



HAL
open science

Prévision de la défaillance et réseaux de neurones : l'apport des méthodes numériques de sélection de variables

Philippe Du Jardin

► **To cite this version:**

Philippe Du Jardin. Prévision de la défaillance et réseaux de neurones : l'apport des méthodes numériques de sélection de variables. domain_other. Université Nice Sophia Antipolis, 2007. Français. NNT: . tel-00475200

HAL Id: tel-00475200

<https://theses.hal.science/tel-00475200v1>

Submitted on 21 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE NICE – SOPHIA-ANTIPOLIS

INSTITUT D'ADMINISTRATION DES ENTREPRISES

Thèse

Pour l'obtention du titre de

Docteur en Sciences de Gestion

Présentée par

Philippe du Jardin

PRÉVISION DE LA DÉFAILLANCE
ET RÉSEAUX DE NEURONES :
L'APPORT DES MÉTHODES
NUMÉRIQUES DE SÉLECTION
DE VARIABLES

Soutenue le 15 décembre 2007

devant le jury composé de :

M. Patrick Boisselier

Professeur à l'IAE de Nice

Directeur

M. Jean-Louis Chandon

Professeur à l'IAE d'Aix-en-Provence

Rapporteur

M. Gérard Dreyfus

Professeur à l'École Supérieure de Physique et de Chimie Industrielles

Suffragant

M. Éric Séverin

Professeur à l'IAE de Valenciennes

Rapporteur

M. Robert Teller

Professeur à l'IAE de Nice

Suffragant

L'Université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

REMERCIEMENTS

Il y a cinq ou six ans, Patrick Boisselier débutait son travail de sape pour me convaincre qu'il était temps que je m'y mette. Pour avoir trouvé les mots qui m'ont conduit aujourd'hui à écrire ces lignes, il faut l'en remercier, mais ce n'est pas tout ;

Dominique Dufour, qui travaillait déjà sur la défaillance à cette époque, m'a permis d'entrevoir le sujet de cette thèse et a aplani bien des difficultés en me fournissant les données sur lesquelles j'allais travailler. Pour ceci, qu'il en soit remercié, mais ce n'est pas tout ;

Patrick Boisselier, encore lui, pour m'avoir consacré plus que du temps tout au long de ces années, pour ses conseils et observations toujours justes, qu'il en soit remercié, mais ce n'est pas encore tout pour lui. Pour son amitié, sa présence et son soutien, parce que cela n'est pas rien, il doit en être remercié, mais ce n'est toujours pas tout ;

MM. les professeurs Jean-Louis Chandon, Gérard Dreyfus, Eric Séverin et Robert Teller, pour avoir bien voulu accepter de faire partie du jury de cette thèse, qu'ils en soient vivement remerciés, mais ce n'est pas tout ;

M. Bernard Olivero, professeur à l'IAE de Nice, et l'école doctorale de Nice, pour leur aide méthodologique, mais aussi pour la pression qu'ils ont su exercer sur le respect des échéances, qu'ils en soient remerciés, mais ce n'est pas tout ;

M. Olivier Oger, pour m'avoir octroyé du temps et des moyens à l'Edhec, qu'il en soit remercié, mais c'est encore loin d'être tout ;

M. Jean-Louis Chandon, cette fois pour les cours qu'il m'a dispensés il y a déjà bien longtemps, qui m'ont fait « tomber » dans l'analyse de données et découvrir les réseaux de neurones, et sans qui cette thèse ne serait pas ce qu'elle est, qu'il en soit remercié ;

Michèle, pour sa vision pointue des ratios, pour les ouvrages qu'elle m'a prêtés, et que je n'ai toujours pas rendus, et pour tout le reste, qu'elle en soit remerciée, mais ce n'est pas tout ;

Régine, pour m'avoir permis de trouver l'introuvable et d'accéder à l'inaccessible grâce à son réseau de documentalistes, qu'elle en soit vivement remerciée, mais ce n'est pas tout ;

Hervé, qui a eu la lumineuse idée de me commander un portable, m'a permis d'écrire en toute tranquillité les lignes de codes dont j'ai eu besoin pour mes calculs, qu'il en soit alors grandement remercié ;

Bill Gates et VBA, sans qui je n'aurais pu réaliser une grande partie des calculs présentés dans cette thèse, sont ici remerciés, mais ce n'est pas tout ;

