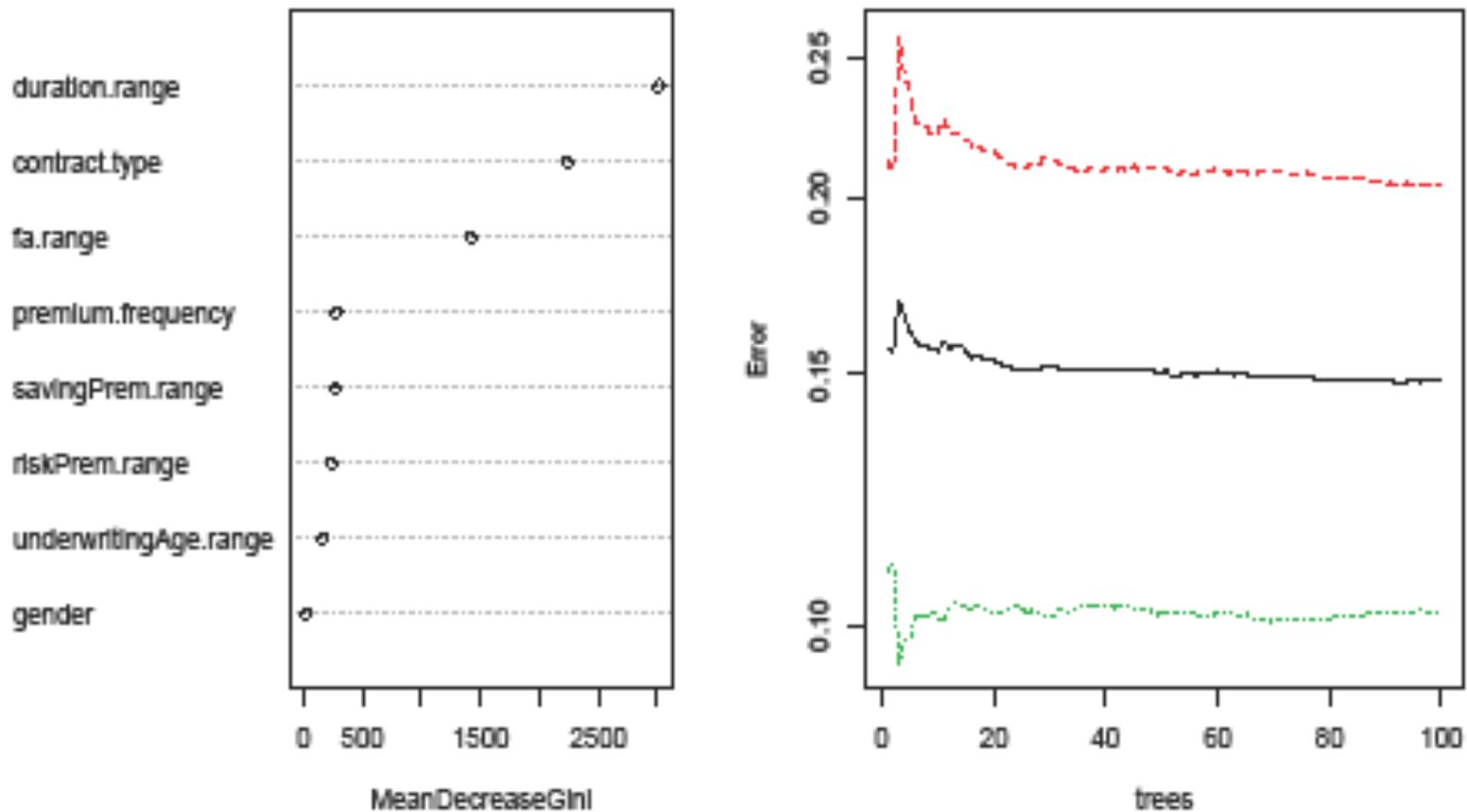
Arbre final élagué retenu



Random forests & « out-of-bag » error

([8], [9])

Figure 5: On the left, the importance of explanatory variables. On the right, the number of trees required to stabilize the *out-of-bag* errors : the black line is the overall error, the green line is the error of the category “surrender” and the red one for the category “no surrender”.



