

Evénements Rares et Modélisation de la Solvabilité des Sociétés d'Assurance Vie

M. Ridha Hsini
CEREMADE, Université ParisIX-Dauphine
e-mail: hsini@ceremade.dauphine.fr

December 24, 2003

Abstract

On se propose de détecter dans cet article l'ensemble d'indicateurs financiers qui peuvent jouer un rôle sur la santé financière des sociétés d'assurance Vie et peuvent, selon leur niveau, leur faire éviter ou, au contraire, les faire basculer en déficit.

Les méthodes économétriques liées à la modélisation de ce genre de problème sont ainsi décrites, et on met l'accent sur les méthodes appliquées aux événements rares correspondant au fait que la probabilité d'appartenir à une classe particulière d'un échantillon constitué de deux ou plusieurs classes d'observations, est très faible.

mots-clés : événements rares, régression logistique, échantillon pondéré, “choice-based sampling”, “separate sampling”, ratios financiers, pouvoir discriminant, corrélation parfaite, redressement.

1 Introduction

Ces dernières années, l'étude de la solvabilité des sociétés d'assurance a attiré l'attention des autorités de tutelle dans plusieurs pays. Les Etats Unis étaient les plus exposés à ce problème vue la complexité de leur réglementation dûe au fait que chaque état dispose de sa propre législation.

Différents types de modèles ont été donc établis et avaient pour objectif soit la détermination d'un capital minimum exigé soit la détection des ratios financiers les plus importants dans l'étude de la santé financière des sociétés d'assurance. C'est le cas du système IRIS (Insurance Regulatory Information System), FAST (Financial Analysis Tracking System) et RBC (risk-based capital), voir Hsini (1999).

Ces démarches diffèrent du système européen qui impose depuis des années aux sociétés d'assurance de disposer d'un capital minimum appelé "*marge de solvabilité*" qui, malgré tous ses inconvénients tels que la non prise en compte des différents types de risques, a réussi jusqu'ici à éviter trop de faillites sur le marché qui reste parmi les plus sûrs et les plus protégés dans le monde.

On propose ici une étude économétrique sur un ensemble de ratios financiers liés à l'activité des sociétés d'assurance vie et qui sont susceptibles d'expliquer le mieux leur santé financière. Ces méthodes ont déjà été presque toutes appliquées dans le premier volet portant sur les sociétés d'assurance non-vie.

Cependant, on va essayer dans cet article de mettre le point sur la particularité des **événements rares** qui apparaissent lorsque l'échantillon étudié est composé de plusieurs sous-populations dont l'une au moins se trouve *très mal représentée* par rapport à l'ensemble. On montrera ainsi que les observations de cette sous-population *rapporteraient plus d'informations* que les autres. En effet, elles *contribuent le moins à l'erreur standard* commise lors de l'estimation des paramètres.

Lorsqu'on est en présence d'un événement rare, d'autres méthodes d'échantillonnage et d'estimation s'imposent.

Dans ce travail on commence tout d'abord par étudier les résultats obtenus lorsqu'on applique la méthode d'échantillonnage aléatoire classique. Dans un second temps *on rajoute à la population initiale d'autres observations fictives identiques aux premières* pour pallier la dissymétrie entre les différentes sous-populations qui la composent. La méthode consiste à dupliquer un certain nombre de fois les observations initiales de façon à obtenir le même nombre d'observations dans les deux classes.

On applique ensuite la méthode d'échantillonnage la mieux adaptée aux problèmes d'événements rares; il s'agit d'une méthode de stratification endogène appelée "**choice-based sampling**" ou "**separate sampling**" qui consiste à sélectionner les observations par rapport à la variable endogène, formant ainsi une sous-population à classes en proportions à peu près égales.

On propose enfin d'ajuster la deuxième méthode de pondération *pour corriger le biais engendré surtout par l'amplification fictive de la taille du groupe mal représenté dans la population.*

Le travail est présenté comme suit. On commence par décrire les données utilisées et les ratios financiers supposés jouer un rôle sur l'évolution de la santé financière des sociétés d'assurance vie. On étudie ensuite les méthodes théoriques appliquées aux phénomènes d'événements rares. Dans une quatrième partie on aborde l'étude empirique en commençant par les deux premiers types d'échantillonnage, aléatoire et pondéré. Une sous-section est réservée à la méthode "choice-based sampling". On propose ensuite une méthode pour redresser la méthode de pondération avant de finir par des conclusions.

2 Données

2.1 Echantillon

On dispose des données comptables fournies par la "*Commission de Contrôle des Sociétés d'Assurance*". Ces données concernent les sociétés d'assurance Vie pour la période allant de 1992 à 1998. On dispose également des informations concernant les sociétés qui ont cessé leur activité durant toute cette période jusqu'à l'année 2000. Mais par manque d'autres informations, on fait par la suite l'hypothèse que toute société qui cesse son activité est considérée comme déficitaire et appartient ainsi à la classe 1.

Pour chaque année d'observation, on met dans la classe déficitaire les sociétés qui cessent leur activité pendant les trois années qui suivent, ce qui nous permet de gagner en taille au niveau de la classe déficitaire (on a en moyenne entre 10 et 20 observations pour chaque année d'observation). Quatre types d'échantillons seront utilisés :

-Dans un premier temps, On considère l'échantillon complet constitué en majorité par les observations de la classe 0 (solvable), ce qui provoque une dissymétrie importante entre les deux classes.

-Pour pallier ce problème de dissymétrie entre les deux classes, on décide de pondérer les observations de façon à obtenir le même nombre d'observations par classe. Les poids appliqués sont donc proportionnels à la taille de chaque classe ($\frac{n}{n_1}$ pour la classe 1, et $\frac{n}{n_0}$ pour la classe 0, où n_1 et n_0 représentent les tailles respectives des deux classes 1 et 0 et $n = n_0 + n_1$).

-On procède ensuite à un échantillonnage selon la méthode de "choice-based sampling" qui représente la méthode la plus adaptée lorsqu'il y a une forte disproportion entre les classes d'une population, autrement lorsqu'il s'agit d'étudier un événement rare.

-Enfin, on suggère un redressement de l'échantillon obtenu par la deuxième méthode d'échantillonnage pondéré. En effet, cette méthode de pondération a créé des observations fictives et a fait supposer qu'on a le même nombre d'observations dans chaque classe, ce qui est loin d'être vrai.

Pour se ramener à la réalité et corriger le biais engendré par cette méthode, on constitue d'abord une population à partir de la première *en dupliquant un certain nombre de fois les observations de la classe déficitaire*. Ainsi pour se ramener aux données initiales on pondère chaque observation de la classe déficitaire par l'inverse du nombre de duplications..

2.2 Choix des ratios financiers

Les ratios financiers utilisés dans ce travail sont spécifiquement liés à l'activité d'assurance sur la vie et couvrent dans leur ensemble le fonctionnement de ce marché.

Ces ratios sont définis comme suit :

symb	variable	définition
TCP	taux de croissance des primes	$(\text{Primes émises})_t / (\text{Primes émises})_{t-1}$
TCEP	taux de cessions (de primes)	$\text{Cessions}_t / (\text{Primes émises})_t$
TS	taux de sinistralité	$(\text{Charge des sinistres})_t / (\text{Primes émises})_t$
CON	taux de conservation	$(\text{Charge nette des sinistres})_t / (\text{Charge brute des sinistres})_t$
TCS	taux de croissance des sinistres	$(\text{Charge des sinistres})_t / (\text{Charge des sinistres})_{t-1}$
TCCP	taux de croissance des cessions	$\text{Cessions}_t / \text{Cessions}_{t-1}$
TFAA		$(\text{Frais d'acquisition et d'administration})_t / (\text{Primes émises})_t$
TACH		$(\text{Autres charges})_t / (\text{Primes émises})_t$
CAIPU		$(\text{Pr. émises contrats capitalisation primes uniques})_t / (\text{Primes émises})_t$
CAIPP		$(\text{Pr. émises contrats capi. primes périodiques})_t / (\text{Primes émises})_t$
CUCPU		$(\text{Pr. émises contrats en unité de compte P.U.})_t / (\text{Primes émises})_t$
CUCPP		$(\text{Pr. émises contrats en unité de compte P.P.})_t / (\text{Primes émises})_t$
CL441		$(\text{Pr. émises contrats L-441})_t / (\text{Primes émises})_t$
DIND		$(\text{Pr. émises contrats dommages corporels individuels})_t / (\text{Primes émises})_t$
DCCOL		$(\text{Pr. émises contrats dommages corporels collectifs})_t / (\text{Primes émises})_t$
CCDECE		$(\text{Pr. émises contrats collectifs décès})_t / (\text{Primes émises})_t$
CCVIE		$(\text{Pr. émises contrats collectifs vie})_t / (\text{Primes émises})_t$
TCE	taux de couv. des eng.	voir définition
RF	résultat financier	voir définition
TCRF	taux de croissance du résul. fin.	$(\text{résultat financier})_t / (\text{résultat financier})_{t-1}$
FPR		$(\text{fonds propres})_t / (\text{Primes émises})_t$
FRF		$(\text{fonds propres})_t / (\text{résultat financier})_t$
FSIN		$(\text{fonds propres})_t / (\text{charge des sinistres})_t$
RSOL	ratio de solvabilité	$(\text{fonds propres})_t / (\text{provisions techniques brutes})_t$
RTECP		$(\text{résultat technique})_t / (\text{Primes émises})_t$
REXEP		$(\text{résultat de l'exercice})_t / (\text{Primes émises})_t$
TCF	taux de croiss. des fonds	$(\text{fonds propres})_t / (\text{fonds propres})_{t-1}$
PPLPL		$(\text{Produits des Placements})_t / \text{Placements}_t$
RFPL		$(\text{Résultat Financier})_t / \text{Placements}_t$
PREPPL		$(\text{produits de réalisation})_t / (\text{Produits des Placements})_t$
PARPPL		$(\text{participations au résultat})_t / (\text{Produits des Placements})_t$
IMMPPL		$(\text{placements immobiliers})_t / (\text{Produits des Placements})_t$
AUTPPL		$(\text{autres revenus des placements})_t / (\text{Produits des Placements})_t$

Table 1: définition des ratios financiers par rapport au nouveau plan comptable

symb	variable	définition
TCP	taux de croissance des primes	Primes émises _t / Primes émises _{t-1}
TCEP	taux de cessions (de primes)	Cessions _t / Primes émises _t
TS	taux de sinistralité	(Charge des sinistres) _t / Primes émises _t
CON	taux de conservation	(Charge nette de sinistres) _t / (Charge brute de sinistres) _t
TCS	taux de croissance des sinistres	(Charge de sinistres) _t / (Charge de sinistres) _{t-1}
TCCP	taux de croissance des cessions	Cessions _t / Cessions _{t-1}
TCOM		Commissions _t / Primes émises _t
TCH		Charges _t / Primes émises _t
COMP		(Pr. émises contrats assurance complémentaire) _t / Primes émises _t
CAPI		(Pr. émises contrats capitalisation) _t / Primes émises _t
CAPACAV		(Pr. émises contrats capitalisation ACAV) _t / Primes émises _t
AAFF		(Pr. émises contrats autres affaires) _t / Primes émises _t
CGBACAV		(Pr. émises contrats grandes branches ACAV) _t / Primes émises _t
CCDECE		(Pr. émises contrats collectifs décès) _t / Primes émises _t
CCACAV		(Pr. émises contrats collectifs ACAV) _t / Primes émises _t
TCEN	taux de couv. des eng.	voir définition
RESFI	résultat financier	voir définition
TCRF	taux de croissance du résul. fin.	(résultat financier) _t / (résultat financier) _{t-1}
FOPR		(fonds propres) _t / Primes émises _t
FORF		(fonds propres) _t / (résultat financier) _t
FOSIN		(fonds propres) _t / charge des sinistres _t
RSOL	ratio de solvabilité	(fonds propres) _t / (provisions techniques brutes) _t
RTECP		(résultat technique) _t / Primes émises _t
REXEP		(résultat de l'exercice) _t / Primes émises _t
TCFON	taux de croiss. des fonds	(fonds propres) _t / (fonds propres) _{t-1}
PPLPL		(Produits des Placements) _t / Placements _t
RFPL		(Résultat Financier) _t / Placements _t

Table 2: définition des ratios financiers par rapport à l'ancien plan comptable

2.2.1 Quelques définitions

*le **taux de couverture des engagements** est définie par :

-Par rapport au nouveau plan comptable

$$TCE = \frac{PL + PLCUC + CCC + CRA}{PROV + PCUC + Per + ESSCP}, \text{ où,}$$

PL = placements, PLCUC = placements des contrats en unités de compte, CCC = comptes courants et caisse, CRA = compte de régularisation de l'actif,

PROV = provisions techniques brutes, PCUC = provisions pour contrats en unités de compte, Per = personnel, ESSCP = état-sécurité sociale-collectivités publiques.

-Par rapport à l'ancien plan comptable

$$TCE = \frac{\text{ensemble des valeurs} + \text{banques et chèques postaux}}{\text{personnel} + \text{état} + \text{provisions techniques brutes}}$$

*le **résultat financier** est défini par :

-par rapport au nouveau plan comptable

$$RF = RP + RPI + RAP + ARF - FF + APP - ACP$$

où RP = revenus des participations, RPI = revenus des placements immobiliers, RAP = revenus des autres placements, ARF = autres revenus financiers, FF = frais financiers, APP = autres produits de placements, ACP = autres charges de placements.

-par rapport à l'ancien plan comptable

$$RF = PVC + PP - CP - RESCAPI - DPROVDEP - MVC,$$

où PVC = plus-values de cessions, PP = produits des placements, CP = charges des placements, RESCAPI = réserves de capitalisation, DPROVDEP = dotations aux provisions pour dépréciation et MVC = moins-values de cessions.

*Pour le calcul du **ratio de solvabilité**, un ajustement a été appliqué aux provisions techniques. En effet, on a enlevé de ces provisions trois quart des provisions pour contrats en unités de compte. Cet ajustement n'a pas été appliqué sur les données correspondant à l'ancien plan comptable pour manque de données.

3 Méthode théorique

3.1 Événements rares

Pour étudier le problème d'insolvabilité des sociétés d'assurance, on considère une variable binaire Y prenant la valeur 1 si la société est déficitaire (événement) et 0 sinon (non-événement). On parle ainsi d'événement rare lorsque le nombre d'observations appartenant à la classe 1 est très réduit (des dizaines, voire des centaines de fois inférieur) par rapport à celui des observations de la classe 0, ce qui signifie que *la probabilité d'appartenance à la classe 1 est très faible par rapport à celle de la classe 0*.

La méthode adoptée pour ce genre de problème est la régression logistique qui permet d'écrire la probabilité $p(y/x, \theta)$ d'appartenir à la classe 1 sous la forme :

$$p(y/x, \theta) = \frac{1}{1 + \exp -(\alpha + x'\theta)} = F(x'\theta) = p,$$

où α est une constante, $\theta = (\theta_1, \dots, \theta_p)'$ et $x = (x_1, x_2, \dots, x_p)$. La fonction log-vraisemblance s'écrit sous la forme :

$$l(y/x, \theta) = \sum_{i=1}^n [y_i \ln(F(x'_i\theta)) + (1 - y_i) \ln(1 - F(x'_i\theta))]$$

Les deux dérivées premières s'en déduisent ainsi :

$$\frac{\partial l(y/x, \theta)}{\partial \theta} = \sum_{i=1}^n \left[\frac{y_i}{F(x'_i\theta)} - \frac{1 - y_i}{1 - F(x'_i\theta)} \right] f(x'_i\theta)x'_i$$

$$\frac{\partial^2 l(y/x, \theta)}{\partial \theta \partial \theta'} = - \sum_{i=1}^n \left[\frac{y_i}{F^2(x'_i\theta)} + \frac{1 - y_i}{(1 - F(x'_i\theta))^2} \right] f^2(x'_i\theta)x_i x'_i + \sum_{i=1}^n \left[\frac{y_i - F(x'_i\theta)}{F(x'_i\theta)(1 - F(x'_i\theta))} \right] f'(x'_i\theta)x_i x'_i$$

Ceci nous permet de calculer la matrice de variance-covariance de l'estimateur du maximum de vraisemblance du vecteur des paramètres θ qui vaut :

$$\text{Var}(\hat{\theta}) = - \left(\mathbb{E} \left[\frac{\partial^2 l(Y/x_i, \theta)}{\partial \theta \partial \theta'} \right] \right)^{-1} = \sum_{i=1}^n \left[\frac{f^2(x_i\theta)}{F(x_i\theta)(1 - F(x_i\theta))} x_i x'_i \right]^{-1}$$

Or dans le cas du modèle logit, on a : $f(x_i\theta) = F(x_i\theta)[1 - F(x_i\theta)]$. La variance se simplifie donc pour s'écrire :

$$\text{Var}(\hat{\theta}) = \sum_{i=1}^n [F(x_i\theta)(1 - F(x_i\theta))x_i x'_i]^{-1} = \sum_{i=1}^n [p_i(1 - p_i)x_i x'_i]^{-1}$$

On en déduit ensuite l'écart type qui mesure l'erreur standard de l'estimateur, ce qui permet de voir la particularité des événements rares et les problèmes y afférents. En effet, dans le cas des événements rares (où les p_i sont en principe très faibles pour les observations de la classe 1) *la régression logistique, malgré tous ses avantages, surestime les p_i pour la classe 1 par rapport à la classe 2*. Ainsi, le facteur $p_i(1 - p_i)$ sera plus grand pour les observations de la classe 1 que

pour celles de la classe 2 (voir King & Zing (2001)).

Les observations de la classe 1 contribuent donc le moins à la variance (qui dépend de l'inverse du facteur $p_i(1 - p_i)$) et par conséquent à l'erreur standard. Ceci signifie que **toute observation supplémentaire de la classe 1 apporterait plus d'information qu'une observation supplémentaire de la classe 2.**

3.2 Echantillonnage stratifié

Lorsque le problème étudié concerne des événements rares, les méthodes statistiques basées sur l'échantillonnage aléatoire ne sont plus appropriées. En effet, si l'on tire au hasard dans la population, l'échantillon obtenu contiendrait peu ou pas d'observations de la classe représentant l'événement rare. Les paramètres estimés ne reflèteraient plus la réalité des choses vu qu'ils ne seraient pas basés sur les différentes catégories ou classes de la population totale. D'autres méthodes d'échantillonnage sont donc nécessaires lorsqu'il s'agit pour nous d'étudier un événement rare. C'est ainsi que les économètres ont développé la méthode de stratification endogène ou **"choice-based sampling"** connue en épidémiologie sous le nom de **"case-control sampling"**.

Dans le cas d'événements rares, la méthode de "choice-based sampling" consiste à prendre à-peu-près le même nombre d'observations appartenant aux différentes classes de la population initiale et le choix de ces observations se fait **au hasard**. On s'intéresse ainsi aux distributions $p(x/y = j)$, $j=1, \dots, J$. Pour $Y = j$ fixé, la probabilité correspondante peut s'écrire selon la règle de Bayes :

$$p(x/y = j) = \frac{p(y = j/x, \theta) \cdot g(x)}{\sum_z p(y = j/z, \theta) \cdot g(z)}$$

où $g(\cdot)$ est la distribution marginale de X . La vraisemblance correspondant à l'échantillon considéré s'écrit ainsi :

$$L = \prod_{s=1}^J \prod_{k=1}^{n_s} \left(\frac{H_s p(y_k = s/x_k, \theta) g(x_k)}{\pi_s} \right)^{n_s}$$

où $\pi_s = \sum_x p(y = s/x, \theta) g(x)$ représente la probabilité marginale d'appartenir à la classe s , $s=1,2$, H_s la probabilité de tirage dans chaque classe fixée par l'économètre, et n_s la nombre d'observation tirées dans la classe s .

Cette vraisemblance calculée sous la méthode "choice-based sampling" est la même que celle calculée selon ce que Anderson (1972) appelle **"separate sampling"** où l'échantillon étudié contient des observations tirées au hasard dans chaque classe de la population initiale.

L'un des problèmes qui se pose avec l'échantillonnage retrospectif ou par stratification endogène est que la loi marginale des x (variables exogènes) ne peut plus être ignorée dans la vraisemblance comme dans le cas d'échantillonnage aléatoire ou prospectif (tirage conditionnellement aux exogènes) puisqu'elle apparait maintenant au dénominateur. Une méthode appropriée d'estimation est ainsi nécessaire.

3.2.1 Estimation

Sous l'hypothèse que π_s est connue, Manski et Lerman (1977) ont proposé un estimateur convergent et de loi asymptotiquement normale, il s'agit de l'estimateur du maximum de vraisemblance exogène pondéré (the weighted exogenous maximum likelihood estimator, WEMLE) qui consiste à remplacer la log-vraisemblance calculée ci-dessus par :

$$l = \sum_{s=1}^J \sum_{k=1}^n w_k \ln(p(y_k = s/x_k, \theta)),$$

où les poids w_k sont telles que $w_k = \frac{\pi_s}{H_s}$.

Lorsqu'elle est inconnue, la probabilité marginale π_s peut être estimée à partir d'un **échantillon auxiliaire** tiré dans la population totale, ou tout simplement par la proportion de chaque classe dans la population considérée. On peut ensuite utiliser le même estimateur de vraisemblance pondérée. Manski et McFadden (1981) et Cosslett (1981) ont proposé d'autres estimateurs plus efficaces que le premier. Celui de Cosslett est plus difficile à calculer.