Stéphanie et Lisa, pour m'avoir donné accès à bon nombre de codes sources en C++ d'une grande variété d'algorithmes neuronaux, m'ont fait gagner un temps précieux, qu'elles en soient remerciées ;

Alexandra, Anne et Clémentine, avec qui j'ai partagé mon quotidien pendant cette dernière année de travail, pour l'aide qu'elles ont su m'apporter et qui m'a permis de consacrer autant de temps que je l'ai souhaité à l'écriture, qu'elles en soient chaleureusement remerciées, mais ce n'est pas tout ;

Georges, pour avoir transformé un PDF en quelques kilos de papier imprimés avec toujours autant de précision, est ici remercié, mais ce n'est pas tout ;

Car tout ceci n'est rien encore :

Elisabeth, par tout ce qui ne peut être écrit ici et plus encore, a su ne pas me tenir rigueur d'avoir passé plus de temps avec des réseaux de neurones qu'avec elle, et ceci vaut plus que tout*.

* Enfin, je ne saurais oublier de remercier Yves Tabourier, dont l'ouvrage « De l'autre côté de Merise » m'a inspiré cette façon de rédiger mes remerciements.

SOMMAIRE

INTRODUCTION	4
PARTIE 1 : LE RÔLE DES RÉSEAUX DE NEURONES DANS LA PRÉVISION DE LA DÉFAILLANCE	11
CHAPITRE 1 : PRÉVISION DE LA DÉFAILLANCE – UN ÉTAT DE L'ART	12
CHAPITRE 2 : CONSTRUCTION DE MODÈLES DE PRÉVISION À L'AIDE DE RÉSEAUX DE NEURONES	84
CHAPITRE 3 : MÉTHODES DE SÉLECTION DE VARIABLES	117
PARTIE 2 : LES MÉTHODES DE SÉLECTION DE VARIABLES ET LES MODÈLES NEURONAUX DE DÉFAILLANCE	135
CHAPITRE 4 : MÉTHODOLOGIE	136
CHAPITRE 5 : STRUCTURES ET PERFORMANCES DES MODÈLES	158
CHAPITRE 6 : APPORT DES MODÈLES À LA COMPRÉHENSION DE LA DÉFAILLANCE	182
CONCLUSION	208
ANNEXES	212
BIBLIOGRAPHIE	219
TABLE DES MATIÈRES	237

INTRODUCTION

La défaillance d'entreprise a fait l'objet d'un nombre considérable de travaux de recherche tentant de mettre en évidence les différents déterminants pouvant affecter l'existence des firmes. La variété des modèles utilisés, mais aussi des cadres théoriques de référence, illustre l'absence de consensus sur la façon d'appréhender ce phénomène et d'en proposer une représentation intelligible mais surtout les difficultés à formuler un modèle général d'interprétation. Une telle variété souligne aussi deux logiques d'études qui ont présidé aux travaux de modélisation. La première, qui puise ses racines dans la pensée stratégique, managériale, organisationnelle et économique, a tenté de construire des modèles de vulnérabilité ou de risque tendant plutôt à rendre compte des causes et des processus conduisant à la défaillance qu'à prévoir celle-ci et son horizon. La seconde, d'inspiration financière, a plutôt voulu s'inscrire dans une perspective de prévision de la défaillance et de son stade ultime, la faillite, que dans une démarche de formalisation du cheminement qui y conduit.

Bien que cette seconde voie domine largement la recherche depuis les années 30, elle semble avoir centré tous ses efforts dans la multiplication des modèles statistiques. Les définitions couramment admises de la défaillance, comme les hypothèses qui sous-tendent les méthodes de prévision – primauté de la recherche des causes trouvant leur traduction dans les documents comptables et financiers, existence d'un lien entre rentabilité, endettement et défaillance –, ont effectivement conduit les efforts de modélisation à faire massivement appel à une classe de techniques particulières issue de l'analyse de données. Parce que la défaillance a souvent été définie du point de vue du droit et appréhendée au travers de faits juridiques, comme l'ont justement fait remarqué Balcaen et Ooghe (2006), la prévision dans ce domaine est depuis longtemps considérée comme réductible à un problème de classification. En effet, quel que soit le critère employé pour caractériser juridiquement l'état de défaillance – liquidation ou règlement judiciaire en France, liquidation ou réorganisation en Grande-Bretagne, liquidation ou redressement en Allemagne, demande de mise sous protection de la loi sur la faillite au titre du chapitre 7 ou du chapitre 11 aux Etats-Unis ... –, celui-ci s'incarne en général dans un événement qui peut être daté et qui va traduire un « état » financier. Dès lors que l'on cherche à caractériser cet état particulier, par opposition à un état de normalité, et à prévoir l'appartenance d'une entreprise à l'un ou à l'autre, on en revient à chercher une règle de classement. On comprend alors que le recours aux techniques susceptibles d'opérer des

classifications ait particulièrement fait florès. C'est précisément ce schéma qui a servi de fondement à un vaste courant de recherche qui trouve ses origines dans les travaux de Beaver (1966) et Altman (1968).