Matrices de confusion (CART)

Table 1: The confusion matrix for T_{max} on the validation sample.

	observed Y = 0	observed Y = 1
predicted Y = 0	4262	1004
predicted Y = 1	728	5644

Table 2: The confusion matrix for the pruned tree on the validation sample.

	observed Y = 0	observed Y = 1
predicted Y = 0	4188	1078
predicted Y = 1	664	5708

Table 3: The confusion matrix of the classifier by the Random Forest.

	observed Y = 0	observed Y = 1
predicted Y = 0	10327	2608
predicted Y = 1	1592	13979

Taux d'erreur correspondant :

- Arbre maximal : 14,88 %
- Arbre élagué : 14,96 %
- Forêts aléatoires : 14,73 %

Coef. de régression logistique

Coef. (var. type)	modality : correspondance	coefficient estimate	std error	p-value	effect
β_0 (continuous)		10.63398	1.48281	7.42e-13	> 0
$\beta_{duration}$ (categorical)	1 : [0,12[(in month)	0 (reference)			nul
	2 : [12,18[-1.31804	0.15450	< 2e - 16	< 0
	3 : [18,24[-2.66856	0.14016	< 2e - 16	< 0
	4 : [24,30[-2.75744	0.14799	< 2e - 16	< 0
	5 : [30,36[-3.09368	0.14294	< 2e - 16	< 0
	6 : [36,42[-3.54961	0.15080	< 2e - 16	< 0
	7 : [42,48[-3.72161	0.14980	< 2e - 16	< 0
	8 : [48,54[-4.10431	0.15772	< 2e - 16	< 0
	9 : > 54	-5.49307	0.14037	< 2e - 16	< 0
$\beta_{premium\ frequency}$ (categorical) (in month)	Monthly	0 (reference)			nul
	Bi-monthly	0.92656	0.62071	0.135504	> 0
	Quarterly	-0.03284	0.10270	0.749148	< 0
	Half-yearly	-0.22055	0.16681	0.186128	< 0
	Annual	0.43613	0.10690	4.51e-05	> 0
$\beta_{underwriting\ age}$ (categorical)	Single	-0.28494	0.38155	0.455177	< 0
	1 : [0,20[(years old)	0 (reference)			nul
	2 : [20,30[0.28378	0.13912	0.041376	> 0
	3 : [30,40[-0.01146	0.13663	0.933163	< 0
	4 : [40,50[-0.26266	0.14077	0.062054	< 0
	5 : [50,60[-0.42098	0.15136	0.005416	< 0
	6 : [60,70[-0.66396	0.19531	0.000675	< 0
7 : > 70	-0.75323	0.23417	0.001297	< 0	
$\beta_{face\ amount}$ (categorical)	1* :	0 (reference)			nul
	2* :	-5.79014	1.46592	7.82e-05	< 0
	3* :	-7.14918	1.46631	1.08e-06	< 0
$\beta_{risk\ premium}$ (categorical)	1* :	0 (reference)			nul
	2* :	0.36060	0.11719	0.002091	> 0
	3* :	0.26300	0.14041	0.061068	> 0
$\beta_{saving\ premium}$ (categorical)	1* :	0 (reference)			nul
	2* :	0.93642	0.13099	8.74e-13	> 0
	3* :	1.32983	0.14955	< 2e - 16	> 0
$\beta_{contract\ type}$ (categorical)	PP con PB	0 (reference)			nul
	PP sin PB	-16.79213	114.05786	0.882955	< 0
	PU con PB	-7.48389	1.51757	8.16e-07	< 0
	PU sin PB	-12.43284	1.08499	< 2e - 16	< 0
β_{gender}	Female	0 (reference)			nul
	Male	-0.08543	0.04854	0.078401	< 0

* Note : for confidentiality reasons, the real ranges of the face amount, the risk premium and saving premium are omitted.