En parallèle à ces travaux économétriques, d'autres travaux ont été effectués dans le *domaine épidémiologique* et ont beaucoup apporté à ce domaine. On peut citer par exemple le travail de Prentice & Pyke (1979). Celui de Anderson (1972) était plus général mais reste une référence dans ce domaine.

Lorsque la probabilité conditionnelle $p(y = s/x, \theta)$ est du type logistique ($p(y = s/x, \theta) = \frac{1}{1 + e^{-x'_i \theta}}$), on peut démontrer (voir Hsini (1999)) *qu'on peut ajuster le modèle par une correction a priori de la constante θ_0 qui devient :*

$$\theta_0^* = \theta_0 - \log\left(\frac{p_1}{p_2}\right),$$

où p_1 est ici la proportion d'observations sélectionnées dans la classe 1, et p_2 la proportion d'observations sélectionnées dans la classe 0. Pour les mêmes raisons évoquées dans Hsini (1999), on vérifie que la diminution est uniquement de $-\log(p_2)$.

Cette dernière décennie s'est développée une nouvelle méthode statistique d'inférence et estimation appelée **méthode de vraisemblance empirique**. Il s'agit d'une méthode non-paramétrique qui nous laisse la possibilité d'utiliser les méthodes classiques d'inférence et estimation telles que la méthode du maximum de vraisemblance, les équations estimantes ou la méthode des moments généralisés (GMM). *Notons tout de même que cette méthode est une reformulation des méthodes classiques.*

Considérons le cas simple de deux classes représentées par la variable binaire Y et notons

$$g_1(x) = p(x/y = 1) \text{ et } g_0(x) = p(x/y = 0),$$

où $x' = (x_1, x_2, \dots, x_p)$ est le vecteur des variables explicatives. On suppose toujours que la probabilité $p(y = 1/x, \theta)$ est du type logistique, on vérifie ainsi que l'on a :

$$g_1(x) = \frac{\exp(\alpha + \theta'x)}{\pi_1 [1 + \exp(\alpha + x'\theta)]} \cdot g(x),$$

et

$$g_0(x) = \frac{1}{\pi_0 [1 + \exp(\alpha + x'\theta)]} \cdot g(x).$$

ce qui implique que

$$g_1(x) = \exp(\alpha^* + x'\theta) \cdot g_0(x),$$

où $\alpha^* = \alpha + \log\left(\frac{1 - \pi_1}{\pi_1}\right)$ et sera estimé par $\hat{\alpha} + \log\left(\frac{n_0}{n_1}\right)$, où $\hat{\alpha}$ est l'estimateur du maximum de vraisemblance de α et n_0 et n_1 sont les tailles respectives des deux sous-échantillons. Notons $n_s(x)$, $s = 0, 1$ le nombre d'observations associées à x dans la classe s , la vraisemblance peut s'écrire sous la forme :

$$L = \prod_{s=1}^2 \prod_x [g_s(x)]^{n_s(x)}$$

Une autre façon d'écrire la vraisemblance serait de dire que tout se passe comme si l'on dispose de deux échantillons indépendants, le premier $(x_1, x_2, \dots, x_{n_0})$ tiré dans $g_0(x)$, et le second $(z_1, z_2, \dots, z_{n_1})$ (ou $(x_{n_0+1}, x_{n_0+2}, \dots, x_n)$) tiré dans $g_1(x)$ définie comme ci-dessus par $g_1(x) = h(x, \alpha^*, \theta) \cdot g_0(x)$, où $h(x, \alpha^*, \theta) = \exp(\alpha^* + x'\theta)$. Voir aussi Qin & Zhang (1997).

Notons $dG_0(x) = g_0(x)$ et supposons que cette distribution soit définie par des poids $p_i = dG_0(x_i)$ associés aux différents points observés dans l'échantillon. La vraisemblance s'écrit :

$$L(p, \alpha^*, \theta) = \prod_{i=1}^{n_0} p_i \prod_{j=1}^{n_1} \exp(\alpha^* + x'_j \theta) p_j = \prod_{i=1}^n p_i \prod_{j=1}^{n_1} \exp(\alpha^* + x'_j \theta),$$

d'où la log-vraisemblance :

$$l(p, \alpha^*, \theta) = \sum_{i=1}^n \log(p_i) + \sum_{j=1}^{n_1} (\alpha^* + x'_j \theta).$$

On s'intéresse à maximiser la fonction $l(p, \alpha^*, \theta)$ sous les contraintes :

$$p_i \geq 0, \sum_{i=1}^n p_i = 1, \text{ et } \sum_{i=1}^n p_i [h(x_i, \alpha^*, \theta) - 1] = 0,$$

où la dernière contrainte résulte du fait que $\sum_{i=1}^n p_i h(x_i, \alpha^*, \theta) = 1$. Le Lagrangien s'écrit donc

$$\mathbb{L}(p, \alpha^*, \theta) = \sum_{i=1}^n \log(p_i) + \sum_{j=1}^{n_1} (\alpha^* + x'_j \theta) - \mu \left[\sum_{i=1}^n p_i - 1 \right] - \lambda \sum_{i=1}^n p_i [h(x_i, \alpha^*, \theta) - 1],$$

et les conditions du premier ordre par rapport aux poids p_i et α^* sont données par :

$$\frac{\partial \mathbb{L}}{\partial p_i} = \frac{1}{p_i} - \mu - \lambda h(x_i, \alpha_0^*, \theta) = 0$$

$$\frac{\partial \mathbb{L}}{\partial \alpha^*} = n_1 - \lambda \sum_{i=1}^n p_i h(x_i, \alpha^*, \theta) = 0$$

La première condition implique : $n - \mu - \lambda = 0$, et on déduit que $\lambda = n_1$ à partir de la deuxième condition, ce qui donne $\mu = n_0$. On montre ainsi que :

$$\begin{aligned} p_i &= \frac{1}{n_0 + n_1 h(x_i, \alpha^*, \theta)} \\ &= \frac{1}{n_0} \cdot \frac{1}{1 + \rho h(x_i, \alpha^*, \theta)} \\ &= \frac{1}{n} \cdot \frac{1}{\frac{n_0}{n} + \frac{n_1}{n} h(x_i, \alpha^*, \theta)} \\ &= \frac{1}{n} \cdot \frac{1}{\pi_0^* + \pi_1^* h(x_i, \alpha^*, \theta)} \end{aligned}$$

où $\rho = \frac{n_1}{n_0}$, $\pi_0^* = \frac{n_0}{n}$ et $\pi_1^* = \frac{n_1}{n}$. π_0^* et π_1^* sont donc des estimateurs des probabilités marginales π_0 et π_1 . Les conditions du premier ordre par rapport à α^* et θ s'écrivent sous la forme :

$$\begin{aligned} \frac{\partial \mathbb{L}}{\partial \alpha^*} &= n_1 - \frac{\lambda}{n_0} \sum_{i=1}^n \frac{\exp(\alpha^* + z_i' \theta)}{1 + \rho \exp(\alpha^* + z_i' \theta)} \\ &= n_1 - \sum_{i=1}^n \frac{\rho \exp(\alpha^* + z_i' \theta)}{1 + \rho \exp(\alpha^* + z_i' \theta)} \\ &= \sum_{i=1}^{n_1} \frac{\partial \ln h}{\partial \alpha^*} - \sum_{i=1}^n \frac{\rho \exp(\alpha^* + z_i' \theta)}{1 + \rho \exp(\alpha^* + z_i' \theta)} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathbb{L}}{\partial \theta} &= \sum_{j=1}^{n_1} x_j - \frac{\lambda}{n_0} \sum_{i=1}^n \frac{z_i' \exp(\alpha^* + z_i' \theta)}{1 + \rho \exp(\alpha^* + z_i' \theta)} \\ &= \sum_{j=1}^{n_1} x_j - \sum_{i=1}^n \frac{z_i' \rho \exp(\alpha^* + z_i' \theta)}{1 + \rho \exp(\alpha^* + z_i' \theta)} \\ &= \sum_{j=1}^{n_1} \frac{\partial \ln h}{\partial \theta} - \sum_{i=1}^n \frac{z_i' \rho \exp(\alpha^* + z_i' \theta)}{1 + \rho \exp(\alpha^* + z_i' \theta)} \\ &= 0 \end{aligned}$$

3.2.2 propriétés asymptotiques

Sous certaines conditions de régularité, on peut montrer que les estimateurs sont convergents et asymptotiquement normaux. Notons $Q_1(\alpha^*, \theta) = \frac{1}{n} \frac{\partial \mathbb{L}}{\partial \alpha^*}$ et $Q_2(\alpha^*, \theta) = \frac{1}{n} \frac{\partial \mathbb{L}}{\partial \theta}$. Les dérivées partielles de Q_1 et Q_2 respectivement par rapport à α^* et θ s'écrivent :

$$\frac{\partial Q_1}{\partial \alpha^*} = -\frac{1}{n} \sum_{i=1}^n \frac{\rho \exp(\alpha^* + z'_i \theta)}{[1 + \rho \exp(\alpha^* + z'_i \theta)]^2} \rightarrow \mathbb{E}_{G_0} \left[\frac{\rho}{1 + \rho} \frac{\exp(\alpha^* + z' \theta)}{1 + \rho \exp(\alpha^* + z' \theta)} \right], n \rightarrow \infty,$$

$$\frac{\partial Q_1}{\partial \theta} = -\frac{1}{n} \sum_{i=1}^n \frac{\rho z'_i \exp(\alpha^* + z'_i \theta)}{[1 + \rho \exp(\alpha^* + z'_i \theta)]^2} \rightarrow \mathbb{E}_{G_0} \left[\frac{\rho}{1 + \rho} \frac{z' \exp(\alpha^* + z' \theta)}{1 + \rho \exp(\alpha^* + z' \theta)} \right], n \rightarrow \infty,$$

$$\frac{\partial Q_2}{\partial \alpha^*} = -\frac{1}{n} \sum_{i=1}^n \frac{\rho z'_i \exp(\alpha^* + z'_i \theta)}{[1 + \rho \exp(\alpha^* + z'_i \theta)]^2} \rightarrow \mathbb{E}_{G_0} \left[\frac{\rho}{1 + \rho} \frac{z' \exp(\alpha^* + z' \theta)}{1 + \rho \exp(\alpha^* + z' \theta)} \right], n \rightarrow \infty,$$

$$\frac{\partial Q_2}{\partial \theta} = -\frac{1}{n} \sum_{i=1}^n \frac{\rho z_i z'_i \exp(\alpha^* + z'_i \theta)}{[1 + \rho \exp(\alpha^* + z'_i \theta)]^2} \rightarrow \mathbb{E}_{G_0} \left[\frac{\rho}{1 + \rho} \frac{z z' \exp(\alpha^* + z' \theta)}{1 + \rho \exp(\alpha^* + z' \theta)} \right], n \rightarrow \infty,$$

où l'on a remplacé $\frac{n_1}{n}$ par $\frac{\rho}{1 + \rho}$, et où les tendances vers l'espérance mathématique résultent de *la loi des grands nombres* appliquée à chaque série de données. On obtient ensuite par *un développement de Taylor* :

$$\begin{aligned} 0 &= Q_1(\hat{\alpha}^*, \hat{\theta}) \\ &= Q_1(\alpha_0^*, \theta_0) + \frac{\partial Q_1(\alpha_0^*, \theta_0)}{\partial \alpha^*} (\hat{\alpha}^* - \alpha_0^*) + \frac{\partial Q_1(\alpha_0^*, \theta_0)}{\partial \theta} (\hat{\theta} - \theta_0) + o_p(|\hat{\alpha}^* - \alpha_0^*| + |\hat{\theta} - \theta_0|) \end{aligned}$$

$$\begin{aligned} 0 &= Q_2(\hat{\alpha}^*, \hat{\theta}) \\ &= Q_2(\alpha_0^*, \theta_0) + \frac{\partial Q_2(\alpha_0^*, \theta_0)}{\partial \alpha^*} (\hat{\alpha}^* - \alpha_0^*) + \frac{\partial Q_2(\alpha_0^*, \theta_0)}{\partial \theta} (\hat{\theta} - \theta_0) + o_p(|\hat{\alpha}^* - \alpha_0^*| + |\hat{\theta} - \theta_0|) \end{aligned}$$

où α_0^* et θ_0 sont les vraies valeurs des paramètres α^* et θ . Ceci permet d'écrire :

$$\begin{aligned} \begin{pmatrix} \hat{\alpha}^* - \alpha_0^* \\ \hat{\theta} - \theta_0 \end{pmatrix} &=^a - \begin{bmatrix} \frac{\partial Q_1}{\partial \alpha^*} & \frac{\partial Q_1}{\partial \theta} \\ \frac{\partial Q_2}{\partial \alpha^*} & \frac{\partial Q_2}{\partial \theta} \end{bmatrix}_{(\alpha_0^*, \theta_0)}^{-1} \begin{pmatrix} Q_1(\alpha_0^*, \theta_0) \\ Q_2(\alpha_0^*, \theta_0) \end{pmatrix} \\ \Rightarrow \begin{pmatrix} \hat{\alpha}^* - \alpha_0^* \\ \hat{\theta} - \theta_0 \end{pmatrix} &=^a - \frac{1 + \rho}{\rho} \begin{bmatrix} \Lambda & \Gamma \\ \Gamma & \Delta \end{bmatrix}^{-1} \begin{pmatrix} Q_1(\alpha_0^*, \theta_0) \\ Q_2(\theta^*, 0) \end{pmatrix} \end{aligned}$$

où “= *a*” signifie “*asymptotiquement égal à*”,

$$\Lambda = \mathbb{E} \left[\frac{\exp(\alpha^* + \theta'z)}{1 + \rho \exp(\alpha^* + \theta'z)} \right], \quad \Gamma = \mathbb{E} \left[\frac{z' \exp(\alpha^* + \theta'z)}{1 + \rho \exp(\alpha^* + \theta'z)} \right],$$

$$\Delta = \mathbb{E} \left[\frac{z z' \exp(\alpha^* + \theta'z)}{1 + \rho \exp(\alpha^* + \theta'z)} \right]$$

Sous réserve que $\begin{bmatrix} \Lambda & \Gamma \\ \Gamma & \Delta \end{bmatrix}^{-1}$ existe, on peut montrer que :

$$\sqrt{n}(\hat{\alpha}^* - \alpha_0^*) = \sqrt{n} \frac{(1 + \rho)}{\rho} [-(\Lambda - \Gamma' \Delta^{-1} \Gamma)^{-1} Q_1(\alpha_0^*, \theta_0) + \sqrt{n} (\Lambda - \Gamma' \Delta^{-1} \Gamma)^{-1} \Gamma' \Delta^{-1} Q_2(\alpha_0^*, \theta_0)]$$

et

$$\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n} \frac{(1 + \rho)}{\rho} [\Delta^{-1} \Gamma (\Lambda - \Gamma' \Delta^{-1} \Gamma)^{-1} Q_1(\alpha_0^*, \theta_0) - (\Delta^{-1} + \Delta^{-1} \Gamma (\Lambda - \Gamma' \Delta^{-1} \Gamma)^{-1} \Gamma' \Delta^{-1}) Q_2(\alpha_0^*, \theta_0)]$$

-loi asymptotique du vecteur $\begin{pmatrix} \sqrt{n} Q_1(\alpha_0^*, \theta_0) \\ \sqrt{n} Q_2(\alpha_0^*, \theta_0) \end{pmatrix}$:

On se trouve ici dans une situation particulière de **mélange de lois**. Le calcul de l'espérance ou la variance de toute fonction $\phi(t, \alpha^*, \theta)$ d'une observation prise dans la population totale peut être fait en fonction de l'espérance ou la variance des deux lois mélangées. Considérons la variable aléatoire T de la loi mélange. On a ainsi les résultats suivants :

$$\sum_{i=1}^n \phi(t_i, \alpha^*, \theta) = \sum_{i=1}^{n_0} \phi(x_i, \alpha^*, \theta) + \sum_{i=n_0+1}^n \phi(x_i, \alpha^*, \theta)$$

$$\sum_{i=1}^n p_i \cdot \phi(t_i, \alpha^*, \theta) = \sum_{i=1}^{n_0} p_i \cdot \phi(x_i, \alpha^*, \theta) + \sum_{i=n_0+1}^n p_i \phi(x_i, \alpha^*, \theta)$$

ce qui donne $\mathbb{E}_G(\phi(t, \alpha^*, \theta)) = \pi_0^* \cdot \mathbb{E}_{G_0}(\phi(x, \alpha^*, \theta)) + \pi_1^* \cdot \mathbb{E}_{G_1}(\phi(x, \alpha^*, \theta))$, où G représente la loi mélange, G_0 et G_1 sont les deux lois respectives des deux populations du mélange. La même chose est vraie pour le calcul de la variance.

Rappelons que les estimateurs de vraisemblance empirique des fonctions de répartition associées aux deux lois sont données par :

$$\hat{G}_0(x) = \sum_{k=1}^n \hat{p}_k \cdot I_{[x_k \leq x]} = \frac{1}{n_o} \sum_{k=1}^n \frac{1}{1 + \rho h(x_k, \hat{\alpha}^*, \hat{\theta})} I_{[x_k \leq x]}$$

$$\hat{G}_1(x) = \sum_{k=1}^n \hat{p}_k \cdot h(x_k, \hat{\alpha}^*, \hat{\theta}) \cdot I_{[x_k \leq x]} = \frac{1}{n_o} \sum_{k=1}^n \frac{h(x_k, \hat{\alpha}^*, \hat{\theta})}{1 + \rho h(x_k, \hat{\alpha}^*, \hat{\theta})} I_{[x_k \leq x]}$$

Les estimateurs non paramétriques des mêmes fonctions sont :

$$G_{n_0}(x) = \frac{1}{n_0} \sum_{k=1}^{n_0} I_{[x_k \leq x]} = \frac{1}{n_0} \sum_{k=1}^{n_0} I_{]-\infty, x_k]}(x)$$

$$G_{n_1}(x) = \frac{1}{n_1} \sum_{k=n_0+1}^n I_{[x_k \leq x]} = \frac{1}{n_1} \sum_{k=n_0+1}^n I_{]-\infty, x_k]}(x)$$

On pourra ainsi montrer (voir Hsini (2001)) que l'estimateur de vraisemblance empirique est plus efficace que l'estimateur non paramétrique.

On s'intéresse maintenant au calcul de matrice de variance-covariance du vecteur score

$$\sqrt{n} \begin{pmatrix} \sum_{i=1}^{n_1} \frac{\partial \ln h}{\partial \alpha^*} - \sum_{i=1}^n \frac{\rho \exp(\alpha^* + z'_i \theta)}{1 + \rho \exp(\alpha^* + z'_i \theta)} \\ \sum_{j=1}^{n_1} \frac{\partial \ln h}{\partial \theta} - \sum_{i=1}^n \frac{z'_i \rho \exp(\alpha^* + z'_i \theta)}{1 + \rho \exp(\alpha^* + z'_i \theta)} \end{pmatrix},$$

On note tout d'abord que ce vecteur est d'espérance nulle. On applique ensuite le *théorème central limite* pour montrer que ce vecteur tend asymptotiquement vers une loi normale de matrice de variance-covariance :

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

où

$$\Sigma_{11} = \rho^2 \pi_0^* \text{Var}_{G_0} \left(\frac{\exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)} \right) + \pi_1^* \text{Var}_{G_1} \left(\frac{\partial \ln h}{\partial \alpha^*} - \frac{\rho \exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)} \right)$$

$$\begin{aligned} \Sigma_{12} &= \pi_0^* \text{Cov}_{G_0} \left(-\frac{\rho \exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)}, -\frac{z' \rho \exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)} \right) \\ &\quad + \pi_1^* \text{Cov}_{G_1} \left(\frac{\partial \ln h}{\partial \alpha^*} - \frac{\rho \exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)}, \frac{\partial \ln h}{\partial \theta} - \frac{z' \rho \exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)} \right) \end{aligned}$$

$$\Sigma_{22} = \rho^2 \pi_0^* \text{Var}_{G_0} \left(\frac{z \exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)} \right) + \pi_1^* \text{Var}_{G_1} \left(\frac{\partial \ln h}{\partial \theta} - \frac{z \rho \exp(\alpha^* + z'\theta)}{1 + \rho \exp(\alpha^* + z'\theta)} \right)$$

On en déduit la loi asymptotique de $\begin{pmatrix} \sqrt{n}(\hat{\alpha}^* - \alpha_0^*) \\ \sqrt{n}(\hat{\theta} - \theta^*) \end{pmatrix}$, et l'on a :

$$\begin{pmatrix} \sqrt{n}(\hat{\alpha}^* - \alpha_0^*) \\ \sqrt{n}(\hat{\theta} - \theta^*) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Omega)$$

$$\text{où } \Omega = \left(\frac{1+\rho}{\rho} \right)^2 S \Sigma S' \text{ et } S = \begin{bmatrix} \Lambda & \Gamma \\ \Gamma & \Delta \end{bmatrix}^{-1}.$$

4 Etude empirique

Dans cette section on procède à deux types d'échantillonnage. Le premier consiste à conserver la totalité des observations, et le deuxième conserve aussi toutes les observations en appliquant un poids spécifique à chaque classe de façon à ramener les deux catégories d'observations en proportions égales. Les poids utilisés sont $\frac{n}{n_1}$ pour la classe 1 et $\frac{n}{n_0}$ pour la classe 2, où n est la taille de l'échantillon et n_1 et n_0 sont les tailles respectives des deux classes. L'échantillon ainsi construit contient deux fois la taille de l'échantillon initial.