Deux grandes étapes ont marqué l'histoire de ce type de modèle de prévision.

La première débute dans les années 30, avec notamment les travaux de Ramster (1931), Fitzpatrick (1932), Winakor et Smith (1935) et Merwin (1942), et constitue le fondement de la contribution des ratios à la problématique d'étude de la défaillance en montrant leur capacité à refléter le degré de santé financière d'une entreprise. Il faudra attendre Beaver (1966) pour obtenir la toute première analyse univariée des ratios, proposant une première justification statistique de leur capacité à rendre compte de la défaillance. Un peu plus tard, Altman (1968) introduira l'analyse multivariée, avec l'analyse discriminante linéaire, dans l'optique de construire un véritable modèle de prévision. Cette façon de procéder fut ensuite largement reprise par Deakin (1972), Edminster (1972), Blum (1974), Altman et al. (1977), El Hennawy et Morris (1983), Karels et Prakash (1987), Laitinen (1991), Altman et al. (1995), Shirata (1998), Grice et Ingram (2001), Kuruppu et al. (2003)...

La méthode consiste à calculer un score s par entreprise, représentant une note de risque, qui s'exprime comme la combinaison linéaire d'un ensemble de variables explicatives x_i , $s = w_0 + \sum_{i=1}^n x_i w_i$, w représentant le vecteur des coefficients w_i de la fonction à estimer, et à comparer ce score à un seuil délimitant la frontière entre les entreprises saines et les entreprises susceptibles de défaillir. Selon la position du score obtenu par rapport à ce seuil, l'entreprise sera classée dans l'une ou l'autre des deux catégories. L'analyse discriminante permet d'estimer le vecteur w en recherchant la combinaison optimale de paramètres qui séparent au mieux les moyennes du score dans les groupes ; ainsi, avec \bar{s}_1 et \bar{s}_2 représentant ces moyennes, la fonction recherchée est celle qui maximise le ratio $(\bar{s}_1 - \bar{s}_2)^2 / V(s)$, avec $V(s)$ correspondant à la moyenne pondérée des variances du score dans chacun des groupes.

Mais l'analyse discriminante a très rapidement buté sur les limites inhérentes aux modèles linéaires (indépendance des prédicteurs, aucune prise en compte des effets conjugués de plusieurs variables, acceptation implicite du phénomène de compensation des variables...) ainsi que sur le caractère très contraignant de ses conditions d'utilisation (multinormalité des variables, égalité des matrices de variance-covariance de chaque groupe, connaissance des probabilités a priori d'appartenance aux classes, absence de colinéarité entre les variables)

qui, dans bien des cas, n'étaient jamais respectées. C'est pour dépasser cette double contrainte que l'emploi d'autres techniques a alors été envisagé.

La deuxième étape débute donc dans les années 70, avec notamment les travaux de Ohlson (1980) consacrés à l'utilisation d'une méthode probabiliste, la régression logistique, qui permet, comme l'analyse discriminante, de calculer un score s , avec $s = 1/(1 + e^{-\sum_{i=1}^n x_i w_i})$, qui s'exprime ici sous la forme d'une probabilité, et qui dérive l'estimation des coefficients w_i à partir de la technique du maximum de vraisemblance. Bien que n'imposant pas d'hypothèses aussi fortes que celles requises par l'analyse discriminante sur la distribution des variables, la régression logistique n'est pas exempte pour autant de toute contrainte. Elle suppose notamment et implicitement que la distribution de la variable dépendante suive une distribution logistique, ce que rien ne garantit par ailleurs.