« Odd-ratios » associés

Odd-ratio	Ref.	Other modalities							
Duration	[0,12]	[12,18]	[18,24]	[24,30]	[30,36]	[36,42]	[42,48]	[48,54]	> 54
<i>nb surrenders</i>	3062	1740	1187	791	728	400	365	244	682
<i>empirical OR</i>		10.56	2.89	2.69	1.82	1.16	0.96	0.68	0.19
<i>modeled OR</i>		0.27	0.07	0.06	0.05	0.03	0.02	0.02	0.004
Premium freq.	Monthly	Bi-monthly	Quarterly	Half-Yearly	Annual	Single			
<i>nb surrenders</i>	2790	12	323	92	595	5387			
<i>empirical OR</i>		2.22	0.93	0.66	2.39	1.60			
<i>modeled OR</i>		2.52	0.97	0.80	1.55	0.75			
UW. age	[0,20[[20,30[[30,40[[40,50[[50,60[[60,70[> 70		
<i>nb surrenders</i>	258	1719	2165	2002	1490	1088	477		
<i>empirical OR</i>		1.16	1.06	1.25	1.63	2.67	3.28		
<i>modeled OR</i>		1.32	0.99	0.77	0.67	0.51	0.47		
Face amount	#1*	#2*	#3*						
<i>nb surrenders</i>	5361	684	3154						
<i>empirical OR</i>		0.14	0.12						
<i>modeled OR</i>		0.003	0.0008						
Risk prem.	#1*	#2*	#3*						
<i>nb surrenders</i>	3941	2987	2271						
<i>empirical OR</i>		1.50	0.92						
<i>modeled OR</i>		1.43	1.30						
Saving prem.	#1*	#2*	#3*						
<i>nb surrenders</i>	3331	1762	4106						
<i>empirical OR</i>		1.90	2.09						
<i>modeled OR</i>		2.55	3.78						
Contract type	PP con PB	PP sin PB	PU con PB	PU sin PB					
<i>nb surrenders</i>	3840	0	5357	2					
<i>empirical OR</i>		0	4.75	0.0008					
<i>modeled OR</i>		5.6e-08	0.0006	3.9e-06					

* Note : for confidentiality reasons, the real ranges of the face amount, the risk premium and saving premium are omitted.

Matrices de confusion (LOGIT), performance

Mauvaises prévis. d'Odd-ratios (malgré bonne tendance). Mais,

- compromis goodness-of-fit / **qualité prévisions (bon ici)**
- Étape de la catégorisation de variable cruciale
- Erreur de mauvaise classification : 15,03 %