Cette méthode a été aussi appliquée par Hsini (1999). *Son grand avantage est d'obtenir un échantillon suffisamment grand contenant autant d'observations dans la classe 1 que dans la classe 0, faisant ainsi disparaître la disproportion importante entre les deux classes considérées.* Cependant, on a dupliqué plusieurs fois chaque observation (surtout celles de la classe 1), ce qui signifie que l'on crée des observations fictives qui sont aussi de simples "copies" des observations initiales. La question reste donc posée sur l'efficacité de cette méthode.

Une méthode plus efficace pour rajouter des observations fictives serait d'étudier les caractéristiques de chacune des variables explicatives x_i au niveau des deux classes considérées et de lui faire construire un intervalle de confiance contenant ses différentes valeurs. Ainsi, au lieu de prendre les mêmes valeurs pour les observations fictives, on prend au hasard des valeurs dans l'intervalle de confiance correspondant. Les observations rajoutées auront les mêmes caractéristiques que les observations initiales sans être de simples copies.

4.1 Tests univariés

Le pouvoir discriminant des ratios considérés doit être maintenant testé pour décider du maintien ou pas de chacun d'entre eux. On fait ainsi, pour chaque ratio, un test simple d'égalité de la moyenne sur les deux classes de sociétés. Le niveau de risque toléré sera plus ou moins élevé en fonction des données. C'est ainsi qu'on a allé jusqu'au niveau 30% pour certains ratios et même au-delà pour l'année 98. En effet, pour cette année d'étude, peu de ratios étaient significatifs pour la distinction entre les deux états de sociétés et ceci à cause de la taille très réduite pour la classe déficitaire.

Il faut noter ici qu'on a décidé d'exclure certains ratios de l'étude pour gagner le plus d'observations appartenant à la classe déficitaire. En effet, ces ratios présentent plusieurs données manquantes et les observations concernées sont automatiquement exclues, ce qui nous fait perdre entre 2 et 10 observations dans la classe déficitaire et ceci coûte très cher car, comme on l'a souligné plus haut pour l'étude des événements rares, ce sont les observations de la classe déficitaire qui nous rapportent le plus d'informations sur le problème étudié.

Une autre méthode aurait consisté à remplacer les données manquantes par la moyenne par exemple, mais on n'a pas opté pour cette alternative partant du principe que ceci ne reflète pas forcément la réalité du marché d'assurance.

4.1.1 Tests sur l'échantillon non pondéré

Pour les données non pondérées et concernant la période correspondant au nouveau plan comptable (1995-1998) on note un pouvoir discriminant assez élevé du taux de sinistralité (TS) pour

les deux années d'étude 95 et 96. Le ratio CCVIE est retenue pour toute la période. D'autres ratios tels que le taux de cession des primes (TCEP), CAPIPU, CCDECE, FSIN, RSOL, RTECP, REXEP, PPLPL, RFPL, PREPPL ainsi que IMMPL avaient un pouvoir discriminant pour au moins deux années d'étude.

Pour la période correspondant à l'ancien plan comptable (1992-1994), plusieurs ratios ont été retenus pour l'année 1993 parmi lesquels trois ratios avaient un pouvoir discriminant très élevé, il s'agit du taux de charges (TCH), du "résultat d'exploitation/chiffre d'affaires" (REXPP) et du "résultat de l'exercice/chiffre d'affaires" (REXEP). On note d'autre part que le résultat financier et le "résultat d'exploitation/chiffre d'affaires" ont été retenus pour toute la période concernée. Parmi les autres ratios, plusieurs ont été retenus pour deux années d'observation.

variables discriminantes						
variable	données 92		données 93		données 94	
	F	Prob > F	F	Prob > F	F	Prob > F
TCP					1.8204	0.1802
TS	3.4261	0.0675				
TCOM					1.1634	0.2833
TCH	3.2750	0.0737	17.4391	0.0001		
CAPI	1.3794	0.2433	2.9885	0.0877		
AAFF			5.6621	0.0197		
CCOLD			1.0439	0.3100		
CCACAV			1.3064	0.2564		
RF	1.4919	0.2251	1.4655	0.2296	1.0955	0.2977
TCRF					1.9313	0.1676
TCF	5.3512	0.0230				
TCE			1.8642	0.1759	1.6199	0.2059
FSIN			1.9911	0.1621	0.8354	0.3628
FPR			5.5897	0.0205		
FRF			2.4897	0.1185		
RSOL			1.4243	0.2362	0.6252	0.4309
REXPP	1.0918	0.2989	15.3547	0.0002	0.2419	0.6239
REXEP			13.5123	0.0004	1.5715	0.2128
RFPL	3.6499	0.0593			1.8704	0.1744

Table 3: test de Fisher sur les données non pondérées

variables discriminantes								
	données 95		données 96		données 97		données 98	
variable	F	Prob > F	F	Prob > F	F	Prob > F	F	Prob > F
CON	2.5056	0.1161			1.7943	0.1828		
TS	8.0858	0.0053	7.6064	0.0068				
TCS			1.6209	0.2055				
TFAA			6.8510	0.0100				
TACH			5.8589	0.0170				
TCEP	3.4370	0.0662			1.3758	0.2430	0.8679	0.3534
CAPIPU			1.5666	0.2132	1.4497	0.2309		
CUCPP	2.0338	0.1565						
CCOLD					0.7806	0.3786	0.7739	0.3808
CCVIE	1.9004	0.1706	4.0936	0.0453	7.1570	0.0085	0.7750	0.3804
RF							0.6705	0.4145
TCRF					0.9589	0.3294		
TCF			7.8783	0.0059				
TCEN	2.4513	0.1201						
FSIN			1.0566	0.3061	1.0619	0.3048	0.6293	0.4291
FPR			1.2775	0.2607				
RSOL	2.0453	0.1553	2.3849	0.1252	2.6229	0.1079		
RTECP			2.8951	0.0915			0.6789	0.4116
REXEP			3.4623	0.0653			0.6166	0.4338
PPLPL	2.2141	0.1394			0.9925	0.3210	1.4725	0.2273
RFPL	3.2559	0.0737					2.2508	0.1362
PREPPL	3.0785	0.0819	2.7687	0.0988				
IMMPPL	5.5744	0.0199	1.9726	0.1628			1.0039	0.3184

Table 4: test de Fisher sur les données non pondérées (suite)

4.1.2 Tests sur l'échantillon pondéré

Une fois la symétrie établie entre les deux classes, on note une augmentation très significative du pouvoir discriminant des ratios retenus. Afin de ne pas garder trop de variables discriminantes, on baisse le niveau de risque en moyenne à 15%, mais malgré tout on obtient entre 9 et 15 variables. L'étape suivante permettra de supprimer certains ratios de l'étude.

Pour la période 92-94 (ancien plan comptable), on note la présence des variables résultat financier, RF, taux de couverture des engagements, et "résultat d'exploitation/chiffre d'affaires" sur toute la période. La plupart des autres ratios tels que TCH, CAPI, CCACAV, TCRF, FSIN, FPR, FRF, RSOL, REXEP, PPLPL et RFPL ont été retenus pour deux années d'observation.

Pour la période 95-98, les variables taux de conservation, taux de croissance des sinistres, taux de cession des primes, "contrats collectifs-vie/chiffre d'affaires", "fonds propres/charge des sinistres", ratio de solvabilité, "résultat technique/chiffre d'affaires", "résultat de l'exercice/chiffre d'affaires" et "immobiliers/produits des placements" ont été retenus pour au moins trois fois durant la période.

variables discriminantes						
	données 92		données 93		données 94	
variable	F	Prob > F	F	Prob > F	F	Prob > F
TCP			3.0173	0.0862	6.1722	0.0146
TS					2.1871	0.1422
TCS	4.4881	0.0373				
TCOM					2.1360	0.1469
TCH	7.8646	0.0063	15.0186	0.0002		
CAPI	3.0600	0.0841	5.4571	0.0220		
AAFF			4.2442	0.0426		
CCACAV	2.6484	0.1076	6.7706	0.0110		
CCDECE			4.3413	0.0404		
RF	8.1748	0.0054	7.4435	0.0078	4.4002	0.0384
TCRF	3.2246	0.0764			1.9714	0.1633
TCF	4.5565	0.0359				
TCE	2.2715	0.1358	7.1259	0.0092	6.3450	0.0133
FSIN			7.6015	0.0072	3.5790	0.0613
FPR			8.3570	0.0049	3.8898	0.0512
FRF	2.7169	0.1033	6.7468	0.0112		
RSOL			5.8568	0.0178	2.7275	0.1016
REXEP			15.6543	0.0002	4.8108	0.0305
REXPP	2.3241	0.1314	8.8835	0.0038	0.8965	0.3459
PPLPL			3.9146	0.0513	1.9406	0.1666
RFPL	14.3727	0.0003			1.7794	0.1851

Table 5: test de Fisher sur les données pondérées

variables discriminantes								
	données 95		données 96		données 97		données 98	
variable	F	Prob > F	F	Prob > F	F	Prob > F	F	Prob > F
CON	7.1780	0.0085	2.5417	0.1136	6.5810	0.0115		
TS	5.5385	0.0203	5.2233	0.0241				
TCS	1.9600	0.1642	4.7376	0.0315			2.0508	0.1547
TFAA			4.0879	0.0455				
TACH			3.7385	0.0556				
TCEP	9.0275	0.0033	5.1099	0.0255			6.5551	0.0117
CAPIPU			2.2131	0.1395	2.1009	0.1497		
CUCPU							2.7233	0.1015
DOCOI	2.9148	0.0905						
CCVIE	2.8525	0.0939	5.7248	0.0183	10.8049	0.0013	2.3166	0.1306
RF							5.0917	0.0259
TCRF					2.9684	0.0874		
TCF			6.3539	0.0131			3.8492	0.0521
FSIN	2.9176	0.0903	3.0758	0.0821	3.9636	0.0487	4.1930	0.0428
FPR							2.8426	0.0944
FRF	1.9290	0.1676						
RSOL	1.8292	0.1789	3.0032	0.0857	3.6068	0.0598		
RTECP			3.9681	0.0487	1.8790	0.1729	4.5751	0.0345
REXEP			4.0014	0.0478	1.9429	0.1658	4.0994	0.0451
PPLPL	2.1260	0.1475			2.0688	0.1528	4.7681	0.0309
RFPL	3.4788	0.0647					10.4875	0.0016
PREPPL	8.0617	0.0053	7.6153	0.0067				
IMPPPL	3.9853	0.0483	2.5270	0.1146			7.6820	0.0065

Table 6: test de Fisher sur les données pondérées (suite)

4.2 Traitement de la colinéarité

Une fois sélectionnées, les variables discriminantes doivent être soumises à d'autres tests statistiques permettant de détecter les phénomènes de colinéarité parfaite entre elles. Si un tel problème se manifeste, un choix s'impose pour décider des variables qu'il faut garder et de celles qu'il faut exclure de l'étude. Pour cette étape on applique deux méthodes qui se complètent. La première consiste à calculer la matrice des corrélations simples entre toutes les variables retenues, et la seconde consiste à calculer pour chaque variable retenue son coefficient de détermination par rapport aux autres variables et de voir si ce coefficient est proche de zéro (cas d'absence de corrélation parfaite) ou au contraire il est proche de 1 (cas de présence de corrélation parfaite).

Une autre méthode pour pallier ce problème de corrélation parfaite consisterait à faire une analyse en composantes principales (ACP) sur l'ensemble des variables discriminantes retenues pour chaque année d'observation. On obtient ainsi des composantes principales qui sont des combinaisons linéaires des variables initiales.

Le grand avantage de cette méthode est que les nouvelles variables obtenues (les composantes principales) ne sont pas corrélées entre elles, ce qui résoud parfaitement le problème de corrélation parfaite, mais un problème apparaîtrait pour le praticien lorsqu'il s'agira pour lui d'interpréter les modèles. En effet, tout ce qu'on pourrait expliquer est l'effet de chaque composante principale, ce qui n'aurait pas beaucoup de sens vu que ce qui nous intéresse est l'effet des ratios financiers et non pas leurs combinaisons. On n'a pas donc opté pour cette méthode.

On dresse ci-dessous les coefficients de détermination (ou corrélation multiple) obtenus pour chaque année d'observation. Les matrices de corrélations simples seront fournies en annexe.

4.2.1 Données non pondérées

variable	données 92		données 93		données 94	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
TCP					0.1151	0.0329
TS	0.1954	0.1392				
TCOM					0.1007	0.0173
TCH	0.4812	0.4450	0.5700	0.4974		
CAPI	0.1083	0.0461	0.1466	0.0024		
AAFF			0.5575	0.4828		
CCDECE			0.4112	0.3117		
CCACAV			0.1481	0.0041		
RF	0.0303	-0.0374	0.1310	-0.0159	0.0628	-0.0241
TCRF					0.8529	0.8393
TCF	0.1255	0.0645				
TCEN			0.4971	0.4121	0.6620	0.6307
FSIN			0.7170	0.6691	0.8530	0.8393
FPR			0.7295	0.6838		
FRF			0.6640	0.6072		
RSOL			0.3050	0.1875	0.9180	0.9104
REXEP			0.7875	0.7515	0.8481	0.8340
REXPP	0.4749	0.4383	0.7055	0.6557	0.8319	0.8162
RFPL	0.2130	0.1581			0.8653	0.8528

Table 7: coefficient de détermination sur les données non pondérées

Pour décider entre les variables, on commence d'abord par étudier simultanément la matrice des corrélations simples et les coefficients de détermination. Lorsque deux variables sont très

variable	données 95		données 96		données 97		données 98	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
CON	0.6507	0.6189			0.5203	0.4881		
TS	0.3732	0.3163	0.9509	0.9459				
TCS			0.2544	0.1778				
TFAA			0.9421	0.9361				
TACH			0.9751	0.9725				
TCEP	0.6370	0.6040			0.5186	0.4863	0.2066	0.1370
CAPIPU			0.3829	0.3194	0.0386	-0.0260		
CUCPU							0.2130	0.1440
CUCPP	0.3776	0.3210						
CCDECE					0.3081	0.2616	0.3488	0.2916
CCVIE	0.0763	-0.0077	0.0492	-0.0584	0.0374	-0.0273	0.1334	0.0574
RF							0.0703	-0.0113
TCRF					0.0489	-0.0151		
TCF			0.2472	0.1698				
TCE	0.6559	0.6246						
FSIN					0.2390	0.1878	0.6821	0.6542
RSOL	0.6504	0.6151	0.1694	0.0840	0.3412	0.2969		
RTECP			0.8967	0.8861			0.4443	0.3956
REXEP			0.9024	0.8924			0.6748	0.6463
PPLPL	0.23770	0.00307			0.0856	0.0242	0.8368	0.8225
RFPL	0.8749	0.8635					0.8359	0.8215
PREPPL	0.1958	0.1227	0.0725	-0.0229				
IMMPPL	0.2558	0.1807	0.0818	-0.0126			0.0286	-0.0567

Table 8: coefficient de détermination sur les données non pondérées (suite)

corrélées, on sait qu'il ne faut pas les garder toutes les deux en même temps, faute de quoi on obtiendrait un modèle instable. Le coefficient de détermination nous montre comment réagit chaque variable dans l'espace engendrée par les autres.

Ceci semble très intéressant pour notre étude du moment où le but est d'obtenir une régression sur toutes ces variables, et le fait d'observer les différentes interactions entre les variables ne peut qu'améliorer le modèle du moment où l'on exclut les variables qui interviennent dans le phénomène de corrélation parfaite.

La suppression d'une variable ne peut pas non plus être automatique ou simple à faire et doit tenir compte de plusieurs facteurs. Ainsi lorsqu'on veut choisir entre deux variables, on tient compte surtout du pouvoir discriminant de chacune d'entre elles (celle ayant un **pouvoir discriminant le plus élevé** a le plus de chance d'être retenue). Mais ceci est insuffisant et il faudra aussi étudier les interactions de chacune d'entre elles avec le reste des variables et de retenir **la moins corrélée** par rapport aux autres. Le dernier test consiste à faire deux versions de modèles contenant chacune l'une des deux variables en question. On retient ainsi le modèle qui apparait le mieux adapté (en considérant comme critère soit le niveau de risque toléré, soit les pourcentages des bien classées par exemple).

Commençons par la période 92-94, on note ainsi que pour l'année 92 trois variables ont été exclues. Il s'agit du taux de charges (en liaison avec l'ensemble des variables), capitalisation (puisque l'on choisit de garder le taux de sinistralité qui est plus discriminant) et "résultat de l'exercice/chiffre d'affaires" (en liaison avec le ratio "résultat financier/placements").

Pour l'année 93, on garde tout d'abord le taux de charges qui était le plus discriminant, on fait ensuite un choix entre les deux ratios "résultat de l'exercice/chiffre d'affaires" et le "résultat

données 92	données 93	données 94
		TCP
TS		
	TCH	
		TCOM
	CAPI	
RF		RF
		TCRF
TCF		
		TCE
	FPR	
	FRF	
	REXPP	
		REXEP
RFPL		

Table 9: variables présélectionnées sur les données non pondérées

d’exploitation/chiffre d’affaires” qui sont très corrélés entre eux (on a donc redondance de l’information) et on opte pour le premier qui est moins corrélé aux autres.

Pour cette année le choix n’était pas simple à faire vu le nombre élevé des variables présélectionnées et les fortes corrélations observées entre elles. On s’est donc basé sur les différents critères décrits ci-dessus pour faire un choix. C’est ainsi qu’ont été exclues les variables taux de couverture des engagements (corrélée avec “fonds propres/chiffre d’affaires”), autres affaires et contrats collectifs-décès (corrélées avec les variables retenues), “fonds propres/charge des sinistres” et “fonds propres/résultat financier” (corrélées avec “fonds propres/chiffre d’affaires” qui est plus discriminant), et enfin le ratio de solvabilité corrélé avec certaines variables retenues telles que “fonds propres/chiffre d’affaires”.

Pour l’année 94, on retient d’abord la variable la plus discriminante, taux de croissance du résultat financier et on exclut le ratio “résultat financier/placements”. On retient ensuite le taux de croissance du chiffre d’affaires et le taux de couverture des engagements, ce qui permet d’exclure certaines variables redondantes telles que le ratio “fonds propres/charge des sinistres”, le ratio de solvabilité et le “résultat d’exploitation/chiffre d’affaires”.

On étudie maintenant la période correspondant au nouveau plan comptable, 95-98. Cinq variables sur onze ont été exclues pour l’année 95. Le taux de conservation lié au taux de cession des primes, la branche contrats en unité de compte-PP liée au taux de sinistralité, le taux de couverture des engagements lié au ratio de solvabilité, et les deux ratios “produits des placements/placements” et “résultat financier/placements” liés au ratio “produits de réalisation/produits des placements”.

Pour l’année 96, le taux de sinistralité est retenu au dépend du taux d’accroissement des fonds, taux des frais d’acquisition et d’administration, et taux de charges. Le ratio “résultat technique/primes” est conservé au dépend du ratio “résultat de l’exercice/primes”, capitalisation-PU, ratio de solvabilité, taux de croissance des sinistres et du ratio “produits de réalisation/produits des placements”. On retient ensuite les deux variables contrats collectifs-vie et “immobiliers/produits des placements”.

Pour l’année 97, seulement trois variables ont été exclues sur neuf. Il s’agit tout d’abord du taux

données 95	données 96	données 97	données 98
		CON	
TS	TS		
TCEP			TCEP
		CAPIPU	
CCVIE	CCVIE	CCVIE	CCVIE
		TCRF	
			FSIN
RSOL		RSOL	
	RTECP		RTECP
		PPLPL	PPLPL
PREPPL			
IMPPPL	IMPPPL		

Table 10: variables présélectionnées sur les données non pondérées (suite)

des cessions qui est fortement corrélé au taux de conservations, le ratio “fonds propres/charge des sinistres” corrélé au ratio de solvabilité, et la branche contrats collectifs-décès corrélée entre autres au ratio de solvabilité et la branche contrats collectifs-vie.

Enfin, pour l’année 98, on opte pour le ratio “produits des placements/placements” au dépend du “résultat financier/placements”, le taux de cessions au dépend des variables contrats en unité de compte-PU, contrats collectifs-vie et résultat financier, les contrats collectifs-vie au dépend des contrats en unité de compte-PU, et enfin, les deux ratios “fonds propres/chiffre d’affaires” et “résultat technique/chiffre d’affaires” au dépend du “résultat de l’exercice/chiffre d’affaires”.