La prise de conscience, à la suite de cette série de travaux, du fait que la forme fonctionnelle reliant les prédicteurs à la variable à expliquer est vraisemblablement plus complexe que celle qui peut être modélisée par une fonction discriminante ou logistique a conduit beaucoup d'auteurs à envisager l'utilisation de toute une série de méthodes non paramétriques alternatives : arbre de classification, système expert, modèle discriminant non paramétrique, k plus proches voisins... et réseaux de neurones. Parmi toutes ces méthodes, les réseaux de neurones, et plus particulièrement les réseaux non bouclés à apprentissage supervisé, occupent une place prépondérante. À ceci, deux raisons peuvent être avancées. Tout d'abord, l'hypothèse de linéarité que l'on retrouve notamment dans une fonction discriminante ne tient pas (Laitinen et Laitinen, 2000), suggérant qu'il est alors pertinent de mettre en œuvre des méthodes de régression non linéaires, comme les réseaux, plus à même d'appréhender la véritable nature des relations entre la variable à expliquer et ses déterminants. Ensuite, parmi les méthodes non linéaires, les réseaux non bouclés apparaissent comme étant des approximateurs universels particulièrement parcimonieux. Ils sont en effet capables d'appréhender tout type de fonction et de réaliser des modèles de précision équivalente à ceux produits par des méthodes traditionnelles non linéaires, mais avec moins de paramètres ajustables, ou bien des modèles plus précis à partir d'un même nombre de paramètres (Dreyfus et al., 2004).

C'est donc cette double propriété qui a fait de ce type de réseau une solution efficace pour concevoir des modèles de prévision de la défaillance. Là aussi, une prévision s'opère par le

calcul d'un score que l'on compare ensuite à un seuil. Le score s est donné par la fonction suivante, pour un réseau de type *Perceptron multicouches* : $s = f(\sum_{j=1}^p w_j f(\sum_{i=1}^n x_i w_{ij}))$, avec un réseau disposant d'une couche cachée, sans « biais », et d'un neurone de sortie, w représentant la matrice de poids du réseau, f la fonction d'activation, le plus souvent représentée par une fonction logistique ou une tangente hyperbolique, n le nombre de variables et p le nombre de neurones de la couche cachée.

Pour autant, la performance d'un modèle, c'est-à-dire sa capacité prédictive, ne dépend pas que de la famille de méthode qui l'a produit. Elle dépend aussi et très étroitement des données employées pour son ajustement. Or, les travaux de recherche structurés autour de l'utilisation de réseaux de neurones réalisés jusqu'à présent, se sont appuyés essentiellement sur des variables choisies selon des considérations ne faisant pas intervenir les spécificités des réseaux. En effet, soit les variables ont été retenues parce que faisant partie des indicateurs traditionnels de l'analyse financière validés historiquement sur la base de tests univariés et rien ne garantit que cette seule appartenance à « l'analyse financière » soit suffisante pour produire des résultats optimaux dans un contexte non linéaire ; soit elles l'ont été parce qu'issues d'un processus de sélection automatique, généralement optimisé pour d'autres méthodes que les réseaux et là encore rien ne garantit l'optimalité du choix. Dans ce dernier cas, la technique employée s'appuie très souvent sur des tests permettant de sélectionner le sous-ensemble maximisant l'écart entre les centroïdes des groupes ou alors celui minimisant le ratio variance inter-groupe sur variance intra-groupe... autant de critères qui sont plus spécifiquement adaptés à l'analyse discriminante ou sur d'autres tests qui eux, le sont plus avec la régression logistique. L'analyse de la littérature que nous avons effectuée, qui porte sur 190 études et qui est présentée au chapitre suivant, illustre bien cette situation.

Les ensembles de variables ainsi retenus sont donc sous-optimaux du point de vue d'un réseau, puisque le critère de sélection qui fonde leur légitimité n'a pas forcément de sens dans ce contexte. Rien ne permet effectivement d'affirmer qu'un ensemble de prédicteurs, optimisé pour être employé avec une méthode particulière, conservera ses propriétés quand il sera utilisé avec une autre. Or, si la recherche d'un optimum dans ce domaine va presque de soi lorsqu'il s'agit de construire une règle de classification avec une méthode traditionnelle, il doit en être de même avec les réseaux de neurones.