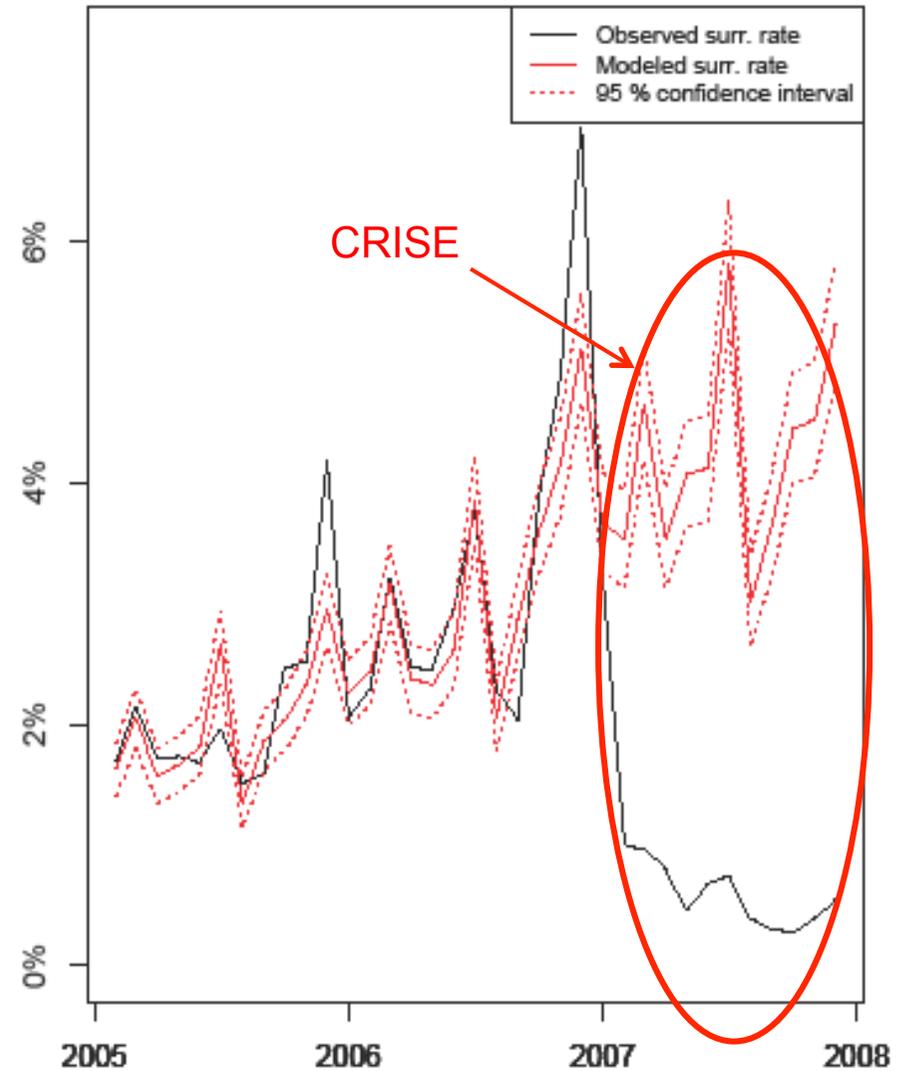
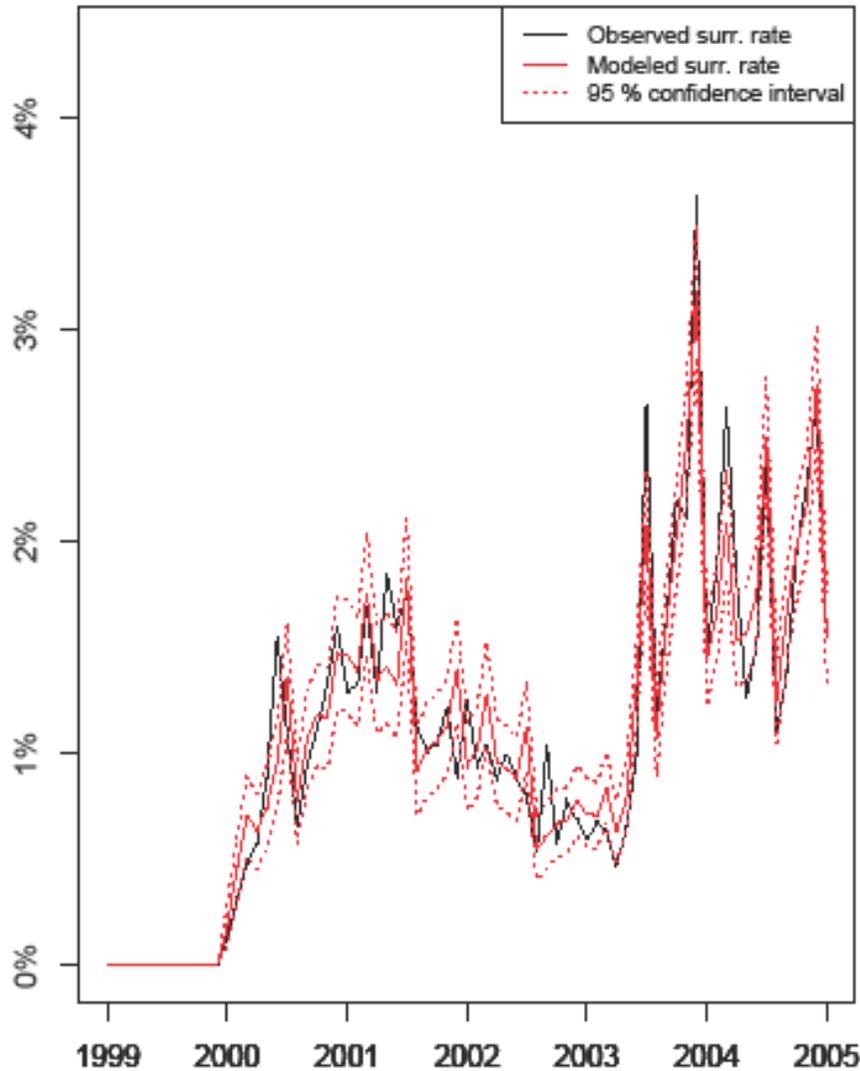
Table 5: The confusion matrix (LR model).

	observed Y = 0	observed Y = 1
predict Y = 0	#correct rejections 4153	#misses 637
predict Y = 1	#false risky policyholder 1113	#success 5735

	T_{max}	T_{pruned}	$T_{RandomForest}$	LR
Se	84.9%	84.1%	84.3%	90%
Sp	85.4%	86.3%	86.7%	78.9%
(1-Se)	15.1%	15.9%	15.7%	10%

Modél. dynamique (date & IBEX en + en covariables)

(Logit)



Quelques conclusions...