4.2.2 données pondérées

La même démarche est appliquée sur les données pondérées pour réduire le nombre de variables explicatives afin d’éviter le problème de corrélation parfaite et la redondance de l’information apportée par celles-ci.

variable	données 92		données 93		données 94	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
TCP			0.5021	0.4011	0.1668	0.0704
TS	0.5826	0.5170			0.2172	0.1266
TCS	0.2874	0.1754				
TCOM					0.3011	0.2202
TCH	0.7990	0.7674	0.8812	0.8571		
CAPI	0.4762	0.3939	0.3476	0.2153		
AAFF			0.8863	0.8632		
CCDECE			0.5164	0.4183		
CCACAV	0.1392	0.0039	0.2176	0.0588		
RF	0.1616	0.0298	0.1552	-0.0162	0.0770	-0.0298
TCRF	0.2690	0.1541			0.9560	0.9509
TCF	0.4369	0.3484				
TCE	0.5077	0.4303	0.5936	0.5112	0.6405	0.5989
FSIN			0.8363	0.8031	0.8936	0.8813
FPR			0.9313	0.9174	0.6612	0.6220
FRF	0.6582	0.6045	0.8800	0.8557		
RSOL			0.6023	0.5216	0.9104	0.9001
REXEP			0.9258	0.9108	0.6023	0.5562
REXP	0.7853	0.7515	0.9309	0.9169		
PPLPL			0.4063	0.2858	0.2642	0.1790
RFPL	0.4162	0.3245			0.9551	0.9499

Table 11: coefficient de détermination sur les données pondérées

Pour l’année 92, on retient tout d’abord le ratio “résultat financier/placements” ce qui per-

met d'exclure certaines variables telles que le taux de couverture des engagements et le taux de croissance des fonds. Le choix est porté ensuite sur les variables taux de sinistralité (on exclut ainsi le taux de charges, capitalisation, et résultat de l'exercice), résultat financier, taux de croissance des sinistres (ce qui permet d'exclure le ratio "fonds propres/résultat financier") et contrats collectifs-ACAV.

Le ratio "résultat d'exploitation/primes" est le plus discriminant sur les données 93. Le choix de ce ratio exclue le taux de charges, "capitalisation/chiffre d'affaires", "autres affaires/chiffre d'affaires", "fonds propres/résultat financier", le taux de couverture des engagements, le ratio de solvabilité et le "résultat de l'exercice/chiffre d'affaires" qui sont aussi très corrélés aux autres variables retenues dans le modèle. On retient aussi le taux d'accroissement du chiffre d'affaires, le résultat financier, le ratio "fonds propres/chiffre d'affaires" et le ratio "produits des placements/placements" en excluant les autres variables.

Les quatre premières variables les plus discriminantes sur les données 94 sont toutes retenues, et ceci permet d'exclure les variables taux de sinistralité, ratio de solvabilité, "fonds propres/sinistres", "fonds propres/primes" et taux de croissance du résultat financier. On retient ensuite le taux de commissions et le ratio "produits des placements/placements".

variable	données 95		données 96		données 97		données 98	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
CON	0.6192	0.5757	0.2789	0.1905	0.5596	0.5260		
TS	0.1039	0.0015	0.9738	0.9706				
TCS	0.3340	0.2579	0.2856	0.1980	0.5658	0.5327	0.2437	0.1612
TFAA			0.9840	0.9820				
TACH			0.9893	0.9879				
TCEP	0.5646	0.5148			0.5940	0.5587	0.2873	0.2095
CAIPU			0.6947	0.6573	0.4196	0.3753		
CUCPU							0.2983	0.2218
DOCOI	0.1209	0.0204						
CCVIE	0.1962	0.1044	0.1568	0.0534	0.1160	0.0486	0.2634	0.1830
RF							0.1522	0.0597
TCRF					0.0758	0.0053		
TCF			0.3499	0.2701			0.3276	0.2543
FSIN	0.4685	0.4077	0.4334	0.3639	0.2159	0.1560	0.5937	0.5494
FPR							0.6947	0.6614
FRF	0.7189	0.6867						
RSOL	0.5510	0.4996	0.4815	0.4179	0.1689	0.1055		
RTECP			0.9581	0.9530	0.9693	0.9669	0.4313	0.3693
REXEP			0.9604	0.9555	0.9712	0.9690	0.7835	0.7599
PPLPL	0.9332	0.9255			0.1976	0.1364	0.5740	0.5275
RFPL	0.9331	0.9254					0.5309	0.4797
PREPPL	0.4091	0.3416	0.2260	0.1311				
IMMPPL	0.5273	0.4732	0.4160	0.3444			0.0747	-0.0262

Table 12: coefficient de détermination sur les données pondérées (suite)

Sur les données 95, on retient d'abord le taux de cessions (ce qui exclut le taux de conservation), le taux de sinistralité, le taux de croissance des sinistres (et l'on exclut "produits des placements/placements"), les dommages corporels-individuels, les contrats collectifs-vie, le ratio "fonds propres/sinistres" (ce qui exclut "fonds propres/résultat financier" et ratio de solvabilité), "produits de réalisation/produits des placements" et "immobiliers/placements".

Pour les données 96, on retient les trois premières variables discriminantes et l'on exclut entre

autres le taux de sinistralité, le taux de frais d'acquisition et d'administration et le taux d'autres charges. On retient aussi le ratio de solvabilité (ce qui exclut les deux résultats technique et de l'exercice ainsi que "fonds propres/sinistres") et le taux de croissance des sinistres.

L'année 97, la variable la plus discriminante est celle représentant les contrats collectifs-vie. Le choix de cette variable implique l'exclusion du ratio de solvabilité. On retient ensuite le taux de conservations qui était très corrélée au taux de cessions. Les autres variables retenues sont capitalisation-PU (et l'on exclut "résultat technique/primes"), taux de croissance du résultat financier, "fonds propres/sinistres" (ce qui exclut "résultat de l'exercice/primes) et enfin le ratio "produits des placements/placements".

Cinq variables discriminantes ont été retenues pour l'année 98. La plus discriminante étant le ratio "résultat financier/placements" et son choix exclue d'autres variables telles que le ratio "produits des placements/placements", le taux de croissance des sinistres ou la branche contrats collectifs-vie qui est liée également au résultat financier retenu aussi dans le modèle. On retient d'autre part le taux de cessions, et les deux ratios "fonds propres/chiffre d'affaires" (qui exclut "fonds propres/sinistres" et "résultat de l'exercice/chiffre d'affaires") et "résultat technique/chiffre d'affaires".

données 92	données 93	données 94
	TCP	TCP
TS		
TCS		
		TCOM
CCACAV		
RF	RF	RF
		TCE
	FPR	
	REXPP	
		REXEP
	PPLPL	PPLPL
RFPL		

Table 13: variables présélectionnées sur les données pondérées

données 95	données 96	données 97	données 98
	CON	CON	
TS			
TCS	TCS		
TCEP			TCEP
DCIND			
		CAPIPU	
CCVIE	CCVIE	CCVIE	
			RF
		TCRF	
	TCF		
FSIN		FSIN	
			FPR
	RSOL		
			RTECP
		PPLPL	
			RFPL
PREPPL	PREPPL		
IMPPPL			

Table 14: variables présélectionnées sur les données pondérées (suite)

4.3 Modèles de la régression logistique

La régression logistique semble aujourd’hui la mieux adaptée à la modélisation des événements rares, et des études comparatives ont montré sa supériorité par rapport à l’analyse discriminante qui apparaît la mieux adaptée pour l’exploration des grands échantillons. D’autres études ont essayé aussi de comparer la régression logistique par rapport aux modèles de durée.

Les modèles adoptés ici sont du type logistique. Les variables présélectionnées pour chaque année d’observation sont introduites dans chaque modèle et les différentes méthodes de sélection des variables explicatives, stepwise, pas-à-pas ascendant et pas-à-pas descendant, sont toutes utilisées pour pouvoir étudier les différentes possibilités.

Comme pour les tests des variables discriminantes, les niveaux de risque tolérés pour l’entrée ou la sortie d’une variable du modèle sont plus ou moins élevés selon le type de données. Les deux statistiques du *rapport de vraisemblance* et du *score* ainsi que *les pourcentages des entreprises bien classées* pour différents *seuils de coupure* sont utilisées pour tester la bonne appropriation de chaque modèle.

4.3.1 Modèles sur les données non pondérées

On note tout d’abord un faible taux des “bien classées” concernant la classe déficitaire pour les différents modèles établis. Ceci provient évidemment de la disproportion importante entre les deux classes d’entreprises. Ainsi, malgré un pourcentage élevé des “bien classées” au niveau global, ce sont surtout les compagnies de la classe solvable qui étaient concernées par ce bon résultat. Au niveau de la classe déficitaire, pour un seuil de coupure égal à 50%, on atteint au meilleur des cas 22% de bien classées.

Deux modèles différents sont établis sur les données 92. Le premier fait intervenir les variables taux de sinistralité, taux de croissance des fonds propres (qui ont un effet positif sur le déficit), et le ratio “résultat financier/produit des placements” qui a un effet négatif sur le déficit. Le deuxième modèle exclue du modèle précédent la variable résultat financier qui a un effet négatif

sur le déficit.

Sur les données 93, on peut d’abord constater sur un premier modèle que le déficit est une fonction croissante des deux variables taux de charges et capitalisation, et qu’il est fonction décroissante du ratio “fonds propres/résultat financier”. Un deuxième modèle rajoute au premier les deux ratios “fonds propres/primes” et “résultat de l’exercice/primes” qui ont tous deux un effet négatif sur le déficit. Le troisième modèle enlève du second la variable “fonds propres/résultat financier” qui n’était pas très significative.

Deux modèles sont aussi établis sur les données 94. Le premier fait intervenir les variables taux de croissance des primes, taux de couverture des engagements et “résultat de l’exercice/primes” qui ont toutes un effet négatif sur le déficit. Le deuxième modèle rajoute au précédent la variable taux de croissance du résultat financier qui a aussi un effet négatif sur le déficit.

variable	pas à pas descendant et stepwise			pas à pas ascendant		
	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-2.6207	0.4179	0.0001	-3.2520	0.9227	0.0004
TS	0.9496	0.4296	0.0271	0.8597	0.4253	0.0433
RF				-2.3422	2.2634	0.3008
TCF	0.3632	0.2501	0.1464	0.3537	0.2495	0.1563
RFPL	-0.7728	0.5358	0.1492	-0.5872	0.5152	0.2543
goodness of fit						
	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	11.823 (3 ddl)		0.0080	14.603 (4 ddl)		0.0056
Score	17.050 (3 ddl)		0.0007	18.418 (4 ddl)		0.0010
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.160	85.0	20.0	91.8	85.0	40.0	89.7
0.200	87.9	20.0	94.8	85.0	10.0	92.8
0.300	88.8	10.0	96.9	87.9	10.0	95.9
0.400	90.7	10.0	99.0	90.7	10.0	99.0
0.500	90.7	10.0	99.0	90.7	10.0	99.0
0.600	90.7	10.0	99.0	90.7	10.0	99.0
0.700	90.7	10.0	99.0	90.7	10.0	99.0

Table 15: modèles sur les données non pondérées 92

	stepwise			pas à pas ascendant			pas à pas descendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-3.2621	0.6494	0.0001	-3.2769	0.6634	0.0001	-3.0568	0.5677	0.0001
TCH	1.3322	0.4705	0.0046	1.5548	0.6771	0.0217	1.5376	0.6725	0.0222
CAPI	0.6008	0.2761	0.0295	0.6668	0.2914	0.0221	0.6823	0.2861	0.0171
FPR				-1.3050	0.8308	0.1162	-1.7492	0.7145	0.0144
FRF	-2.0284	1.0506	0.0535	-1.0244	1.1500	0.3730			
REXEP				-0.8361	0.5996	0.1632	-0.9385	0.5893	0.1112
goodness of fit	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	17.593 (3 ddl)		0.0005	20.790 (5 ddl)		0.0009	19.917 (4 ddl)		0.0005
Score	21.441 (3 ddl)		0.0001	24.236 (5 ddl)		0.0002	24.146 (4 ddl)		0.0001
pourcentage des bien classées									
Cutoff Pr	Total	class1	class2	Total	class1	class2	Total	class1	class2
0.160	87.5	55.6	90.8	88.5	44.4	93.1	88.5	44.4	93.1
0.200	90.6	55.6	94.3	88.5	44.4	93.1	89.6	44.4	94.3
0.300	89.6	33.3	95.4	89.6	44.4	94.3	89.6	44.4	94.3
0.400	88.5	22.2	95.4	89.6	33.3	95.4	89.6	33.3	95.4
0.500	90.6	22.2	97.7	90.6	33.3	96.6	90.6	33.3	96.6
0.600	90.6	22.2	97.7	88.5	11.1	96.6	88.5	11.1	96.6
0.700	89.6	11.1	97.7	88.5	11.1	96.6	88.5	11.1	96.6

Table 16: modèles sur les données non pondérées 93

	pas à pas descendant et stepwise			pas à pas ascendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-2.3322	0.3691	0.0001	-2.3438	0.3723	0.0001
TCP	-0.5290	0.4323	0.2211	-0.5690	0.4452	0.2012
TCRF				-0.1823	0.2789	0.5134
TCE	-0.7367	0.3824	0.0540	-0.7022	0.3839	0.0674
REXEP	-0.5699	0.3118	0.0675	-0.4644	0.3531	0.1884
goodness of fit	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	7.331 (3 ddl)		0.0621	7.754 (4 ddl)		0.1010
Score	6.605 (3 ddl)		0.0856	7.208 (4 ddl)		0.1253
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.100	57.0	58.3	56.8	53.3	50.0	53.7
0.140	81.3	50.0	85.3	81.3	41.7	86.3
0.160	83.2	41.7	88.4	86.0	41.7	91.6
0.180	86.0	33.3	92.6	83.2	8.3	92.6
0.200	85.0	25.0	92.6	83.2	0.0	93.7
0.300	83.2	0.0	93.7	84.1	0.0	94.7

Table 17: modèles sur les données non pondérées 94

Sur les données 95, on a aussi deux modèles. Le premier fait intervenir les variables taux de sinistralité, contrats collectifs-vie, ratio de solvabilité, “immobiliers/produits des placements” (qui ont un effet positif sur le déficit), et le taux de croissance des primes qui a un effet négatif sur le déficit. Le deuxième modèle rajoute au premier le ratio “produits de réalisation/produits des placements” qui a un effet négatif sur le déficit.

Un seul modèle est établi sur les données 96. Il montre que le déficit est une fonction croissante des variables taux de sinistralité, contrats collectifs-vie, “résultat technique/chiffre d’affaires” et “immobiliers/produits des placements”.

On établit ensuite trois modèles sur les données 97. Dans un premier modèle on voit que le déficit est une fonction croissante des variables taux de conservation, capitalisation PU, contrats collectifs-vie et ratio de solvabilité. Le deuxième modèle rajoute au premier le ratio “produits des placements/placements” qui intervient aussi positivement sur le déficit. Enfin le troisième modèle rajoute au premier la variable taux de croissance du résultat financier qui intervient négativement sur le déficit.

Trois versions de modèles sont établis sur les données 98. Un premier modèle montre que le déficit est une fonction croissante des deux ratios CCVIE et RTECP. Le deuxième modèle rajoute au premier le taux de cession des primes et le ratio “fonds propres/charge des sinistres” qui interviennent négativement sur le déficit, et le ratio “produits des placements/placements” qui intervient positivement sur le déficit. Le troisième modèle exclut du second la variable TCEP.

variable	pas à pas descendant et stepwise			pas à pas ascendant		
	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-2.6328	0.5186	0.0001	-2.5992	0.5088	0.0001
TS	1.0300	0.4950	0.0375	0.9627	0.4893	0.0491
TCEP	-2.3861	1.1344	0.0354	-2.2400	1.1348	0.0484
CCVIE	0.4393	0.2356	0.0623	0.3801	0.2382	0.1106
RSOL	0.4266	0.2620	0.1034	0.3969	0.2629	0.1311
PREPPL				-0.4042	0.3899	0.2999
IMPPPL	0.5457	0.3703	0.1406	0.5952	0.4503	0.1862
goodness of fit	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	22.713 (5 ddl)		0.0004	24.012 (6 ddl)		0.0005
Score	20.569 (5 ddl)		0.0010	22.368 (6 ddl)		0.0010
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.100	60.0	77.8	56.9	60.8	77.8	57.8
0.200	71.7	33.3	78.4	78.3	44.4	84.3
0.300	80.0	11.1	92.2	82.5	22.2	93.1
0.400	83.3	11.1	96.1	82.5	11.1	95.1
0.500	85.8	11.1	99.0	84.2	11.1	97.1
0.600	86.7	11.1	100.0	86.7	11.1	100.0
0.700	86.7	11.1	100.0	86.7	11.1	100.0

Table 18: modèles sur les données non pondérées 95

les trois méthodes de sélection			
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-2.0243	0.3078	0.0001
TS	1.3392	0.5718	0.0192
CCVIE	0.6197	0.2242	0.0057
RTECP	0.6728	0.3163	0.0334
IMMPPL	0.3464	0.1991	0.0819
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	17.431 (4 ddl)		0.0016
Score	19.375 (4 ddl)		0.0007
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.100	62.9	72.2	61.3
0.200	75.8	38.9	82.1
0.300	84.7	38.9	92.5
0.400	84.7	16.7	96.2
0.500	85.5	11.1	98.1
0.600	85.5	11.1	98.1
0.700	84.7	5.6	98.1

Table 19: modèle sur les données non pondérées 96

variable	stepwise			pas à pas ascendant			pas à pas descendant		
	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-2.3311	0.3818	0.0001	-2.2496	0.3850	0.0001	-2.4320	0.4033	0.0001
CON	0.9190	0.6873	0.1812	0.9116	0.6826	0.1817	0.8301	0.6809	0.2228
CAPIPU	0.2645	0.2021	0.1907	0.2526	0.2038	0.2152	0.2783	0.2008	0.1658
CCVIE	0.5906	0.2243	0.0084	0.6327	0.2296	0.0059	0.6417	0.2301	0.0053
TCRF							-0.7887	0.6078	0.1944
RSOL	0.5137	0.2176	0.0182	0.5020	0.2185	0.0216	0.5219	0.2284	0.0223
PPLPL				1.0070	0.8807	0.2529			
goodness of fit									
	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	13.554 (4 ddl)		0.0089	14.696 (5 ddl)		0.0117	15.533 (5 ddl)		0.0083
Score	15.542 (4 ddl)		0.0037	16.719 (5 ddl)		0.0051	16.327 (5 ddl)		0.0060
pourcentage des bien classées									
Cutoff Pr	Total	class1	class2	Total	class1	class2	Total	class1	class2
0.100	70.1	56.3	72.1	70.9	62.5	72.1	70.9	56.3	73.0
0.140	81.1	37.5	87.4	80.3	43.8	85.6	80.3	37.5	86.5
0.200	82.7	37.5	89.	81.1	37.5	87.4	81.1	31.3	88.3
0.300	83.5	31.3	91.0	84.3	37.5	91.0	83.5	31.3	91.0
0.400	81.1	0.0	92.8	81.1	6.3	91.9	85.0	25.0	93.7
0.500	85.8	0.0	98.2	85.8	0.0	98.2	83.5	0.0	95.5

Table 20: modèles sur les données non pondérées 97

	stepwise			pas à pas ascendant			pas à pas descendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-2.8632	0.4295	0.000	-5.2339	1.6074	0.0011	-4.9936	1.5216	0.0010
TCEP				-0.7987	0.9916	0.4206			
CCVIE	0.4040	0.3143	0.1986	0.4714	0.3285	0.1513	0.5073	0.3324	0.1269
FS				-8.9728	5.2592	0.0880	-8.5516	5.0498	0.0904
RTECP	0.8159	0.6482	0.2081	1.0310	0.8104	0.2033	1.1977	0.8063	0.1374
PPLPL				0.4907	0.3744	0.1900	0.4277	0.3539	0.2268
goodness of fit	χ^2 for Cov	P-Value	χ^2 for Cov	P-Value	χ^2 for Cov	P-Value	χ^2 for Cov	P-Value	
-2 Log L	2.531 (2 ddl)	0.2821	9.501 (5 ddl)	0.0907	8.361 (4 ddl)	0.0792			
Score	1.878 (2 ddl)	0.3909	3.358 (5 ddl)	0.6450	2.629 (4 ddl)	.6218			
pourcentage des bien classées									
Cutoff Pr	Total	class1	class2	Total	class1	class2	Total	class1	class2
0.060	69.4	50.0	70.7	62.1	75.0	61.2	9.7	75.0	58.6
0.080	76.6	12.5	81.0	74.2	62.5	75.0	67.7	50.0	69.0
0.100	78.2	0.0	83.6	77.4	37.5	80.2	78.2	25.0	81.9
0.120	83.1	0.0	88.8	83.1	25.0	87.1	84.7	25.0	88.8
0.140	87.1	0.0	93.1	83.1	12.5	87.9	85.5	12.5	90.5
0.200	91.9	0.0	98.3	87.9	12.5	93.1	89.5	12.5	94.8
0.220	91.9	0.0	98.3	88.7	12.5	94.0	89.5	0.0	95.7

Table 21: modèles sur les données non pondérées 98

4.3.2 modèles sur les données pondérées

Pour améliorer les modèles, la seule solution est de rétablir l'équilibre entre les deux groupes d'observations. La première méthode utilisée consiste à surpondérer les observations de la classe déficitaire dont la taille était très réduite. Cette méthode peut être critiquée et l'on peut se poser trop de questions sur son efficacité du moment où l'on se base sur des observations fictives et sur une symétrie en réalité inexistante entre les deux classes. C'est ainsi qu'on proposera en dernière section une première correction de cette méthode.