Bien que les techniques de sélection de variables adaptées aux modèles neuronaux soient souvent mises en œuvre conjointement avec ces derniers dans des domaines très variés, comme la reconnaissance de formes, le data mining, l'imagerie... (Liu et Yu, 2005), elles le sont beaucoup moins dans le champ de la finance d'entreprise et encore moins dans celui de la prévision de défaillance à l'aide de modèles neuronaux. Lorsque c'est le cas, les expérimentations conduites le sont de manière « minimaliste », car utilisant la plupart du temps la même technique de sélection : un algorithme génétique appliqué à un nombre de variables réduit et testée sur de petits échantillons, limitant par là même la portée des enseignements qui peuvent en être tirés (Back et al., 1994, 1996 A, 1996 B, 1997 ; Wallrafen et al., 1996 ; Kiviluoto, 1998 ; Kim et al., 2003 ; Sexton et al., 2003 ; Brabazon et Keenan, 2004). Seules de très rares études ont employé des méthodes propres aux réseaux, mais sur de petits échantillons, avec peu de variables, en n'offrant jamais de comparatifs entre différentes méthodes de cette catégorie (Tyree et Long, 1996 ; Charalambous et al., 2000 ; Sen et al., 2004 ; Bose et Pal, 2006). Or, les techniques candidates, bien moins lourdes d'emploi, sont suffisamment nombreuses pour qu'il soit possible de construire des cadres d'expérimentation mettant en parallèle plusieurs techniques, afin de s'affranchir des limites posées par l'emploi d'une seule d'entre elles. Il est, à cet égard, symptomatique de constater que la plupart des études font grand cas de la nécessité de comparer des résultats issus de différents classifieurs sans pour autant tenir le même raisonnement quant aux modalités de choix des données sur lesquelles porteront les classements.

L'objet de notre travail s'inscrit donc dans cette perspective. Il vise à montrer en quoi les techniques de sélection de variables adaptées aux réseaux offrent une alternative utile à celles traditionnellement employées par les études expérimentales pour améliorer le degré de précision des modèles de défaillance, en comparant plusieurs couples « méthodes de prévision – technique de sélection ». Il vise aussi à montrer comment la détection des variables pertinentes permet d'approfondir la connaissance du problème traité au travers de l'image ou des images qu'elles peuvent renvoyer de l'état de défaillance.

Nous présenterons donc tout d'abord, dans une première partie, un état de l'art des travaux relatifs à notre problématique au travers de trois chapitres.

Le premier est consacré aux études expérimentales conduites jusqu'à présent et vise à définir le périmètre général de la nôtre ainsi que les points de comparaison avec les travaux précédents. Ce chapitre met plus particulièrement l'accent sur ce que l'on entend par

« défaillance » et sur le type de défaillance que l'on peut raisonnablement envisager de prévoir, sur les facteurs à l'origine de ce phénomène et la façon dont ces derniers peuvent se traduire dans des données analysables, et en définitive sur les variables susceptibles de rendre compte de cet événement particulier pouvant affecter la vie d'une firme.

Le deuxième chapitre, lui, est dévolu plus spécifiquement à la famille de réseaux de neurones généralement employée pour réaliser des tâches de prévision de la défaillance et aux différents algorithmes auxquels on peut avoir recours. Ce chapitre a pour objet de mettre en exergue l'intérêt de ces méthodes pour appréhender des problématiques de prévision, notamment dans des circonstances où la théorie fait défaut, et où l'on s'en remet aux données pour tenter de faire émerger des constantes, mais aussi de montrer les limites et toutes les difficultés de mise en œuvre qui les caractérisent. Il a aussi pour objet de mettre en lumière et de justifier les choix expérimentaux que nous avons effectués, et notamment ceux qui ont présidé à la sélection d'un type d'algorithme intervenant dans le processus d'apprentissage.

Le troisième chapitre est centré sur la problématique liée aux opérations de sélection de variables ainsi que sur les méthodes qui y sont associées. Il s'agit de définir les caractéristiques générales des méthodes employées en amont de beaucoup d'opérations de classification ou de régression réalisées avec des méthodes statistiques « traditionnelles », et les mettre en regard des techniques de sélection adaptées aux réseaux de neurones. Ce chapitre, lui aussi, a pour finalité de délimiter le cadre général d'expérimentation que nous avons conçu et d'argumenter les choix opérés.

Cette partie, parce qu'elle constitue une synthèse des principaux travaux conduits dans les trois principaux domaines sur lesquels repose la problématique de cette thèse, représente d'abord la base méthodologique permettant de légitimer nos principaux choix et plus particulièrement, ceux afférents aux techniques de prévision à comparer, aux critères et aux méthodes de sélection à utiliser, ainsi que ceux relatifs à la constitution des échantillons de données, au choix des variables explicatives, et à la définition des protocoles de sélection de ces mêmes variables et de validation des résultats.