- CART pourrait être utilisé pour sélectionner les var.expl. les + discriminantes, ensuite en input du Logit (**réduction de l'espace**)
- CART/LOGIT ont un taux erreur faible, l'une ou l'autre des méthodes étant préférée en fonction du critère de performance à maximiser.
- Principales variables explicatives: **fiscalité**, **garanties**, **réseau de distribution**, **richesse** de l'assuré (et autres suivant les pays...)
- Ces méthodes permettent de découvrir des profils-type risqués:
 - Personnes très jeunes ou âgées,
 - À primes périodiques (+ prime fréquente et + la proba. de rachat est grande),
 - Personnes pauvres ou de niveau de vie moyen,
 - Personnes dont l'ancienneté atteint la possibilité de **défiscalisation**.
- Ces méthodes ne suffisent pas en période de crise...et problème: historiquement on n'a jamais vu des taux d'intérêt aussi bas...même s'il y a déjà eu des crises (2000); depuis 3 ans les souscriptions des nouvelles affaires se font à ces taux anormalement bas...crainte de **rachat massif** !

Annexes

Crises de corrélation

Articles associés :

* « From deterministic to stochastic surrender risk models: impact of correlation crises on economic capital »

Co-auteur : Stéphane Loisel, ISFA, Université Lyon 1

* « Les comportements de rachat en Assurance Vie en régime de croisière et en période de crise »

Co-auteurs : Stéphane Loisel (ISFA) & M.P. Gonon (Actuaris)

Idée d'origine des crises de corrélation

Environnement

- La situation économique joue un rôle majeur sur le comportement des assurés

Perturbation macroéco.

- Taux de rachat global augmente quand nous observons des perturbations des marchés financiers