Les résultats obtenus concernant les données pondérées montrent une nette amélioration des différents modèles. Si l'on se restreint aux pourcentages des bien classées, on constate que pour un seuil de coupure de 50%, on atteint dans certains cas jusqu'à 60% au niveau de la classe déficitaire.

Le premier modèle est établi sur les données 92. Il montre que le déficit est d'une part fonction croissante du taux de sinistralité et d'autre part fonction décroissante des variables taux de croissance des primes, contrats collectifs-ACAV, résultat financier et "résultat financier/placements".

Un seul modèle est aussi établi sur les données 93 et montre que le déficit est fonction décroissante de toutes les variables retenues : taux de croissance du chiffre d'affaires, résultat financier, "fonds propres/chiffres d'affaires" et "résultat d'exploitation/chiffre d'affaires".

Même chose pour les données 94 où le modèle montre que le déficit est fonction décroissante des variables taux de croissance des primes, résultat financier, taux de couverture des engagements et "résultat de l'exercice/primes". Le ratio "produits des placements/placements" intervient positivement sur le déficit.

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-2.9435	0.7566	0.0001
TS	0.7591	0.2516	0.0026
TCS	-3.0833	1.1265	0.0062
CCACAV	-0.9117	0.4350	0.0361
RF	-5.0410	1.7058	0.0031
RFPL	-1.3218	0.3744	0.0004
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	95.633 (5 ddl)		0.0001
Score	51.762 (5 ddl)		0.0001
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	58.5	80.0	56.0
0.400	62.8	70.0	61.9
0.500	68.1	50.0	70.2
0.600	79.8	50.0	83.3
0.700	84.0	30.0	90.5

Table 22: modèle sur les données pondérées 92

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-1.3256	0.3480	0.0001
TCP	-0.7462	0.3710	0.0443
RFIN	-2.6119	0.8222	0.0015
FPR	-1.2135	0.2961	0.0001
REXPP	-1.7843	0.3059	0.0001
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	102.724 (4 ddl)		0.0001
Score	55.135 (4 ddl)		0.0001
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	57.1	83.3	54.0
0.400	67.0	66.7	67.0
0.500	73.2	58.3	75.0
0.600	79.5	50.0	83.0
0.700	84.8	41.7	90.0

Table 23: modèle sur les données pondérées 93

les 3 méthodes de sélection			
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-0.5842	0.1893	0.0020
TCP	-0.4985	0.2146	0.0202
RF	-0.6463	0.3151	0.0402
TCE	-1.1432	0.3038	0.0002
REXEP	-0.8208	0.2224	0.0002
PPLPL	0.4934	0.2351	0.0358
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	54.391 (5 ddl)		0.0001
Score	43.113 (5 ddl)		0.0001
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	34.9	83.3	28.9
0.400	56.0	83.3	52.6
0.500	68.8	50.0	71.1
0.600	75.2	33.3	80.4
0.700	81.7	16.7	89.7

Table 24: modèle sur les données pondérées 94

Sur les données 95, un premier modèle montre que le déficit est fonction croissante des variables taux de sinistralité, contrats collectifs-vie et “immobilier/produits des placements”, et qu’il est fonction décroissante des variables taux de croissance des primes et “produits de réalisation/produits des placements”. Le deuxième modèle rajoute au premier les deux variables dommages corporels individuels et “fonds propres/charge des sinistres” qui interviennent négativement sur le déficit.

Le seul modèle établi sur les données 96 montre que le déficit est fonction croissante des variables taux de conservation, contrats collectifs-vie, taux de croissance du résultat financier et ratio de solvabilité. Ce modèle montre aussi que le déficit est fonction décroissante des variables taux de croissance des sinistres et “produits de réalisation/produits des placements”.

Le modèle sur les données 97 montre que le déficit est fonction croissante du taux de conservation et fonction décroissante du ratio “fonds propres/ charge des sinistres”.

On établit enfin un seul modèle sur les données 98. Ce modèle montre que les variables taux de croissance des primes, résultat financier et “fonds propres /primes” interviennent négativement sur le déficit et que le ratio “résultat technique/primes” intervient positivement.

	pas à pas descendant et stepwise			pas à pas ascendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-0.6824	0.2186	0.0018	-0.7958	0.2339	0.0007
TS	0.9621	0.2989	0.0013	0.9494	0.3103	0.0022
TCEP	-1.6975	0.5140	0.0010	-1.7446	0.5269	0.0009
DCIND				-0.4604	0.2827	0.1035
CCVIE	0.3404	0.1320	0.0099	0.3121	0.1350	0.0208
FSIN				-0.5182	0.3033	0.0875
PREPPL	-0.4791	0.2105	0.0229	-0.4018	0.2166	0.0635
IMMPPL	0.4270	0.2256	0.0584	0.4043	0.2365	0.0873
goodness of fit						
	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	69.421 (5 ddl)		0.0001	75.945 (7 ddl)		0.0001
Score	47.155 (5 ddl)		0.0001	52.232 (7 ddl)		0.0001
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.300	41.5	88.9	33.0	40.7	83.3	33.0
0.400	55.1	77.8	51.0	52.5	72.2	49.0
0.500	72.0	61.1	74.0	71.2	61.1	73.0
0.600	78.8	38.9	86.0	78.0	44.4	84.0
0.700	80.5	22.2	91.0	79.7	27.8	89.0

Table 25: modèles sur les données pondérées 95

les 3 méthodes de sélection			
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-0.7261	0.2094	0.0005
CON	0.7124	0.3120	0.0224
TCS	-1.1229	0.3668	0.0022
CCVIE	0.4224	0.1397	0.0025
TCF	0.7268	0.1828	0.0001
RSOL	0.4292	0.1594	0.0071
PREPPL	-0.8246	0.2627	0.0017
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	77.068 (6 ddl)		0.0001
Score	60.680 (6 ddl)		0.0001
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	46.2	83.3	39.6
0.400	54.6	77.8	50.5
0.500	70.6	55.6	73.3
0.600	79.8	44.4	86.1
0.700	83.2	38.9	91.1

Table 26: modèle sur les données pondérées 96

les 3 méthodes de sélection			
variable	Pa-Est	St-Er	Pr $> \chi^2$
CONST	-0.5524	0.2010	0.0060
CON	0.5734	0.2608	0.0279
CAPIPU	0.2384	0.1124	0.0340
CCVIE	0.5390	0.1393	0.0001
TCRF	-0.6343	0.2788	0.0229
FSIN	-1.3849	0.5757	0.0162
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	52.523 (5 ddl)		0.0001
Score	43.678 (5 ddl)		0.0001
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	32.0	81.3	25.0
0.400	39.1	62.5	35.7
0.500	66.4	37.5	70.5
0.600	81.3	25.0	89.3
0.700	83.6	25.0	92.0

Table 27: modèle sur les données pondérées 97

les trois méthodes de sélection			
variable	Pa-Est	St-Er	Pr $> \chi^2$
CONST	-1.5849	0.4280	0.0002
TCEP	-1.3559	0.4097	0.0009
RF	-1.4426	0.4294	0.0008
FPR	-4.8000	1.6667	0.0040
RTECP	0.8113	0.3589	0.0238
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	62.999 (4 ddl)		0.0001
Score	39.002 (4 ddl)		0.0001
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	39.7	87.5	36.4
0.400	49.2	87.5	46.6
0.500	54.0	62.5	53.4
0.600	65.1	25.0	67.8
0.700	81.0	12.5	85.6

Table 28: modèle sur les données pondérées 98

4.4 Données stratifiées

On procède maintenant à la constitution d'un échantillon selon la méthode "choice-based sample". On retient d'abord toutes les observations appartenant à la classe déficitaire, et on tire **au hasard** le même nombre d'observations dans la classe solvable. L'échantillon obtenu contient ainsi autant d'observations de la classe déficitaire que de la classe solvable, mais le nombre total d'observation reste faible.

4.4.1 Tests univariés

Pour détecter les variables les plus discriminantes à introduire ultérieurement dans les différents modèles, on procède comme précédemment par un test de comparaison de la moyenne de chaque variable sur les deux classes, mais vue la taille très réduite de l'échantillon retenu, on procède maintenant par une méthode non paramétrique basée sur *le test de Wilcoxon*. Les résultats obtenus sont dressés ci-dessous pour chaque année d'observation. On note le nombre plus ou moins élevé des variables retenues en premier lieu comme discriminantes. Le niveau de signification est limité en moyenne à 20%.

Pour la période couvrant l'ancien plan comptable, on note la présence des ratios liées aux charges et commissions, les contrats collectifs ACAV, le résultat financier, le taux de couverture des engagements ainsi que les résultats d'exploitation et de l'exercice.

variables discriminantes				
variable	Z-Wil	Prob > Z	T-Stud	Prob > T
TCH	1.64163	0.10067	1.7856	0.0893
CAPI	-1.19940	0.23037	0.3716	0.7157
CCACAV	-1.49488	0.13495	-1.0233	0.3303
RF	-1.83862	0.06597	-1.4749	0.1558
TCF	-1.51398	0.13003	0.4846	0.6335
TCE	-2.16695	0.03024	-2.0212	0.0569
FSIN	-1.23231	0.21783	-0.9551	0.3529
REXPP	-1.96995	0.04884	-1.3966	0.1779

Table 29: tests de Wilcoxon et Student sur les données 92

variables discriminantes				
variable	Z-Wil	Prob > Z	T-Stud	Prob > T
TCOM	1.47224	0.14096	1.5271	0.1410
TCH	1.29904	0.19393	1.5564	0.1339
CCACAV	-2.18345	0.02900	-1.7318	0.0980
CCOMPLEM	1.42540	0.15404	1.4480	0.1617
RF	-1.58771	0.11235	-1.8529	0.0774
TCE	-2.33827	0.01937	-1.3752	0.1829
REXPP	-2.80015	0.00511	-2.8228	0.0099
REXEP	-1.29904	0.19393	-1.4945	0.1492

Table 30: tests de Wilcoxon et Student sur les données 93

Pour la période 95-98, plusieurs indicateurs sont retenus. On note la présence du taux de conservation, du taux de sinistralité, du taux de cession des primes, du résultat financier et son taux d'accroissement, des branches contrats collectifs décès et vie, capitalistaion-PU, dommages corporels individuels (97), contrats en unité de compte-PP (98). On retient aussi les indicateurs liés aux placements tels que PREPPL, IMPPPL et AUTPPL.

variables discriminantes				
variable	Z-Wil	Prob > Z	T-Stud	Prob > T
TCP	-1.76092	0.07825	-1.4550	0.1602
TCEP	-1.30017	0.19354	0.4034	0.6906
CCOMP	1.25670	0.20886	0.7604	0.4551
RF	-1.93412	0.05310	-1.9164	0.0684
TCE	-1.81865	0.06896	-1.4793	0.1532
FPR	-1.35677	0.17485	-1.2612	0.2207
FRF	-1.52998	0.12602	-1.5961	0.1257
REXPP	-1.64545	0.09988	-1.7479	0.0957
REXEP	-1.41451	0.15721	-2.3101	0.0306

Table 31: tests de Wilcoxon et Student sur les données 94

variables discriminantes				
variable	Z-Wil	Prob > Z	T-Stud	Prob > T
TCP	-2.13561	0.03271	-2.5686	0.0148
CON	1.95078	0.05108	2.3341	0.0256
TS	1.69267	0.09052	1.4130	0.1668
TCCP	-2.60723	0.00913	-1.9257	0.0690
TCEP	-2.64031	0.00828	-2.1056	0.0427
CAPIPU	2.10230	0.03553	1.4447	0.1577
FSIN	-1.62939	0.10323	-0.9139	0.3672
PREPPL	-1.61367	0.10660	-1.7466	0.0897

Table 32: tests de Wilcoxon et Student sur les données 95

variables discriminantes				
variable	Z-Wil	Prob > Z	T-Stud	Prob > T
CON	2.13200	0.03301	1.7320	0.0929
TS	1.61885	0.10548	1.8175	0.0785
TCEP	-1.58866	0.11214	-1.6894	0.1009
CAPIPU	1.35578	0.17517	1.2568	0.2179
CCDECE	-1.20357	0.22876	-0.3009	0.7655
RF	-1.30885	0.19058	-0.5220	0.6054
TCRF	-1.96328	0.04961	-1.4871	0.1468
IMMPPL	-1.24330	0.21376	1.1595	0.2548

Table 33: tests de Wilcoxon et Student sur les données 96

variables discriminantes				
variable	Z-Wil	Prob > Z	T-Stud	Prob > T
CON	1.83737	0.06616	1.0897	0.2845
TCEP	-1.79599	0.07250	-1.5305	0.1364
DCIND	-2.00919	0.04452	-1.1437	0.2618
CCDECE	-1.48132	0.13852	-0.2356	0.8153
CCVIE	1.52921	0.12621	2.5243	0.0171
FSIN	1.41333	0.15756	1.7112	0.0974
IMMPPL	-1.91313	0.05573	-1.4296	0.1632
AUTPPL	1.56409	0.11780	0.9902	0.3302

Table 34: tests de Wilcoxon et Student sur les données 97

variables discriminantes				
variable	Z-Wil	Prob > Z	T-Stud	Prob > T
TCCP	-1.74284	0.08136	-1.0369	0.3203
TCEP	-2.25961	0.02385	-2.7291	0.0163
CUCPP	2.13852	0.03247	1.3372	0.2230
TCF	-1.41778	0.15625	-1.4205	0.1773
REXEP	-1.41778	0.15625		
RFPL	1.41778	0.15625	1.6505	0.1211
PARTPPL	-1.62783	0.10356	-2.0505	0.0639

Table 35: tests de Wilcoxon et Student sur les données 98

4.4.2 Traitement de la colinéarité

Toujours comme pour les autres types de données (pondérées ou pas), on applique les mêmes principes pour traiter le problème de colinéarité parfaite entre les variables retenues par les tests univariés.

variable	données 92		données 93		données 94	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
TCP					0.2494	-0.1259
TCOM	0.1717	-0.1906				
TCH	0.6557	0.4549	0.7319	0.6147		
TCEP					0.3642	0.0463
CAPI	0.2554	-0.1789				
CAPACAV	0.1954	-0.2739				
CCACAV			0.2674	-0.0532		
CCOMP			0.1955	-0.1564	0.3694	0.0541
RF	0.0951	-0.4327	0.2804	-0.0345	0.1648	-0.2528
TCF	0.1920	-0.2793				
TCE	0.3020	-0.1051	0.3902	0.1235	0.3442	0.0164
FSIN	0.1977	-0.2704				
FPR					0.5742	0.3613
FRF					0.7162	0.5743
REXEP			0.7684	0.6671	0.7371	0.6056
REXPP	0.5813	0.3370	0.7967	0.7078	0.7709	0.6563

Table 36: coefficient de détermination sur les données stratifiées

Concernant la période 92-94, seulement trois variables ont été retenues pour l'année 92. D'abord la variable au pouvoir discriminant le plus élevé, le taux de couverture des engagements. On exclut ainsi le taux de charges et le ratio "résultat d'exploitation/chiffre d'affaires". On choisit ensuite le résultat financier et le taux d'accroissement des fonds, ce qui exclut le ratio "fonds propres/charge des sinistres" et la branche capitalisation.

Sur les données 93, on retient les deux premières variables les plus discriminantes en excluant le ratio "résultat de l'exercice/chiffre d'affaires", le taux de charges, la branche contrats collectifs-ACAV et le résultat financier. On retient aussi le taux de commissions et la branche contrats complémentaires.

Enfin pour l'année 94, on retient le résultat financier, le taux d'accroissement du chiffre d'affaires et les deux ratios "fonds propres/chiffre d'affaires" et "résultat d'exploitation/chiffre d'affaires".

Sur les données 95, on retient uniquement les trois premières variables discriminantes. En effet, la liaison entre ces variables était quasiment nulle et des liaisons fortes apparaissent avec

variable	données 95		données 96		données 97		données 98	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
TCP	0.5257	0.4309						
CON	0.6847	0.6216	0.9855	0.9817	0.6938	0.6081		
TS	0.1621	-0.0055	0.2348	0.0364				
TCCP							0.2327	-0.3428
TCEP	0.6696	0.6035	0.9844	0.9804	0.7194	0.6409	0.5247	0.1683
CAPIPU	0.0853	-0.1355	0.5281	0.4058				
CUCPU							0.2150	-0.3738
DCIND					0.2322	0.0172		
CCDECE			0.2459	0.0503	0.2854	0.0853		
CCVIE					0.2726	0.0689		
RF			0.1106	-0.1200				
TCRF			0.5553	0.4400				
TCF							0.6594	0.4039
FSIN	0.4190	0.3028			0.2595	0.0522		
REXEP							0.7200	0.5100
RFPL							0.6924	0.4616
PREPPL	0.2543	0.87002						
PARPPL							0.7738	0.6042
IMMPPL			0.0303	-0.1775	0.2962	0.0991		
AUTPPL					0.3607	0.1818		

Table 37: coefficient de détermination sur les données stratifiées (suite)

les autres variables.

Le choix de la variable taux de conservation sur les données 96 exclut le taux de cessions et la branche contrats collectifs-décès. On retient aussi le ratio de sinistralité et la branche capitalisation-PU (ce qui exclut le taux d'accroissement du résultat financier).

Pour l'année 97, seulement trois variables ont été retenues, le taux de cessions, les contrats collectifs-vie et "immobiliers/produits des placements".

On retient enfin pour l'année 98 deux variables, le taux d'accroissement des cessions et le ratio "particuliers/produits des placements".

données 92	données 93	données 94
		TCP
	TCOM	
	CCOMPL	
RF		RF
TCF		
TCE	TCE	
		FPR
	REXPP	REXPP

Table 38: variables présélectionnées sur les données stratifiées

données 95	données 96	données 97	données 98
TCP			
	CON		
	TS		
			TCCP
TCEP		TCEP	
CAIPU	CAIPU		
		CCVIE	
	RF		
	IMMPPL	IMMPPL	
			PARTPPL

Table 39: variables présélectionnées sur les données stratifiées (suite)

4.4.3 Modèles sur les données stratifiées

Notons tout d'abord que le nombre des variables explicatives retenues est maintenant très réduit vu la taille relativement faible de l'échantillon complet. Un résultat surprenant montre que le pourcentage des bien classées au niveau de la classe déficitaire s'améliore nettement et dépasse même celui obtenu sur les modèles des données pondérées. Pour un seuil de coupure de 50%, on atteint maintenant jusqu'à près de 80% de compagnies bien classées dans la classe déficitaire et ceci malgré le nombre faible de variables explicatives retenues dans les différents modèles.

Les trois méthodes de sélection en pas à pas retiennent le même modèle pour l'années 92. Ce modèle montre que les variables résultat financier et taux de couverture des engagements interviennent négativement sur le déficit.

La même chose est aussi vraie pour les données 93 où le seul modèle retenu montre que le déficit est fonction décroissante des deux variables taux de couverture des engagements et "résultat d'exploitation/chiffre d'affaires".

Sur les données 94, on retient deux modèles. Le premier fait intervenir les deux variables résultat financier et "fonds propres/chiffre d'affaires" qui ont un rôle modérateur sur le déficit. Le deuxième rajoute la variable taux de croissance du chiffre d'affaires qui intervient aussi négativement sur le déficit.

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-1.9455	1.3327	0.1444
RF	-7.7263	4.6347	0.0955
TCE	-3.7062	2.2893	0.1055
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	11.418 (2 ddl)		0.0033
Score	5.246 (2 ddl)		0.0726
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	80.0	54.5	38.5
0.400	61.9	70.0	54.5
0.500	61.9	70.0	54.5
0.600	66.7	70.0	63.6
0.700	71.4	60.0	81.8

Table 40: modèle sur les données stratifiées 92

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	0.8045	0.7642	0.2925
TCE	-1.2187	0.7565	0.1072
REXPP	-3.2715	1.5097	0.0302
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	13.087 (2 ddl)		0.0014
Score	7.026 (2 ddl)		0.0298
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	70.8	83.3	58.3
0.400	79.2	83.3	75.0
0.500	79.2	75.0	83.3
0.600	70.8	58.3	83.3
0.700	70.8	58.3	83.3

Table 41: modèle sur les données stratifiées 93

	pas à pas descendant et stepwise			pas à pas ascendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-0.1569	0.5259	0.7654	-0.0601	0.5453	0.9123
TCP				-0.7533	0.5689	0.1855
RF	-1.6133	0.9673	0.0954	-1.5899	0.9494	0.0940
FPR	-1.0522	0.6232	0.0913	-1.1236	0.6832	0.1000
goodness of fit						
	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	7.702 (2 ddl)		0.0213	9.929 (3 ddl)		0.0192
Score	6.106 (2 ddl)		.0472	8.073 (3 ddl)		0.0445
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.300	54.2	83.3	25.0	54.2	75.0	33.3
0.400	58.3	75.0	41.7	58.3	66.7	50.0
0.500	62.5	58.3	66.7	66.7	66.7	66.7
0.600	58.3	41.7	75.0	62.5	50.0	75.0
0.700	66.7	41.7	91.7	66.7	50.0	83.3

Table 42: modèles sur les données stratifiées 94

Un seul modèle est retenu sur les données 95. On constate que dans ce modèle, les deux variables taux de croissance des primes et taux de cessions des primes ont un rôle modérateur sur le déficit et que la variable capitalisation-PU a un rôle accélérateur.