Nous présenterons ensuite, dans une seconde partie, le plan d'expérience, puis les résultats qui ont été obtenus.

Le quatrième chapitre présente la façon dont les expérimentations ont été conduites, et décrit en détail les modalités d'échantillonnage et de choix des variables initiales, ainsi que le processus

de création des modèles selon les techniques de prévision et de sélection de variables retenues ici, sans oublier le mode de validation des modèles obtenus. Ce chapitre ne définit donc pas simplement les conditions dans lesquelles les modèles ont été créés, il précise aussi explicitement leur périmètre d'application et de validité.

Le cinquième chapitre analyse la structure et les performances des modèles, et notamment les performances selon la technique de sélection des variables employée, en comparant les techniques, à la fois sur des échantillons de test et des échantillons de validation. Il s'agit d'étudier les points communs entre tous les modèles en se demandant s'ils possèdent une structure de représentation de la défaillance similaire, et si celle-ci éclaire le processus qui conduit les entreprises vers le déclin. En effet, tout dans la littérature suggère implicitement que les modèles ont tous un caractère très conjoncturel et que le sens qu'ils renvoient de la défaillance est en définitive très dépendant des données dont ils sont issus. Mais aucune étude n'a été conduite avec pour objectif de comparer les résultats de différents processus de sélection de variables pour y rechercher des points de convergence et tenter de discerner des variables « fortes » qui traduisent la structure du phénomène plutôt qu'elles ne révèlent ses différentes facettes. Il s'agit aussi de mettre en lumière la nature de la relation pouvant exister entre un couple « technique de sélection de variables – technique de prévision » et un niveau de performance donné, et apprécier la contribution des modes de sélection de variables à la précision des modèles.

Le sixième et dernier chapitre présente une analyse de la façon dont chaque modèle, issu d'un couple particulier « technique de sélection de variables – technique de prévision », est susceptible de rendre compte de la défaillance au travers des profils d'entreprises qu'il permet de révéler. Il s'agit de compléter l'analyse précédente qui fournit des indications sur les dimensions sous jacentes à la défaillance, par une analyse des entreprises elles-mêmes et des catégories qu'elles forment afin, là aussi, de voir s'il est envisageable de dégager quelque constante, notamment dans ce qui pourrait révéler la présence de chemins que les entreprises empruntent lorsqu'elles déclinent et s'approchent de leur fin.

PARTIE 1 : LE RÔLE DES RÉSEAUX DE NEURONES DANS LA PRÉVISION DE LA DÉFAILLANCE

Cette première partie vise à asseoir les principaux choix méthodologiques sur lesquels s'appuient nos expérimentations. Il ne s'agit pas ici de dresser un inventaire exhaustif du contenu de toutes les publications réalisées, notamment dans le domaine de la défaillance d'entreprise. De multiples études comparatives ont été effectuées à cet égard, qui passent en revue toutes les facettes de la problématique. Il suffit de se reporter à Keasey et Watson (1991), Dimitras et al. (1996), Altman et Narayanan (1997), Wong et al. (1997), Adya et Collopy (1998), O'Leary (1998), Zhang et al. (1998), Vellido et al. (1999), Coakley et Brown (2000), Aziz et Dar (2004), Balcaen et Ooghe (2004, 2006), Ravi Kumar et Ravi (2007), pour constater que ce genre d'exercice a été abondamment pratiqué sous la forme d'un recensement des questions soulevées soit par tous les types d'études empiriques pouvant s'intéresser à la défaillance, soit uniquement par celles mettant en jeu des réseaux de neurones pour prévoir cette même défaillance. Il s'agit plutôt de discerner, au travers de la littérature, quels sont les éléments à prendre en compte, les définitions à retenir, les arbitrages à effectuer pour construire un plan d'expérience robuste.

Ce chapitre est donc consacré à la défaillance en elle-même, à ces définitions et aux conditions de son appréhension par des méthodes quantitatives, aux réseaux de neurones en tant que variété appartenant à la sphère des méthodes quantitatives, à leurs propriétés et conditions d'utilisation pour créer des modèles de prévision et enfin, aux techniques de sélection de variables considérées comme processus d'ajustement déterminant la précision de tels modèles.