Apparition d'un cycle

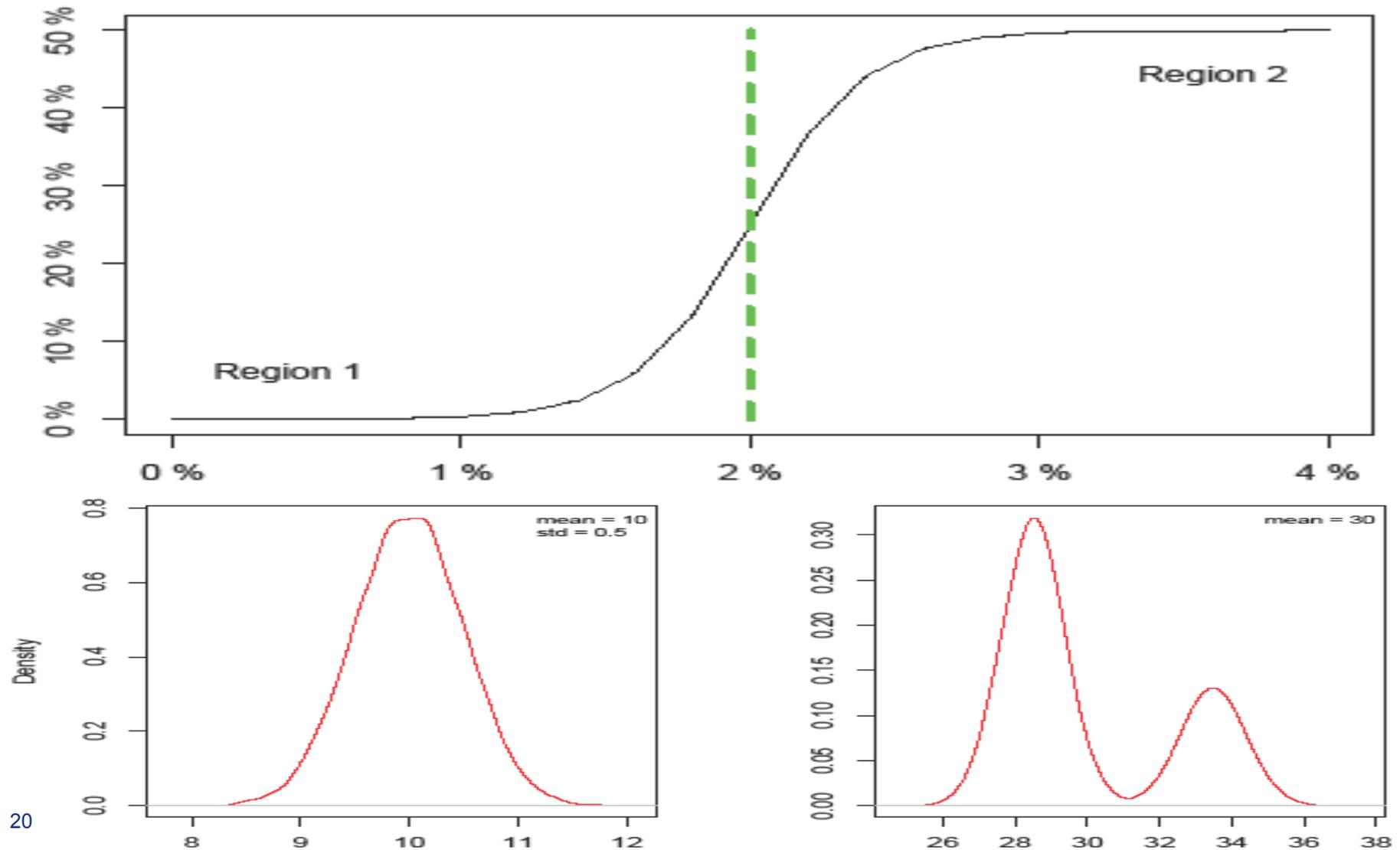
Crise de corrélation

- Corrélation entre les décisions des assurés augmente significativement en crise ou fortes récessions [15]

Modélisation classique du taux de rachat avec une gaussienne (moyenne et variance observées) devient **erronée**.