On retient deux modèles sur les données 96. Le premier montre que le déficit est fonction croissante du taux de sinistralité, et le second rajoute le taux conservation qui intervient aussi positivement.

Sur les données 97 on retient aussi deux modèles. Le premier montre que le déficit est fonction croissante de la variable contrats collectifs-vie. Le deuxième rajoute la variable taux de cessions et montre que le déficit est fonction décroissante de cette variable.

On a enfin deux autres modèles sur les données 98. Le premier fait intervenir le ratio “particuliers/produits des placements” qui intervient négativement. Le second rajoute la variable taux de croissance des cessions des primes qui joue aussi un rôle modérateur.

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-0.1224	0.5777	0.8321
TCP	-1.3348	0.6801	0.0497
TCEP	-2.5816	1.3499	0.0558
CAPIPU	2.0775	1.5922	0.1920
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	18.470 (3 ddl)		0.0004
Score	11.655 (3 ddl)		0.0087
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	69.4	88.9	50.0
0.400	69.4	77.8	61.1
0.500	75.0	77.8	72.2
0.600	72.2	66.7	77.8
0.700	75.0	61.1	88.9

Table 43: modèle sur les données stratifiées 95

	pas à pas descendant et stepwise			pas à pas ascendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	0.2532	0.4412	0.5659	0.1630	0.5441	0.7645
CON				1.2808	1.1060	0.2468
TS	1.7254	1.0909	0.1137	1.9735	1.2504	0.1145
goodness of fit						
	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	5.204 (1 ddl)		0.0225	9.094 (2 ddl)		0.0106
Score	3.181 (1 ddl)		0.0745	5.709 (2 ddl)		0.0576
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.300	47.1	88.2	5.9	52.9	88.2	17.6
0.400	47.1	64.7	29.4	52.9	76.5	29.4
0.500	61.8	52.9	70.6	58.8	58.8	58.8
0.600	61.8	29.4	94.1	64.7	41.2	88.2
0.700	58.8	17.6	100.0	61.8	23.5	100.0

Table 44: modèles sur les données stratifiées 96

	pas à pas descendant et stepwise			pas à pas ascendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	0.4473	0.6439	0.4872	0.6236	0.7769	0.4222
TCEP				-0.9622	0.9327	0.3023
CCVIE	2.0433	1.4288	0.1527	2.6659	1.8407	0.1475
goodness of fit						
	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	7.629 (1 ddl)		0.0057	10.174 (0.0062 ddl)		
Score	5.606 (1 ddl)		0.0179	7.048 (2 ddl)		0.0295
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.300	37.5	75.0	0.0	53.1	93.8	12.5
0.400	50.0	56.3	43.8	53.1	68.8	37.5
0.500	62.5	50.0	75.0	59.4	43.8	75.0
0.600	65.6	31.3	100.0	62.5	37.5	87.5
0.700	65.6	31.3	100.0	62.5	31.3	93.8

Table 45: modèles sur les données stratifiées 97

	pas à pas descendant et stepwise			pas à pas ascendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-0.0596	0.6282	0.9244	0.0944	0.7439	0.8990
TCCP				-1.0037	0.7270	0.1674
PARTPPL	-1.4973	1.0119	0.1390	-2.1896	1.4686	0.1360
goodness of fit						
	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	3.125 (1 ddl)		0.0771	5.496 (2 ddl)		0.0640
Score	2.832 (1 ddl)		0.0924	4.638 (2 ddl)		0.0983
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.300	42.9	66.7	25.0	64.3	83.3	50.0
0.400	64.3	66.7	62.5	64.3	66.7	62.5
0.500	50.0	33.3	62.5	57.1	50.0	62.5
0.600	64.3	33.3	87.5	50.0	33.3	62.5
0.700	57.1	16.7	87.5	57.1	33.3	75.0

Table 46: modèles sur les données stratifiées 98

4.5 Correction de la méthode de pondération

L'une des méthodes d'échantillonnage utilisée dans cet article a consisté à ramener la taille des deux échantillons considérés au même niveau en créant de nouvelles observations à partir des premières. L'inconvénient de cette méthode réside dans le fait qu'on rajoute des observations fictives et qu'on suppose par la suite que les deux populations considérées sont en proportions égales, ce qui est loin d'être vrai. *Par conséquent, les paramètres estimés seront biaisés.*

Rappelons tout d'abord que la méthode a consisté à associer un poids à chaque observation pour que tout se passe comme si l'on avait k fois la même observation dans la population initiale. Cette démarche ne permet pas de corriger ensuite l'estimation des paramètres. En effet, si pour corriger on avait multiplié par l'inverse du poids considéré, ceci n'aurait pas changé grand chose par rapport à la méthode utilisée sur les données non pondérées vu qu'on multiplie et on divise simultanément par le même facteur.

Pour corriger ces défauts, on commence tout d'abord par créer une autre population qui contient effectivement les données dupliquées. Il suffit ensuite de multiplier chaque observation par l'inverse de la fréquence qui lui a été associée pour corriger le biais engendré par la duplication.

La démarche consiste ainsi à garder la même taille de l'échantillon 0 (la fréquence associée est ainsi égale à 1), et pour la classe 1, on duplique chaque observation un certain nombre de fois de façon à avoir une taille à peu près égale à celle de la classe 0. L'échantillon ainsi obtenu correspond à la nouvelle population qui sera utilisée pour la modélisation.

Si l'on considère par exemple une population initiale constituée par 85 observations de la classe 0 et 10 observations de la classe 1, on rajoute 7 autres sous-échantillons contenant les mêmes informations que les observations de la classe 1. On obtient ainsi une population avec 85 observations de la classe 0 et 80 observations de la classe 1. Pour **redresser**, on associe un poids $1/8$ aux observations de la classe 1 et un poids égal à 1 aux observations de la classe 0.

4.5.1 tests univariés

Le test de Fisher est appliqué sur les différentes données pour détecter les variables les plus discriminantes.

Concernant la première période 92-94, on remarque que le résultat financier est retenu sur toute la période. D'autres ratios tels que le taux de charges, la branche capitalisation, le taux de couverture des engagements, le ratio "fonds propres/charge des sinistres", les résultats d'exploitation et de l'exercice et enfin le ratio "résultat financier/placements" sont retenus deux fois sur la période.

Pour la période 95-98, on note surtout la présence du taux de cession des primes, des branches contrats collectifs vie et capitalisation-PU, du ratio de solvabilité, des ratios "produits des placements/placements", "produits de réalisation/produits des placements" et "immobiliers/produits des placements".

variables discriminantes						
	données 92		données 93		données 94	
variable	F	Prob > F	F	Prob > F	F	Prob > F
TCP					3.0806	0.0810
TS	6.1670	0.0140				
TCOM					1.9688	0.1623
TCH	5.8950	0.0163	31.2181	0.0001		
CAPI	2.4830	0.1170	5.3498	0.0221		
AFF			10.1359	0.0018		
CCACAV			2.3386	0.1284		
RF	2.6855	0.1032	2.6234	0.1075	1.8539	0.1751
TCRF					3.2684	0.0723
TCF	9.6322	0.0023				
TCE			3.3371	0.0698	2.7413	0.0996
FSIN			3.5643	0.0610	1.4137	0.2360
FPR			10.0063	0.0019		
FRF			4.4569	0.0365		
RSOL			2.5497	0.1125		
REXPP	1.9652	0.1629	27.4868	0.0001		
REXEP			24.1887	0.0001	2.6594	0.1047
RFPL	6.5698	0.0113			3.1653	0.0769

Table 47: test de Fisher

variables discriminantes								
	données 95		données 96		données 97		données 98	
variable	F	Prob > F	F	Prob > F	F	Prob > F	F	Prob > F
CON	4.0344	0.0460			2.9427	0.0878		
TS	13.0194	0.0004	13.4575	0.0003				
TCS			2.8677	0.0919				
TFAA			12.1210	0.0006				
TACH			10.3657	0.0015				
TCEP	5.5341	0.0197			2.2563	0.1346	1.6139	0.2053
CAPIPU			2.7718	0.0975	2.3775	0.1246		
CUCPP	3.2747	0.0719						
CUCPU							1.4854	0.2242
CCDECE					1.2803	0.2592	1.4391	0.2316
CCVIE	3.0600	0.0819	7.2426	0.0077	11.7375	0.0007	1.4411	0.2312
RF							1.2468	0.2654
TCRF					1.5725	0.2113		
TCF			13.9385	0.0002				
TCE	3.9471	0.0484						
FSIN					1.7415	0.1884	1.1703	0.2805
RSOL	3.2934	0.0711	4.2195	0.0412	4.3016	0.0393		
RTECP			5.1221	0.0247			1.2625	0.2624
REXEP			6.1256	0.0141			1.1466	0.2854
PPLPL	3.5651	0.0605			1.6278	0.2035	2.7382	0.0994
RFPL	5.2425	0.0231					4.1853	0.0419
PREPPL	4.9569	0.0272	4.8985	0.0280				
IMMPPL	8.9757	0.0031	3.4899	0.0632			1.8667	0.1732

Table 48: test de Fisher (suite)

4.5.2 traitement de la colinéarité

On applique les mêmes méthodes étudiées dans les premières parties pour traiter le problème de liaisons entre les variables et pouvoir détecter les indicateurs les plus aptes à être introduits dans les modèles.

On se base surtout sur les corrélations simples entre toutes les variables ainsi que le coefficient de corrélation multiple de chaque variable en fonction des autres, ce qui permet d'exclure les variables redondantes.

La variable la plus discriminantes sur les données 92 est le taux de croissance des fonds propres. On exclut ainsi les deux variables capitalisation et "résultat financier/placements". On retient ensuite le taux de sinistralité, le résultat financier et le le ratio "résultat d'exploitation/chiffre d'affaires" en excluant aussi le taux de charges.

Sur l'ensemble de variables présélectionnées sur les données 93, seulement quatre ont été retenues. D'abord la plus discriminante, le taux de charges, qui était en forte liaison avec les ratios "autres affaires/chiffre d'affaires", "résultat de l'exercice/chiffre d'affaires" et "résultat d'exploitation/chiffre d'affaires". Ces variables ont été donc exclus. On retient ensuite le ratio "capitalisation/chiffre d'affaires", le résultat financier et le ratio "fonds propres/résultat financier". Toutes les autres variables ont été exclus pour cause de fortes liaisons.

Pour l'année 94, on retient tout d'abord le ratio "résultat financier/placements". On exclut ainsi le taux de croissance du résultat financier, le ratio "fonds propres/charge des sinistres", le ratio de solvabilité ainsi que les deux ratios "résultat d'exploitation/primes" et "résultat de l'exercice/primes". Toutes les autres variables ont été retenues.

variable	données 92		données 93		données 94	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
TCP					0.0640	0.0257
TS	0.2399	0.2110				
TCOM					0.1098	0.0733
TCH	0.5805	0.5646	0.6892	0.6640		
CAP	0.0953	0.0610	0.1400	0.0704		
AAFF			0.6007	0.5684		
CCACAV			0.1832	0.1171		
RF	0.0738	0.0386	0.1793	0.1129	0.0623	0.0240
TCRF					0.8546	0.8486
TCF	0.1397	0.1070				
TCE			0.4995	0.4590	0.4291	0.4058
FSIN			0.6401	0.6110	0.6447	0.6301
FPR			0.7141	0.6910		
FRF			0.6086	0.5770		
RSOL			0.5313	0.4934		
REXP	0.4998	0.4808	0.7907	0.7738		
REXEP			0.8082	0.7926	0.6184	0.6028
RFPL	0.1474	0.1150			0.8678	0.8624

Table 49: coefficient de détermination

Les quatre premières variables les plus discriminantes sur les donnée 95 ont été retenues. Il s'agit du taux de sinistralité, du ratio "immobilier/produits des placements", du taux de cessions (ce qui exclut le taux de conservation), et des deux ratios "produits de réalisation/produits des

placements” et “résultat financier/placements” (ce qui exclut les contrats en unité de compte-PU et le ratio “produits des placements/placements”. On retient aussi le taux de couverture des engagements et les contrats collectifs-vie.

Le taux d'accroissement des fonds propres était le plus discriminant sur les données 96. Le choix de ce ratio exclut le taux de sinistralité, le taux de frais d'acquisition et d'administration et le taux d'autres charges. On retient ensuite la branche contrats collectifs-vie, le ratio “résultat technique/chiffre d'affaires” (qui exclut les deux ratios “résultat de l'exercice/chiffre d'affaires” et “produits de réaalisation/produits des placements” ainsi que la branche capitalisation-PU), et enfin le ratio “immobiliers/produits des placements”.

Pour l'année 97, on retient la branche contrats collectifs-vie, le taux de conservations (qui exclut le taux de cessions), le ratio de solvabilité (qui exclut la branche contrats collectifs-décès et le ratio “fonds propres/charge des sinistres”), la branche capitalisation-PU, le taux de croissance du résultat financier et le ratio “produits des placements/placements”.

Enfin, pour l'année 98, on retient tout d'abord le ratio “résultat financier/placements” et l'on exclut le ratio “produits des placements/placements”. On retient ensuite le taux de cessions (qui exclut les deux branches contrats en unité de compte-PU et contrats collectifs-décès), les contrats collectifs-vie et les deux ratios “résultat technique/chiffre d'affaires” et “fonds propres/charge des sinistres”.

variable	données 95		données 96		données 97		données 98	
	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq	R-square	Adj R-sq
CON	0.6639	0.6455			0.5209	0.5016		
TS	0.4079	0.3754	0.9535	0.9509				
TCS			0.2548	0.2134				
TFAA			0.9448	0.9417				
TACH			0.9761	0.9748				
TCEP	0.6491	0.6298			0.5198	0.5005	0.2066	0.1702
CAIPU			0.3813	0.3470	0.0421	0.0035		
CUCPP	0.3988	0.3658						
CUCPU							0.2130	0.1769
CCDECE					0.3076	0.2797	0.3488	0.3189
CCVIE	0.0716	0.0206	0.0326	-0.0211	0.0234	-0.0159	0.1334	0.0936
RF							0.0703	0.0276
TCRF					0.0571	0.0192		
TCF			0.2757	0.2354				
TCE	0.6615	0.6429						
FSIN					0.2153	0.1838	0.6821	0.6675
RSOL	0.6619	0.6433	0.1739	0.1281	0.3449	0.3185		
RTECP			0.8990	0.8933			0.4443	0.4188
REXEP			0.9052	0.8999			0.6748	0.6599
PPLPL	0.8831	0.8767			0.0750	0.0379	0.8368	0.8293
RFPL	0.8716	0.8646					0.8359	0.8284
PREPPL	0.2250	0.1824	0.0746	0.0232				
IMPPPL	0.4069	0.3743	0.0899	0.0394			0.0286	-0.0160

Table 50: coefficient de détermination (suite)

4.5.3 Modèles corrigés sur les données pondérées

La première constatation qu'on peut faire est qu'au niveau des pourcentages des bien classées, ces modèles corrigés sont moins bons que les modèles obtenus sur les données pondérées, mais classent mieux que les modèles obtenus sur les données non pondérées.

Un seul modèle est établi sur les données 92. On peut conclure que le déficit est d'une part fonction croissante taux de sinistralité et du taux de croissance du résultat financier, et d'autre part fonction décroissante du ratio "résultat d'exploitation/chiffre d'affaires".

On a aussi un seul modèle sur les données 93 qui montre que le taux de charges et la branche capitalisation jouent un rôle accélérateur sur le déficit et qu'à l'opposé, le ratio "fonds propres/résultat financier" joue un rôle modérateur.

Trois modèles sont établis sur les données 94. Le premier introduit uniquement deux variables et montre que le taux de couverture des engagements et le ratio "résultat financier/placements" jouent un rôle modérateur sur le déficit. Un deuxième modèle rajoute au premier la variable taux de croissance du chiffre d'affaires qui joue aussi un rôle modérateur sur le déficit. Enfin, le troisième modèle introduit la variable taux de commissions et montre que ce taux joue un rôle accélérateur.

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-1.8732	0.3979	0.0001
TS	1.8851	0.9229	0.0411
TCF	0.8899	0.4671	0.0568
REXPP	-0.7007	0.4058	0.0842
goodness of fit	χ^2 for Cov		P-Value
-2 Log L	12.593 (3 ddl)		0.0056
Score	19.085 (3 ddl)		0.0003
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.180	67.0	40.0	94.8
0.200	62.4	30.0	95.9
0.300	62.9	30.0	96.9
0.400	63.5	30.0	97.9
0.500	58.4	20.0	97.9
0.600	3.8	10.0	99.0
0.700	53.8	10.0	99.0

Table 51: modèle corrigé sur les données pondérées 92

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-2.0402	0.4202	0.0001
TCH	1.9504	0.6888	0.0046
CAPI	0.7499	0.3446	0.0295
FRF	-1.9671	1.0189	0.0535
goodness of fit	χ^2 for Cov		P-Value
-2 Log L	17.593 (3 ddl)		0.0005
Score	21.441 (3 ddl)		0.0001
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.300	70.8	44.4	95.4
0.400	60.1	22.2	95.4
0.500	61.3	22.2	97.7
0.600	61.3	22.2	97.7
0.700	61.3	22.2	97.7

Table 52: modèle corrigé sur les données pondérées 93

	stepwise			pas à pas ascendant			pas à pas descendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-2.0689	0.3202	0.0001	-2.0788	0.3357	0.0001	-2.0695	0.3290	0.0001
TCP				-0.4752	0.3609	0.1880	-0.4676	0.3481	0.1792
TCOM				0.3486	0.2876	0.2255			
TCE	-0.4884	0.3433	0.1548	-0.5177	0.3519	0.1412	-0.4551	0.3326	0.1712
RFPL	-0.5776	0.4161	0.1651	-0.6702	0.4420	0.1294	-0.6393	0.4390	0.1454
goodness of fit	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	4.007 (2 ddl)		0.1348	7.333 (4 ddl)		0.1193	6.017 (3 ddl)		0.1108
Score	4.708 (2 ddl)		0.0950	7.866 (4 ddl)		0.0966	6.720 (3 ddl)		0.0814
pourcentage des bien classées									
Cutoff Pr	Total	class1	class2	Total	class1	class2	Total	class1	class2
0.180	59.1	16.7	95.9	68.5	41.7	91.8	56.9	16.7	91.8
0.200	55.2	8.3	95.9	65.2	33.3	92.8	58.0	16.7	93.8
0.300	56.4	8.3	97.9	55.2	8.3	95.9	55.8	8.3	96.9
0.400	56.9	8.3	99.0	55.8	8.3	96.9	56.4	8.3	97.9
0.500	56.9	8.3	99.0	56.4	8.3	97.9	56.4	8.3	97.9
0.600	53.0	0.0	99.0	56.4	8.3	97.9	56.9	8.3	99.0
0.700	53.0	0.0	99.0	53.0	0.0	99.0	53.0	0.0	99.0

Table 53: modèles corrigés sur les données pondérées 94

Les trois méthodes de sélection des variables aboutissent à un seul modèle sur les données 95. Ce modèle montre que d'une part le déficit est fonction croissante des variables taux de sinistralité, contrats collectifs-vie et "immobiliers/produits des placements", et que d'autre part le déficit est fonction décroissante du taux de cession des primes.

Une seule version de modèle est établie sur les données 96. A partir de ce modèle, on peut conclure que le déficit est fonction croissante des quatre variables contrats collectifs-vie, taux de croissance des fonds propres, "résultat technique/chiffre d'affaires" et "immobiliers/produits des placements".

Trois modèles sont aussi établis sur les données 97. Le premier montre que le déficit est fonction croissante des variables taux de conservation, capitalisation-PU, contrats collectifs-vie et ratio de solvabilité. Le deuxième modèle rajoute au premier le ratio "produits des placements/placements" qui joue aussi un rôle accélérateur sur le déficit. Le troisième modèle exclut du premier le taux de conservation et rajoute le taux de croissance du résultat financier qui joue un rôle modérateur sur le déficit.

Sur les données 98, on établit deux modèles. Le premier retient uniquement une seule variable, le taux de cessions, et montre que le déficit est une fonction décroissante de cette variable. Le second retient quatre variables et montre que le déficit est fonction croissante des variables contrats collectifs-vie, "résultat technique/chiffre d'affaire" et "résultat financier/chiffre d'affaires", et fonction décroissante du ratio "fonds propres/charge des sinistres". Les entreprises bien classées sont cependant en pourcentage très faible, mais cette même remarque a été faite sur tous les autres modèles établis sur les données 98. Cette année d'observation contient le moins d'observations dans la classe déficitaire.