CHAPITRE 1 : PRÉVISION DE LA DÉFAILLANCE – UN ÉTAT DE L'ART

Le terme de défaillance renvoie en première approximation à une réalité d'entreprise qui n'est ni simple à définir, ni simple à appréhender dès lors que l'on souhaite distinguer des firmes défaillantes de firmes qui ne le sont pas. Les définitions courantes s'appuient sur l'idée qu'une entreprise est dite défaillante si elle n'est pas en mesure de faire face à ses obligations vis-à-vis de ses débiteurs (Charreaux, 1996). Cependant, cette définition laisse en suspend le degré de gravité de cette incapacité, car elle ne remet pas systématiquement en cause le devenir de la firme, ce qui suggère que toute prévision d'incident mineur ne présente finalement qu'un intérêt limité, quel que soit le point de vue que l'on adopte : celui du créancier ou du débiteur. Elle laisse aussi planer quelques incertitudes quant à la possibilité de dater l'événement, surtout pour un observateur externe, alors que la datation d'un tel incident est indispensable si l'on souhaite caractériser un état de défaillance par opposition un état de « normalité » et dresser le profil de cet état. Ce n'est qu'en opérant des mesures avant et après la survenue d'un événement de cette nature que l'on peut opposer des profils et donc, les discriminer.

En introduisant les notions de « gravité » et de « datation », on voit bien que l'on oriente fortement le sens du terme vers une définition où il n'y a d'état de défaillance que d'état observable aisément. On voit aussi qu'il faut y ajouter des critères permettant de fixer un seuil de « gravité » ainsi qu'un horizon de temps au-delà desquels l'entreprise sera considérée comme réellement défaillante. Mais alors, quel seuil faut-il adopter ? Dans quel élément de réalité ce seuil peut-il se retrouver ? Comment dater la survenue de la défaillance ? Ces questions, précisément, seront abordées au cours de ce chapitre. Il s'agit notamment de faire le point sur les différentes variations de définition que l'on retrouve dans la littérature, afin d'en adopter une. Il s'agit aussi de discerner, parmi les états de défaillance observables, ceux qui sont raisonnablement prévisibles, donc ceux qui peuvent être discriminés avec une bonne précision. Il s'agit enfin d'analyser les variables les plus à même de porter la trace de ces états, de les restituer, et de définir les modalités de constitution des échantillons expérimentaux les plus propices à la construction d'un modèle.

1.1 DÉFINITION ET HORIZON DE LA DÉFAILLANCE

Malecot (1991) souligne que les statistiques de la défaillance s'appuient toutes sur des données relatives à des catégories dont la définition est régie par le droit car elles ne tiennent compte

finalement que des entreprises qui se trouvent assignées en justice suite au non-paiement d'une dette ou qui risquent de l'être. La défaillance désigne donc d'abord, dans son sens juridique premier, une situation où les firmes sont confrontées à des procédures judiciaires de redressement ou de liquidation, ou à des procédures similaires à l'étranger.

Ainsi, aux Etats-Unis, seules sont recensées les entreprises qui se trouvent sous le coup du *Chapter 11* du code de la faillite (*Federal Bankruptcy Code*), qui renvoie aux procédures judiciaires de redressement (*Reorganization*) ou du *Chapter 7*, qui correspond aux procédures de liquidation (*Liquidation*). Dans le premier cas, on retrouve des entreprises, solvables ou non, qui souhaitent se placer sous la protection de la loi, parce qu'elles envisagent des difficultés à honorer leurs engagements financiers dans le futur, de manière à préparer leur restructuration tout en étant préservées de la menace que représentent les créanciers. Dans le second, on trouve les firmes endettées et incapables de servir leurs dettes qui ont volontairement demandé, ou y ont été forcées par leurs créanciers, à entrer dans un processus de liquidation judiciaire.

En Allemagne, sont recensées les entreprises que l'on retrouve sous le coup de la loi sur l'insolvabilité du 5 octobre 1994 (*Insolvenzordnung*) et qui distingue, d'un côté, les entreprises se considérant comme étant surendettées ou dans une incapacité de payer « imminente », et qui sollicitent l'ouverture de la procédure d'insolvabilité, et de l'autre, celles qui sont confrontées à cette même procédure par assignation d'un créancier. Selon le degré d'insolvabilité constatée et la capacité de l'entreprise à la résorber, elle peut être, comme aux Etats-Unis, liquidée ou redressée.