Changement de distribution

Surrender rate VS delta_r

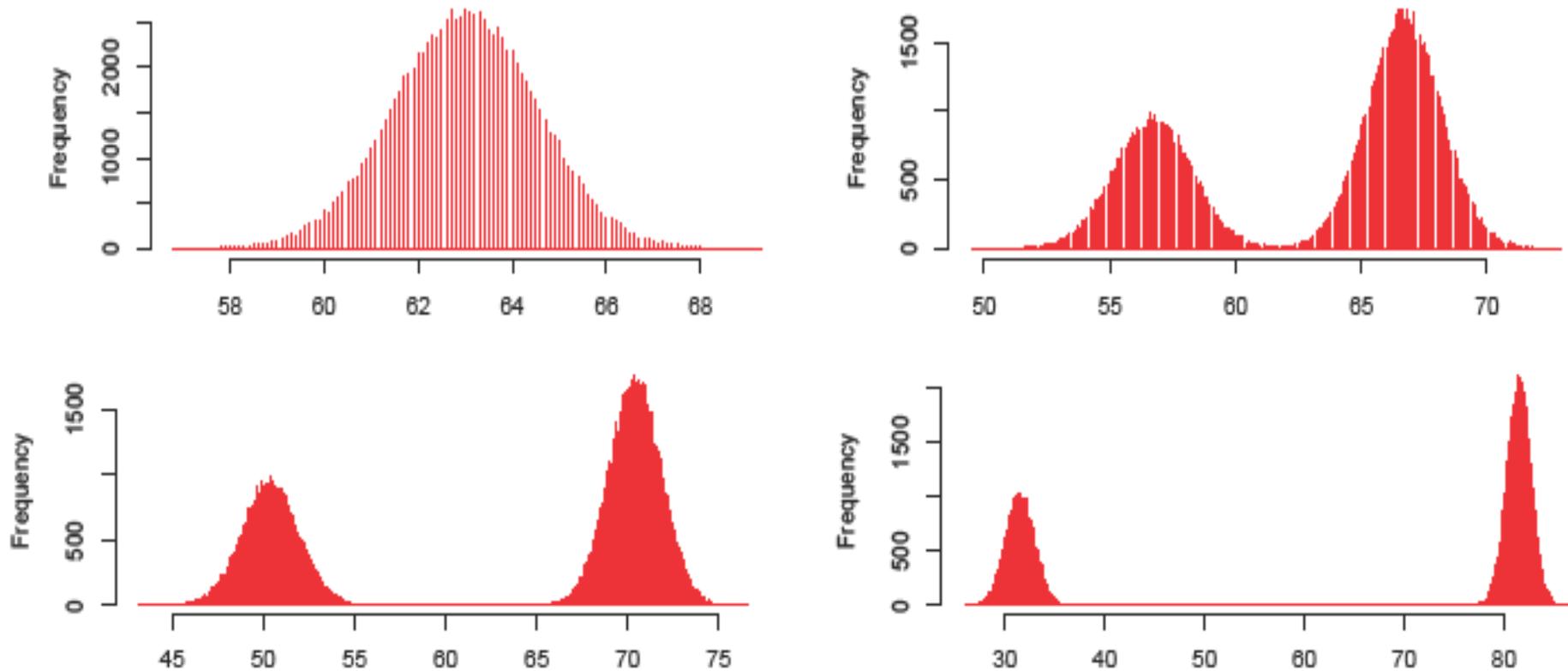


Pourquoi ce changement ?

- **La distribution bi-modale représente 2 comportements antagonistes: presque tous les assurés...**
 - Rachètent leur contrat ;
 - Ou non ;
- **En d'autres termes les assurés ont un comportement moutonnier, expliqué par la panique par exemple [12] ;**
- **Mathématiquement c'est une copule de Mardia, somme linéaire d'une copule d'indépendance et de la borne supérieure de Fréchet: les assurés ont donc la même décision de...**
 - Suivre le consensus de marché avec la proba p_0^2
 - Se comporter indépendamment avec la proba $(1 - p_0^2)$

$$C^M(u, v) = p_0^2 C^F(u, v) + (1 - p_0^2) C^I(u, v)$$

Conséquences sur la distribution du taux de rachat



Impact très important sur d'éventuelles réserves à mettre de côté pour faire face à une vague de rachat par exemple

Bibliographie

■ Books :

- [1] Classification And Regression Trees ; Breiman L. , Friedman J., Olshen R.A. & Stone C.J. ; Chapman and Hall (1984)
- [2] Logistic regression models ; Hilbe J.F. ; Chapman and Hall (2009)
- [3] Applied Logistic Regression 2nd Ed.; Hosmer D.W. & Lemeshow F. ; Wiley (2000)
- [15] Quantitative risk management ; McNeil, A. , Frey, R. et Embrechts, P. (2005) ; Princeton Series In Finance

■ Papers :

- [4] Modeling surrender and lapse rate with economic variables ; Kim C. ; North American Actuarial Journal p. 56-70 (2005)
- [5] Modeling insurance surrenders by the negative binomial model ; Kagraoka Y. ; Working Paper (2005)
- [6] Prévisions par arbres de classification ; Ghattas B. ; Mathématiques et Sciences Humaines 146 (1999)
- [7] Aggrégation d'arbres de classification ; Ghattas B. ; Revue de statistiques appliquées 2 (48), p. 85-98 (2000)
- [8] Random Forests ; Breiman L. ; Machine Learning (45) p. 5-32 (2001)
- [9] Bagging predictors ; Breiman L. ; Machine Learning (24) p. 123-140 (1996)
- [11] Human decision-making behavior and modeling effect ; Kim, C.N., Yang, K.H. et Kim, J. (2008) ; Decision Support Systems 45, 517-527
- [12] Human and social behavior in computational modeling and analysis of egress ; Pan, X., Han, C.S., Dauber, K. & Law (2006) ; Automation In Construction 15, 448-461
- [13] Decision field theory extensions for behavior modeling in dynamic environment using Bayesian belief network ; Lee, S., Son, Y-J et Jin, J. (2008) ; Information Sciences 178, 2297-2314
- [14] Fair valuation of participating policies with surrender options and regime switching ; Kuen, S.T. (2005) ; Insurance : Mathematics And Economics 37, 533-552

■ Memoirs :

- [10] Analyse des rachats d'un portefeuille vie individuelle : approche théorique et application pratique ; Fauvel, S. et Le Pévédic M. ; ENSAE (2007)

Présentation of AXA Global Life / 2010 – V1.0 (2010, June the 29th)

The data used in this presentation are the property of AXA Global Life and cannot be reused without the prior written consent of AXA Global Life.