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-1.8914	0.3635	0.0001
TS	1.5285	0.7856	0.0517
TCEP	-1.9912	0.9651	0.0391
CCVIE	0.4668	0.2588	0.0712
IMPPPL	0.7907	0.5531	0.1528
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	20.061 (4 ddl)		0.0005
Score	18.293 (4 ddl)		0.0011
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.180	70.6	61.1	78.8
0.200	71.1	61.1	79.8
0.300	65.5	33.3	93.3
0.400	59.8	16.7	97.1
0.500	60.8	16.7	99.0
0.600	61.3	16.7	100.0
0.700	61.3	16.7	100.0

Table 54: modèle corrigé sur les données pondérées 95

	les trois méthodes de sélection		
variable	Pa-Est	St-Er	Pr > χ^2
CONST	-1.6878	0.2864	0.0001
CCVIE	0.6755	0.2621	0.0100
TCF	0.8123	0.4081	0.0465
RTECP	0.6550	0.3709	0.0774
IMPPPL	0.3671	0.2442	0.1327
goodness of fit			
	χ^2 for Cov		P-Value
-2 Log L	14.967 (4 ddl)		0.0048
Score	16.880 (4 ddl)		0.0020
pourcentage des bien classées			
Cutoff Pr	Total	class1	class2
0.180	65.4	50.0	81.6
0.200	63.5	44.4	83.5
0.300	67.8	44.4	92.2
0.400	58.3	22.2	96.1
0.500	59.7	22.2	99.0
0.600	54.0	11.1	99.0
0.700	54.0	11.1	99.0

Table 55: modèle corrigé sur les données pondérées 96

variable	stepwise			pas à pas ascendant			pas à pas descendant		
	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-1.9896	0.3208	0.0001	-1.9213	0.3241	0.0001	-1.9624	0.3021	0.0001
CON	0.7585	0.5672	0.1812	0.7524	0.5634	0.1817			
CAIPU	0.3288	0.2513	0.1907	0.3140	0.2534	0.2152	0.4012	0.2480	0.1057
CCVIE	0.7000	0.2658	0.0084	0.7499	0.2721	0.0059	0.8100	0.2732	0.0030
TCRF							-0.8280	0.5329	0.1203
RSOL	0.6544	0.2772	0.0182	0.6395	0.2783	0.0216	0.5219	0.2647	0.0487
PPLPL				0.8159	0.7136	0.2529			
goodness of fit									
	χ^2 for Cov			χ^2 for Cov			χ^2 for Cov		
-2 Log L	13.554 (4 ddl)			14.696 (5 ddl)			13.063 (4 ddl)		
Score	15.542 (4 ddl)			16.719 (5 ddl)			13.807 (4 ddl)		
pourcentage des bien classées									
Cutoff Pr	Total	class1	class2	Total	class1	class2	Total	class1	class2
0.180	67.6	43.8	88.3	67.1	43.8	87.4	66.7	43.8	86.5
0.200	68.1	43.8	89.2	67.1	43.8	87.4	67.6	43.8	88.3
0.300	69.1	43.8	91.0	69.1	43.8	91.0	68.6	43.8	90.1
0.400	64.3	31.3	92.8	66.7	37.5	91.9	61.4	25.0	92.8
0.500	52.7	0.0	98.2	55.6	6.3	98.2	52.7	0.0	98.2
0.600	53.1	0.0	99.1	55.6	6.3	98.2	53.1	0.0	99.1
0.700	53.1	0.0	99.1	53.1	0.0	99.1	53.1	0.0	99.1

Table 56: modèles corrigés sur les données pondérées 97

	pas à pas ascendant			pas à pas descendant		
variable	Pa-Est	St-Er	Pr > χ^2	Pa-Est	St-Er	Pr > χ^2
CONST	-2.8495	0.4841	0.0001	-4.9138	1.5082	0.0011
TCEP	-0.8569	1.0023	0.3926			
CCVIE				0.4828	0.3278	0.1408
FSIN				-8.2941	5.0322	0.0993
RTECP				1.1504	0.8006	0.1508
RFPL				0.3727	0.3677	0.3108
goodness of fit	χ^2 for Cov		P-Value	χ^2 for Cov		P-Value
-2 Log L	1.503 (1 ddl)		0.2202	8.080 (4 ddl)		0.0887
Score	0.903 (1 ddl)		0.3420	2.572 (4 ddl)		0.6317
pourcentage des bien classées						
Cutoff Pr	Total	class1	class2	Total	class1	class2
0.040	57.5	100.0	16.4	64.9	87.5	43.1
0.060	49.1	75.0	24.1	66.7	75.0	58.6
0.080	31.6	0.0	62.1	72.8	75.0	70.7
0.100				64.9	50.0	79.3
0.160				52.6	12.5	91.4
0.200				54.4	12.5	94.8
0.300				55.7	12.5	97.4
0.400				56.1	12.5	98.3
0.440				56.1	12.5	98.3
0.460				50.0	0.0	98.3

Table 57: modèles corrigés sur les données pondérées 98

5 Conclusions

Plusieurs ratios financiers ont été testés dans cette étude pour détecter ceux d'entre eux qui possèdent un pouvoir discriminant entre les deux états de sociétés, solvable et déficitaire. Parmi les ratios discriminants, peu (entre 2 et 6 ratios) ont été préseélectionnés à chaque fois pour être introduits dans les différents modèles. Plusieurs raisons, telles que la forte corrélation constatée entre les variables explicatives utilisées et le nombre assez faible d'observations appartenant à la classe déficitaire, ont été derrière cette limitation du nombre de ratios à introduire dans les modèles.

Deux types de ratios ont été détectés. D'abord ceux qui peuvent accroître le déficit, la compagnie a ainsi intérêt à limiter ces ratios. D'autre part, des ratios qui peuvent au contraire améliorer la santé financière de la compagnie, et on a ainsi intérêt à les accroître.

Différentes méthodes d'échantillonnage ont été utilisées sachant que certaines d'entre elles peuvent renfermer des inconvénients. L'objectif était de pouvoir constater empiriquement la différence entre ces méthodes. D'autre part, certaines méthodes telles que "the separate sampling" et la pondération corrigée, ont été suggérées soit pour pallier la dissymétrie entre les deux classes de la population, soit pour corriger le biais engendré par d'autres.

Si l'on prend comme critère de bonne appropriation d'un modèle le pourcentage des bien classées, on remarque qu'on a en général un faible pourcentage concernant les entreprises de la classe déficitaire, sauf pour les deux méthodes d'échantillonnage par pondération non corrigée et "choice-based sampling" où on atteint, pour un seuil de coupure de 50%, jusqu'à 63% pour la première méthode et 80% pour la seconde.

Rappelons que les ratios utilisés dans ce travail sont basés sur des données comptables publiques, ce qui signifie que les informations disponibles sont assez limitées. On peut donc imaginer que d'autres ratios basées sur *des données confidentielles* peuvent apporter beaucoup plus d'informations sur la santé financière des sociétés d'assurance, mais malheureusement le marché reste très fermé par rapport à la diffusion de l'information et ceci pour diverses raisons telles que *la guerre commerciale*.

D'un autre côté, on peut être satisfait du fait que certains modèles enregistrent entre 50 et 80% de biens classées dans la classe déficitaire pour un seuil de coupure de 50% avec assez peu de ratios financiers, ce qui montre l'importance de ces ratios.

Au niveau théorique, on a mis l'accent sur le problème d'événements rares et les méthodes appropriées lorsqu'on est confronté à ce genre de problème sont exposées. On a ainsi étudié les méthodes d'estimation et les propriétés asymptotiques des estimateurs de *la vraisemblance empirique*.

Appendice I : corrélations simples sur les données non pondérées

	TS	TCH	CAPI	RF	TCF	REXPP	RFPL
TS	1.00000	0.45601	-0.08173	-0.02630	-0.12878	-0.07681	0.10107
TCH		1.00000	-0.20065	-0.11603	-0.15531	-0.46352	0.25789
CAPI			1.00000	-0.03873	0.23706	0.09948	0.01076
RF				1.00000	-0.00422	0.05528	0.02177
TCF					1.00000	0.05730	-0.16046
REXPP						1.00000	0.23179
RFPL							1.00000

Table 58: matrice des corrélations simples sur les données non pondérés 92

	TCH	DEC	ACAV	CAPI	AFF	RF	TCE	FSIN	FPR	FRF	RSOL	EXP	EXE
TCH	1.000	0.022	-0.166	-0.163	0.612	-0.143	0.057	0.175	0.307	0.024	0.156	-0.712	-0.594
CCDE		1.000	-0.103	-0.226	-0.097	-0.152	0.401	-0.030	0.120	0.427	0.349	0.152	0.115
CCAC			1.000	-0.112	-0.070	0.116	0.035	0.043	-0.059	-0.056	0.026	0.017	-0.049
CAPI				1.000	-0.089	-0.041	-0.019	0.070	0.002	-0.039	-0.033	0.043	0.048
AAFF					1.000	-0.039	-0.233	-0.212	-0.468	-0.254	-0.206	-0.516	-0.654
RF						1.000	-0.111	-0.154	-0.122	-0.193	-0.051	0.124	0.046
TCE							1.000	0.405	0.392	0.671	0.660	0.085	0.034
FSIN								1.000	0.583	0.640	0.683	-0.147	-0.103
FPR									1.000	0.530	0.556	-0.184	0.058
FRF										1.000	0.918	0.114	0.071
RSOL											1.000	0.024	-0.023
REXPP												1.000	0.741
REXEP													1.000

Table 59: matrice des corrélations simples sur les données non pondérés 93

	TCP	TCOM	RF	TCRF	TCE	FSIN	RSOL	EXPP	EXEP	RFPL
TCP	1.00000	-0.00961	0.03944	-0.02616	0.02522	-0.09479	-0.04002	0.13725	0.10744	-0.11160
TCOM		1.00000	-0.12831	-0.02239	0.03172	-0.11638	-0.02145	-0.00109	-0.03814	0.00051
RF			1.00000	0.12711	-0.04376	-0.12387	-0.11531	0.07143	0.09338	0.14242
TCRF				1.00000	-0.07488	-0.25768	-0.30999	0.35081	0.47362	0.90987
TCE					1.00000	0.51650	0.71134	-0.39294	-0.35132	-0.17535
FSIN						1.00000	0.81311	-0.67839	-0.63352	-0.49476
RSOL							1.00000	-0.75192	-0.74343	-0.29542
EXPP								1.00000	0.88003	0.30067
EXEP									1.00000	0.39308
RFPL										1.00000

Table 60: matrice des corrélations simples sur les données non pondérées 94

	CON	TS	TCEP	CUCPP	CCVIE	TCE	RSOL	PPLPL	RFPL	PRE	IMM
CON	1.000	0.071	-0.827	0.076	0.119	0.039	0.023	0.057	0.052	0.005	0.098
TS		1.000	0.023	0.238	-0.045	-0.067	-0.042	0.017	0.031	-0.047	0.111
TCEP			1.000	-0.062	-0.084	-0.116	-0.042	-0.052	-0.045	-0.009	-0.059
CUCPP				1.000	-0.075	0.008	0.073	0.881	0.913	-0.090	-0.016
CCVIE					1.000	0.001	0.007	-0.067	-0.039	-0.154	-0.115
TCE						1.000	0.817	-0.026	-0.014	0.020	-0.097
RSOL							1.000	0.033	0.045	0.041	-0.082
PPLPL								1.000	0.903	-0.156	0.022
RFPL									1.000	-0.120	0.016
PREPPL										1.000	-0.005
IMMPPL											1.000

Table 61: matrice des corrélations simples sur les données non pondérées 95

	TS	TCS	FADM	ACH	CAPPU	CCVIE	TCF	RSOL	TECP	EXEP	PRE	IMM
TS	1.000	-0.003	0.908	0.965	0.019	-0.076	0.401	0.012	-0.016	0.025	0.042	0.036
TCS		1.000	-0.026	-0.000	0.331	-0.017	-0.040	-0.013	-0.190	-0.178	-0.055	-0.017
FADM			1.000	0.955	-0.037	-0.053	0.457	0.027	0.141	0.200	0.086	0.030
ACH				1.000	-0.014	-0.057	0.427	0.018	0.002	0.073	0.092	0.020
CAPPU					1.000	-0.079	-0.041	-0.051	0.262	0.300	-0.159	-0.062
CCVIE						1.000	-0.102	-0.085	-0.017	-0.042	-0.057	-0.104
TCF							1.000	-0.001	0.149	0.138	-0.010	0.119
RSOL								1.000	0.105	0.237	0.152	0.173
TECP									1.000	0.872	-0.175	-0.052
EXEP										1.000	-0.016	-0.063
PREPP											1.000	-0.023
IMMPP												1.000

Table 62: matrice des corrélations simples sur les données non pondérées 96

	CON	TCEP	CAPPU	CCDEC	CCVIE	TCRF	FSIN	RSOL	PPPL
CON	1.00000	-0.75702	0.11186	-0.22342	0.03720	-0.10413	-0.15641	-0.18284	0.04876
TCEP		1.00000	-0.10618	0.22340	-0.09427	0.06609	0.09862	0.12537	-0.01093
CAPPU			1.00000	-0.14589	-0.04963	-0.00999	-0.06897	-0.06730	-0.02453
CCDEC				1.00000	-0.09338	-0.10592	0.47048	0.46370	-0.06139
CCVIE					1.00000	0.03537	-0.08701	-0.10638	-0.09392
TCRF						1.00000	-0.08418	0.00047	-0.07916
FSIN							1.00000	0.30162	0.01504
RSOL								1.00000	0.06810
PPPL									1.00000

Table 63: matrice des corrélations simples sur les données non pondérées 97

	TCEP	CUCPU	CCDE	CCVI	RF	FSIN	TECP	EXEP	PPLPL	RFPL	IMM
TCEP	1.000	-0.231	0.297	-0.039	-0.179	-0.017	-0.056	-0.046	-0.012	-0.043	0.033
CUCPU		1.000	-0.232	-0.195	0.030	-0.175	0.042	-0.097	0.054	-0.056	-0.010
CCDE			1.000	-0.062	-0.111	0.442	0.071	0.227	-0.062	-0.049	-0.007
CCVI				1.000	-0.115	-0.006	-0.201	-0.031	-0.113	-0.072	-0.095
RF					1.000	-0.118	0.063	-0.037	-0.015	0.005	0.036
FSIN						1.000	-0.051	0.666	0.041	0.033	-0.092
TECP							1.000	0.424	0.038	0.044	0.064
EXEP								1.000	0.056	0.055	-0.023
PPLPL									1.000	0.904	-0.032
RFPL										1.000	-0.055
IMMPPL											1.000

Table 64: matrice des corrélations simples sur les données non pondérées 98

Appendice II : corrélations simples sur les données pondérées

	TS	TCS	TCH	ACAV	CAP	RF	TCRF	TCF	TCE	FRF	EXPP	RFPL
TS	1.000	-0.072	0.639	0.080	-0.275	-0.083	-0.138	-0.204	-0.053	-0.104	-0.199	0.181
TCS		1.000	-0.084	0.115	-0.034	-0.045	0.351	0.019	0.126	0.273	-0.021	-0.019
TCH			1.000	-0.089	-0.288	-0.153	-0.136	-0.179	-0.070	-0.349	-0.703	0.143
ACAV				1.000	-0.122	0.074	-0.017	-0.040	0.015	0.145	0.183	-0.112
CAPI					1.000	-0.084	0.008	0.545	0.051	0.077	0.146	-0.097
RF						1.000	-0.076	-0.037	-0.027	-0.054	0.131	0.027
TCRF							1.000	0.000	0.008	0.042	-0.001	-0.155
TCF								1.000	-0.022	0.098	0.060	-0.299
TCE									1.000	0.563	0.302	0.514
FRF										1.000	0.324	0.024
EXPP											1.000	0.145
RFPL												1.000

Table 65: matrice des corrélations simples sur les données pondérées 92

	TCP	TCH	DEC	ACA	CAP	AA	RF	TCE	FS	FP	FRF	RS	EXP	EXE	PPP
TCP	1.00	-0.17	0.24	0.02	0.04	-0.21	-0.00	-0.19	0.36	0.04	0.36	-0.30	0.16	0.12	0.05
TCH		1.00	-0.12	-0.16	-0.29	0.85	-0.19	-0.28	-0.18	-0.14	-0.48	-0.26	-0.89	-0.69	0.13
CCDE			1.00	-0.09	-0.27	-0.15	-0.143	0.31	0.00	0.19	0.40	0.33	0.14	0.14	-0.09
ACAV				1.00	-0.10	-0.08	0.17	0.10	0.10	0.02	0.02	0.08	0.10	0.05	0.19
CAPI					1.00	-0.16	-0.04	0.08	0.08	-0.05	0.01	-0.01	0.20	0.08	0.04
AUTA						1.00	-0.11	-0.40	-0.53	-0.85	-0.68	-0.49	-0.87	-0.91	0.11
RF							1.00	-0.08	-0.07	-0.06	-0.07	-0.09	0.18	0.10	0.10
TCEN								1.00	0.42	0.42	0.61	0.91	0.29	0.25	-0.51
FSIN									1.00	0.46	0.65	0.72	0.19	0.20	-0.15
FPR										1.00	0.58	0.55	0.15	0.59	0.14
FRF											1.00	0.91	0.53	0.56	-0.21
RSOL												1.00	0.34	0.37	-0.57
EXPP													1.00	0.74	-0.07
EXEP														1.00	-0.01
PPPL															1.00

Table 66: matrice des corrélations simples sur les données pondérées 93

	TCP	TS	TCOM	RF	TCRF	TCEN	FSIN	FPR	RSOL	EXEP	PPPL	RFPL
TCP	1.000	-0.191	-0.016	0.077	0.018	0.023	-0.036	-0.210	0.007	0.147	-0.126	-0.006
TS		1.000	-0.047	0.003	-0.172	-0.239	-0.057	0.152	-0.017	-0.065	0.041	-0.107
TCOM			1.000	-0.142	0.062	-0.097	-0.131	-0.128	-0.047	-0.182	0.289	0.121
RF				1.000	0.138	0.019	-0.088	-0.070	-0.086	0.132	0.090	0.130
TCRF					1.000	0.019	-0.100	-0.104	-0.179	0.493	0.191	0.969
TCEN						1.000	0.470	0.389	0.65752	-0.239	-0.245	-0.078
FSIN							1.000	0.671	0.815	-0.441	-0.378	-0.249
FPR								1.000	0.651	-0.370	-0.151	-0.072
RSOL									1.000	-0.581	-0.373	-0.193
REXEP										1.000	0.173	0.440
PPPL											1.000	0.280
RFPL												1.000

Table 67: matrice des corrélations simples sur les données pondérées 94

	CON	TS	TCS	TCEP	DCOI	CCVI	FSIN	FRF	RSOL	PPPL	RFPL	PRE	IMM
CON	1.00	0.10	0.04	-0.79	0.01	0.15	0.07	-0.18	0.05	0.05	0.05	-0.08	0.11
TS		1.00	-0.04	-0.01	-0.02	-0.08	-0.14	-0.07	-0.04	-0.00	0.01	-0.11	0.06
TCS			1.00	-0.01	-0.03	0.01	0.12	0.03	0.02	0.70	-0.64	-0.02	-0.07
TCEP				1.00	0.00	-0.08	-0.07	0.18	-0.04	-0.05	-0.05	0.11	-0.05
DCOI					1.00	-0.13	-0.09	-0.06	-0.05	-0.05	-0.04	0.29	0.02
CCVI						1.00	0.01	-0.13	-0.08	-0.08	-0.04	-0.19	-0.15
FSIN							1.00	0.62	0.41	-0.06	-0.03	0.14	-0.15
FRF								1.00	0.65	-0.08	-0.02	0.23	-0.11
RSOL									1.00	-0.05	-0.03	-0.04	-0.11
PPPL										1.00	0.90	-0.16	0.12
RFPL											1.00	-0.13	0.09
PREPPL												1.00	-0.10
IMMPPL													1.00

Table 68: matrice des corrélations simples sur les données pondérées 95

	CON	TS	TCS	FADM	ACH	CPU	CCVI	TCF	FSIN	RSOL	TEC	EXE	PRE	IMM
CON	1.00	-0.33	-0.00	-0.36	-0.36	0.16	0.15	-0.20	0.01	0.06	-0.10	-0.08	-0.17	0.07
TS		1.00	-0.04	0.97	0.98	-0.01	-0.11	0.50	-0.05	-0.04	0.23	0.27	0.13	-0.00
TCS			1.00	-0.02	-0.01	0.16	-0.033	-0.02	0.02	0.11	-0.14	-0.13	0.01	0.11
FADM				1.00	0.98	-0.073	-0.10	0.53	-0.01	-0.01	0.27	0.29	0.17	-0.00
ACHA					1.00	-0.016	-0.10	0.52	-0.01	-0.02	0.26	0.30	0.17	-0.00
CAPU						1.00	-0.08	-0.10	0.13	-0.07	0.55	0.62	-0.17	-0.07
CCVI							1.00	-0.13	-0.12	-0.16	-0.08	-0.11	-0.21	-0.16
TCF								1.00	-0.06	0.05	0.14	0.14	0.13	0.19
FSIN									1.00	0.25	0.14	0.25	0.19	-0.13
RSOL										1.00	0.16	0.18	0.03	0.41
TECP											1.00	0.93	-0.15	-0.13
EXEP												1.00	-0.04	-0.10
PREPP													1.00	0.03
IMMP														1.00

Table 69: matrice des corrélations simples sur les données pondérées 96

	CON	TCEP	CAIPU	CCOLVIE	TCRF	FSIN	RSOL	RTECP	REXEP	PPLPL
CON	1.000	-0.761	0.136	0.114	-0.150	-0.164	-0.106	-0.008	-0.069	0.020
TCEP		1.000	-0.125	-0.145	0.105	0.107	0.058	-0.138	0.068	0.011
CAIPU			1.000	-0.052	-0.021	-0.003	-0.079	0.172	0.264	0.024
CCOLVI				1.000	0.003	-0.103	-0.194	0.012	-0.034	-0.131
TCRF					1.000	-0.097	0.114	0.018	0.029	0.004
FSIN						1.000	0.187	-0.014	0.380	0.009
RSOL							1.000	0.068	0.088	0.386
RTECP								1.000	0.864	0.043
REXEP									1.000	0.058
PPLPL										1.000

Table 70: matrice des corrélations simples sur les données pondérées 97

	TCS	TCEP	CUCPU	CCVI	RF	TCF	FSIN	FPR	TECP	EXEP	PPPL	RFPL	IMM
TCS	1.00	0.38	0.01	-0.07	0.14	0.14	-0.02	-0.03	-0.00	-0.01	-0.00	0.15	0.11
TCEP		1.00	-0.16	-0.05	-0.14	-0.01	0.02	-0.01	-0.12	0.00	-0.00	-0.05	0.07
CUCPU			1.00	-0.31	0.06	-0.06	-0.18	-0.16	-0.01	-0.15	0.20	0.06	-0.04
CCVI				1.00	-0.15	0.20	-0.03	0.08	-0.12	0.02	-0.20	-0.13	-0.11
RF					1.00	0.13	-0.09	-0.07	0.02	-0.04	-0.03	-0.01	0.09
TCF						1.00	0.19	0.43	-0.22	0.16	-0.03	-0.02	0.01
FSIN							1.00	0.57	-0.06	0.64	0.02	0.02	-0.05
FPR								1.00	-0.08	0.71	-0.05	-0.02	-0.03
TECP									1.00	0.33	0.03	0.05	0.02
EXEP										1.00	0.07	0.07	0.00
PPPL											1.00	0.90	-0.04
RFPL												1.00	-0.07
IMM													1.00

Table 71: matrice des corrélations simples sur les données pondérées 98

Appendice III : corrélations simples sur les données stratifiées

	TCH	CAPI	CACAV	RF	TCF	TCEN	FSIN	REXPP
TCH	1.00000	-0.28944	0.06756	-0.20483	-0.16265	-0.37269	-0.32284	-0.71919
CAPI		1.00000	-0.10335	-0.01001	0.39179	-0.02917	0.19709	0.17096
CACAV			1.00000	-0.06967	-0.05381	-0.07147	-0.08146	-0.34364
RF				1.00000	0.00503	-0.03626	0.00564	0.18996
TCF					1.00000	-0.02496	0.26394	0.07304
TCEN						1.00000	0.19768	0.39815
FSIN							1.00000	0.22551
REXPP								1.00000

Table 72: matrice des corrélations simples sur les données stratifiées 92

	TCOM	TCH	CCACAV	CCOMP	RF	TCEN	REXPP	REXEP
TCOM	1.00000	-0.02675	-0.19375	-0.06904	-0.29747	-0.03787	-0.05853	-0.13041
TCH		1.00000	-0.24257	0.34050	-0.20646	-0.41255	-0.80181	-0.63341
CCACAV			1.00000	0.22874	0.05621	0.32525	0.10352	0.02324
CCOMP				1.00000	-0.28421	0.09584	-0.32332	0.07665
RF					1.00000	0.17902	0.33232	0.21694
TCEN						1.00000	0.23686	0.21789
REXPP							1.00000	0.75910
REXEP								1.00000

Table 73: matrice des corrélations simples sur les données stratifiées 93

	TCP	TCEP	CCOMP	RF	TCE	FPR	FRF	REXPP	REXEP
TCP	1.00000	0.12102	-0.09485	0.01824	-0.02441	0.00608	-0.10021	-0.04526	0.23854
TCEP		1.00000	0.33420	-0.17993	0.18100	0.26984	0.43683	0.03511	0.12974
CCOMP			1.00000	-0.17691	0.50874	0.15239	0.25382	0.20589	0.13041
RF				1.00000	-0.06388	-0.18119	-0.17375	0.21619	0.19890
TCE					1.00000	0.29314	0.37485	0.25200	0.14164
FPR						1.00000	0.73359	0.18472	0.08360
FRF							1.00000	0.41664	0.25867
REXPP								1.00000	0.80181
REXEP									1.00000

Table 74: matrice des corrélations simples sur les données stratifiées 94

	TCP	CON	TS	TCEP	CAIPU	FSIN	PREPPL
TCP	1.00000	-0.20315	-0.38374	-0.02126	-0.04875	0.56915	0.35675
CON		1.00000	0.03475	-0.78099	0.11788	0.02865	-0.19426
TS			1.00000	0.09148	0.05393	-0.15237	-0.10682
TCEP				1.00000	-0.06105	-0.03239	0.20223
CAIPU					1.00000	0.08945	-0.20534
FSIN						1.00000	0.36201
PREPPL							1.00000

Table 75: matrice des corrélations simples sur les données stratifiées 95

	CON	TS	TCEP	CAIPU	CCDECE	RF	TCRF	IMPPPL
CON	1.00000	0.06778	-0.99027	0.11532	-0.42186	0.16718	-0.19083	0.06497
TS		1.00000	-0.05283	0.28803	-0.05132	-0.07349	-0.15823	0.06632
TCEP			1.00000	-0.11151	0.39452	-0.14553	0.15329	-0.05264
CAIPU				1.00000	-0.16655	-0.08266	0.54480	-0.06286
CCDECE					1.00000	-0.15837	-0.05439	-0.05981
RF						1.00000	0.03461	-0.01244
TCRF							1.00000	-0.14963
IMPPPL								1.00000

Table 76: matrice des corrélations simples sur les données stratifiées 96

	CON	TCEP	DCIND	CCDECE	CCVIE	FSIN	IMPPPL	AUTPPL
CON	1.00000	-0.81775	0.08082	-0.23235	0.14307	0.13241	0.02654	0.09492
TCEP		1.00000	-0.07427	0.28760	-0.14094	-0.15338	-0.00115	-0.26463
DCIND			1.00000	-0.13125	-0.15621	0.05381	0.40910	-0.24179
CCDECE				1.00000	-0.20484	0.33498	-0.13106	-0.10818
CCVIE					1.00000	-0.02437	-0.30720	0.42347
FSIN						1.00000	-0.19483	0.19311
IMPPPL							1.00000	-0.32938
AUTPPL								1.00000

Table 77: matrice des corrélations simples sur les données stratifiées 97

	TCCP	TCEP	CUCPP	TCF	REXEP	RFPL	PARTPPL
TCCP	1.00000	0.11064	-0.14985	0.04626	0.29423	0.12250	-0.17848
TCEP		1.00000	-0.23025	0.31686	0.26419	-0.48099	0.55280
CUCPP			1.00000	-0.16429	-0.19706	0.20913	-0.21768
TCF				1.00000	0.58569	-0.18079	0.35948
REXEP					1.00000	0.13575	-0.17285
RFPL						1.00000	-0.75002
PARTPPL							1.00000

Table 78: matrice des corrélations simples sur les données stratifiées 98

Appendice IV : corrélations simples sur les données pondérées après correction

	TS	TCH	CAPI	RF	TCF	REXPP	RFPL
TS	1.00000	0.45601	-0.08173	-0.02630	-0.12878	-0.07681	0.10107
TCH		1.00000	-0.20065	-0.11603	-0.15531	-0.46352	0.25789
CAPI			1.00000	-0.03873	0.23706	0.09948	0.01076
RF				1.00000	-0.00422	0.05528	0.02177
TCF					1.00000	0.05730	-0.16046
REXPP						1.00000	0.23179
RFPL							1.00000

Table 79: matrice des corrélations simples sur les données 92

	TCH	CCACAV	CAPI	AAFF	RF	TCE	FSIN	FPR	FRF	RSOL	EXPP	EXEP
TCH	1.000	-0.166	-0.163	0.612	-0.143	0.057	0.174	0.307	0.024	0.156	-0.712	-0.594
CCACAV		1.000	-0.112	-0.070	0.116	0.035	0.043	-0.059	-0.056	0.026	0.017	-0.050
CAPI			1.000	-0.089	-0.041	-0.019	0.070	0.002	-0.039	-0.033	0.043	0.048
AAFF				1.000	-0.039	-0.233	-0.212	-0.468	-0.254	-0.206	-0.516	-0.654
RF					1.000	-0.111	-0.154	-0.122	-0.193	-0.051	0.124	0.046
TCE						1.000	0.405	0.392	0.671	0.660	0.086	0.034
FSIN							1.000	0.583	0.640	0.683	-0.147	-0.103
FPR								1.000	0.529	0.556	-0.184	0.058
FRF									1.000	0.918	.114	0.071
RSOL										1.000	0.024	-0.023
EXPP											1.000	0.741
EXEP												1.000

Table 80: matrice des corrélations simples sur les données 93

	TCP	TCOM	RF	TCRF	TCEN	FSIN	RSOL	REXPP	REXEP	RFPL
TCP	1.00000	-0.00961	0.03944	-0.02616	0.02522	-0.09479	-0.04002	0.13725	0.10744	-0.11160
TCOM		1.00000	-0.12831	-0.02239	0.03172	-0.11638	-0.02145	-0.00109	-0.03814	0.00051
RF			1.00000	0.12711	-0.04376	-0.12387	-0.11531	0.07143	0.09338	0.14242
TCRF				1.00000	-0.07488	-0.25768	-0.30999	0.35081	0.47362	0.90987
TCEN					1.00000	0.51650	0.71134	-0.39294	-0.35132	-0.17535
FSIN						1.00000	0.81311	-0.67839	-0.63352	-0.49476
RSOL							1.00000	-0.75192	-0.74343	-0.29542
REXPP								1.00000	0.88003	0.30067
REXEP									1.00000	0.39308
RFPL										1.00000

Table 81: matrice des corrélations simples sur les données 94

	CON	TS	TCEP	CUCPU	CCVI	TCE	RSOL	PPPL	RFPL	PRE	IMM
CON	1.000	0.071	-0.827	0.076	0.119	0.039	0.023	0.057	0.052	0.005	0.098
TS		1.000	0.023	0.238	-0.045	-0.067	-0.042	0.017	0.031	-0.047	0.111
TCEP			1.000	-0.06	-0.084	-0.116	-0.042	-0.052	-0.045	-0.009	-0.059
CUCPU				1.000	-0.075	0.008	0.073	0.881	0.913	-0.090	-0.016
CCVI					1.000	0.001	0.007	-0.067	-0.039	-0.154	-0.115
TCE						1.000	0.818	-0.026	-0.014	0.020	-0.097
RSOL							1.000	0.033	0.045	0.041	-0.082
PPPL								1.000	0.903	-0.156	0.022
RFPL									1.000	-0.120	0.016
PRE										1.000	-0.005
IMM											1.000

Table 82: matrice des corrélations simples sur les données 95

	TS	TCS	FADM	ACH	CPU	CCVI	TCF	RSOL	TEC	EXE	PRE	IMM
TS	1.000	-0.003	0.908	0.965	0.019	-0.076	0.400	0.012	-0.016	0.025	0.042	0.036
TCS		1.000	-0.026	-0.000	0.331	-0.017	-0.040	-0.013	-0.190	-0.178	-0.055	-0.017
FADM			1.000	0.955	-0.037	0.053	0.457	0.027	0.141	0.200	0.086	0.030
ACH				1.000	-0.014	-0.057	0.427	0.018	0.002	0.073	0.092	0.020
CPU					1.000	-0.079	-0.041	-0.051	0.262	0.300	-0.159	-0.062
CCVI						1.000	-0.102	-0.086	-0.017	-0.043	-0.057	-0.104
TCF							1.000	-0.001	0.149	0.138	-0.010	0.119
RSOL								1.000	0.105	0.237	0.152	0.173
TEC									1.000	0.872	-0.175	-0.052
EXE										1.000	-0.016	-0.063
PRE											1.000	-0.023
IMM												1.000

Table 83: matrice des corrélations simples sur les données 96

	CON	TCEP	CAPIPU	CDECE	CLVIE	TCRF	FSIN	RSOL	PPLPL
CON	1.00000	-0.75702	0.11186	-0.22342	0.03720	-0.10413	-0.15641	-0.18284	0.04876
TCEP		1.00000	-0.10618	0.22340	-0.09427	0.06609	0.09862	0.12537	-0.01093
CAPIPU			1.00000	-0.14589	-0.04963	-0.00999	-0.06897	-0.06730	-0.02453
CCDECE				1.00000	-0.09338	-0.10592	0.47048	0.46370	-0.06139
CCVIE					1.00000	0.03537	-0.08701	-0.10638	-0.09392
TCRF						1.00000	-0.08418	0.00047	-0.07916
FSIN							1.00000	0.30162	0.01504
RSOL								1.00000	0.06810
PPLPL									1.00000

Table 84: matrice des corrélations simples sur les données 97

	TCEP	CUCPU	CCDE	CCVI	RF	FSIN	TECP	EXEP	PPPL	RFPL	IMM
TCEP	1.000	-0.231	0.297	-0.039	-0.179	-0.017	-0.056	-0.046	-0.012	-0.043	0.033
CUCPU		1.000	-0.232	-0.195	0.030	-0.175	0.042	-0.097	0.054	-0.056	-0.010
CCDE			1.000	-0.062	-0.111	0.442	0.071	0.227	-0.062	-0.049	-0.007
CCVI				1.000	-0.115	-0.006	-0.201	-0.031	-0.113	-0.072	-0.095
RF					1.000	-0.118	0.063	-0.037	-0.015	0.005	0.036
FSIN						1.000	-0.051	0.666	0.041	0.033	-0.092
TECP							1.000	0.424	0.038	0.044	0.064
EXEP								1.000	0.056	0.055	-0.023
PPPL									1.000	0.904	-0.032
RFPL										1.000	-0.055
IMM											1.000

Table 85: matrice des corrélations simples sur les données 98

References

- [1] B.K. Atchinson, (1997), "Remarks on the American Risk-Based Capital Model". The Geneva Papers on Risk and Insurance, Vol. 22, No 82.
- [2] Ambrose J. M. and A.M. Carroll, (1994), "Using Best's Ratings in Life Insurer Insolvency Prediction". The Journal of Risk and Insurance, Vol. 61, No 2.
- [3] Ambrose J. M. and J.A. Seward, (1988), "Best's Ratings, Financial Ratios and Prior Probabilities in Insolvency Prediction". The Journal of Risk and Insurance, Vol. 55.
- [4] Anderson J., (1972), "Separate Sample Logistic Discrimination". Biometrika, vol. 59, No. 1, p. 19-35.
- [5] B.K. Atchinson, (1997), "Remarks on the American Risk-Based Capital Model". The Geneva Papers on Risk and Insurance, Vol. 22, No 82.
- [6] AXA, Befec-Price Waterhouse, (1994), "Gestion et Analyse Financière". L'ARGUS Editions, Collection Assurances: Comptabilité, Gestion, Finance.
- [7] Baranoff E. G., T. W. Sager and R. C. Witt, (1999), "Industry Segmentation and Predictor Motifs for Solvency Analysis of the Life/Health Insurance Industry". The Journal of Risk and Insurance, Vol. 66, No. 1, p. 99-123.
- [8] BarNiv R. and J. B. McDonald, (1999), "Review of Categorical Models for Classification Issues in Accounting and Finance". Review of Quantitative Finance and Accounting, Vol. 13, p. 39-62.
- [9] BarNiv R. and J. Hathorn, (1997), "The Merger or Insolvency Alternative in the Insurance Industry". The Journal of Risk and Insurance, Vol. 64, No. 1.
- [10] BarNiv R. and J. B. McDonald, (1992), "Identifying Financial Distress in the Insurance Industry : A Synthesis of Methodological and Empirical Issues". The Journal of Risk and Insurance, Vol. LIX, No. 4.
- [11] BarNiv R. and R. A. Hershbarger, (1990), "Classifying Financial Distress in the Life Insurance Industry". The Journal of Risk and Insurance.
- [12] BarNiv R. and A. Raveh, (1989), "Identifying Financial Distress : A New Nonparametric Approach". Journal of Business Finance and Accounting, Vol. 16, No. 3.
- [13] Browne M. J. and R. E. Hoyt, (1995), "Economic and Market Predictors of Insolvencies in the Property-Liability Insurance Industry". The Journal of Risk and Insurance, Vol. 62, No. 2.
- [14] Cosslett S.R., (1981), "Maximum Likelihood Estimator for Choice-Based Samples". Econometrica, Vol. 49, No. 5, (September), p. 1288-1316.
- [15] Cummins J. D., S. E. Harrington and R. Klein, (1995), "Insolvency Experience, Risk-Based Capital, and Prompt Corrective Action in Property-Liability Insurance". Journal of Banking and Finance, No 19.
- [16] Cummins J. D. and R. Derrig (diteurs), (1991), "Managing the Insolvency Risk of Insurance Companies". Kluwer Academic Publishers, 1991.

- [17] Daykin C. D., T. Pentikainen and M. Pesonen, (1996), "Practical Risk Theory for Actuaries". Chapman & Hall.
- [18] Gouriéroux C., and A. Monfort, (1989), "Econometrics Based on Endogenous Samples". Working Paper (8903), CREST, INSEE.
- [19] Grace M. F., S. E. Harrington and R. W Klein, (1998), "Risk-Based Capital and Solvency Screening in Property-Liability Insurance : Hypotheses and Empirical Tests". The Journal of Risk and Insurance, Vol. 65, No. 2, p. 213-243.
- [20] Hooker N. D., J.R. Bulmer, S.M. Cooper and P.H. Hinton, (1996), "Risk-Based Capital in General Insurance". British Actuarial Journal, Vol II, No 2.
- [21] Hsieh D. A., C. F. Manski, and D. McFadden, (1985), "Estimation of Response Probabilities from Augmented Retrospective Observations". Journal of the American Statistical Association, Vol. 80, No. 391, (September), p. 651-662.
- [22] Hsini M.R., (2001), "Estimation from Endogenously Stratified Samples: Alternative Methods". In progress.
- [23] Hsini M.R., (1999), "Appréciation Econométrique de la Solvabilité des Sociétés d'Assurance Non-Vie". Cahiers du CEREMADE, No. 9936.
- [24] Kim Y. D., D. R. Anderson, T. L. Amburgey and J. C. Hickman, (1995), "The Use of Event History Analysis To Examine Insurer Insolvencies". The Journal of Risk and Insurance, Vol. 62, No. 1.
- [25] King G., L. Zeng, (2001), "Logistic Regression in Rare Events Data". Political Analysis, Vol. 9, No. 2, p. 137-163.
- [26] Lee S. H. and J. L. Urritia, (1996), "Analysis and Prediction of Insolvency in the Property-Liability Insurance Industry : A Comparison of Logit and Hazard Models". The Journal of Risk and Insurance, Vol. 63, No. 1.
- [27] Lollivier S., M. Marpsat, D. Verger, (1996), "L'économétrie et l'étude des comportements: Présentation et mise en oeuvre de modèles de régression qualitatifs". INSEE: Série des documents de travail "Méthodologie Statistique", No. 9606.
- [28] Maddala G. S., (1991), "A Perspective on the Use of Limited-Dependent and Qualitative Variables Models in Accounting Research". The Accounting Review, Vol. 66, No. 4 (October), p. 788-807.
- [29] Manski C. F. and S. R. Lerman, (1977), "The Estimation of Choice Probabilities from Choice-Based Sample". Econometrica, Vol. 45, No. 8 (November), p. 1977-1988.
- [30] Manski C.F., and D. McFadden, (1981), "Alternative Estimators and Sample Designs for Discrete Choice Analysis". In C.F. Manski and D. McFadden, eds. "Structural Analysis of Discrete Data with Econometric Applications". Cambridge, Massachusetts : M.I.T. Press
- [31] Muller E. and M. Reischel, (1996), "The Theoretical Considerations Underlying the Risk Reserve Process and the Practical Assessment and Control of the Capital Required (Risk-Based Capital)". The Geneva Papers on Risk and Insurance, Vol. 21, No. 78.
- [32] Petauton P., (1996), "Théorie et Pratique de l'Assurance Vie". Dunod (2ième édition).

- [33] Pottier S. W., (1998), "Life Insurer Financial Distress, Best's Ratings and Financial Ratios". *The Journal of Risk and Insurance*, Vol. 65, No. 2, p. 275-288.
- [34] Prentice R.L. and R. Pyke, (1979), "Logistic Disease Incidence Models and Case-Control Studies". *Biometrika*, Vol. 66, No. 3, p. 403-411.
- [35] Qin J. and B. Zhang, (1997), "A Goodness-of fit Test for Logistic Regression Models Based on Case-Control Data". *Biometrika*, Vol. 83, No. 3, p. 609-618.
- [36] Qin J. and J. Lawless, (1994), "Empirical Likelihood and General Estimating Equations". *The Annals of Statistics*, Vol. 22, No. 1, p. 300-325.