En Grande-Bretagne, les firmes comptabilisées le sont dès lors qu'elles sont confrontées à l'une des procédures du *Insolvency Act*, que l'on peut regrouper en quatre ensembles. Les deux premiers désignent des procédures classiques de redressement (*Administration*) et de liquidation (*Liquidation*). Le troisième encadre un processus d'élaboration d'un concordat entre le créancier et ses débiteurs (*Company Voluntary Arrangement*) et le dernier, propre à la législation britannique, désigne un protocole (*Administrative Receivership*) spécialement conçu pour protéger les droits des créanciers. Ici aussi, ces procédures visent soit à préserver les entreprises incapables d'honorer leurs dettes (*Administration*) ou à faciliter un arrangement avec les créanciers (*Company Voluntary Arrangement*), soit à les liquider (*Compulsory Liquidation*) en cas de cessation de paiement ou à faciliter le recouvrement d'une dette.

En France, on retrouve dans les statistiques, les entreprises en cessation de paiement (cessation constatée par un tribunal à la demande de l'entreprise, d'un créancier ou du tribunal lui-même) qui ont subi soit un redressement destiné à sauvegarder l'entreprise et à apurer le passif, *via* un plan de continuation de l'activité ou une cession à un tiers, soit une liquidation prononcée dès lors que le redressement s'avère impossible. La réforme de 2005 du droit des procédures collectives, par la loi n° 2005-854 du 26 juillet de la même année sur la sauvegarde des entreprises, qui s'inspire du *Chapter 11* américain, donnera forcément lieu à une autre catégorie de défaillance, mais à ce jour, elle ne transparaît pas encore dans les données.

Au Japon, enfin, sont répertoriées les entreprises qui relèvent de quatre procédures, deux faisant partie de l'arsenal lié à la liquidation (*Hasan* relative à la faillite et *Tokubetsu Seisan* relative à une procédure de liquidation spéciale), et deux appartenant à celui lié à la réorganisation et à la poursuite d'activité (*Kaisha Kosei* traitant la réorganisation réservée aux entreprises cotées et *Minji Saisei* traitant le redressement judiciaire destiné aux seules petites et moyennes entreprises).

Le tableau 1 présenté ci-après résume la façon dont les cinq pays décrits ci-dessus catégorisent les entreprises défaillantes.

Tableau 1 : Catégories de défaillance déterminées par le droit selon les pays

	Etats-Unis	Grande-Bretagne	France	Allemagne	Japon
Liquidation	Chapter 7	Voluntary Liquidation Compulsory Liquidation	Liquidation	Liquidation	Faillite Liquidation spéciale
Redressement-réorganisation	Chapter 11	Administration Administration Receivership Company Voluntary Arrangement	Redressement Sauvegarde	Cession Continuation	Réorganisation Redressement judiciaire civil

Tableau inspiré de Pochet (2001).

On pourrait continuer à détailler ce comparatif international, on finirait toujours par discerner, au travers des variantes dues aux conditions historiques ayant présidé à l'élaboration des textes de droits nationaux, des catégories assez similaires, qui existent d'ailleurs depuis l'avènement du droit romain. En effet, depuis la période romaine, les procédures de traitement des faillites ont toujours été des procédures visant soit à éliminer les entreprises trop faibles pour survivre, soit à permettre le redressement de celles qui disposent encore d'un potentiel de survie. Le droit sanctionne ainsi une situation financière ; celle du débiteur qui n'est pas solvable.

Cette façon d’appréhender la défaillance est en réalité bien pratique puisqu’elle donne un critère simple à utiliser pour isoler deux populations d’entreprises ; celles qui sont sous le coup d’une procédure judiciaire et les autres. On comprend dès lors, puisque toute statistique obéit à une classification juridico-financière, que les modèles de prévision sont généralement élaborés et calibrés selon cette dichotomie. En effet, la démarche de construction étant la plupart du temps une démarche participant d’une logique inductive, les modèles sont issus de procédures d’analyse de données qui s’appuient sur ces deux catégories d’entreprises définies *a priori* ; les données sont réparties selon ces deux groupes mutuellement exclusifs à une période donnée et à l’aide d’une procédure automatisée, on cherche à caractériser le profil de chaque ensemble au travers d’une règle de discrimination, qui une fois validée, fera office de modèle de prévision.

Pour autant, cette définition n’est pas exempte de toute critique. En effet, les modèles dits de prévision de la défaillance élaborés selon cette typologie, sont avant tout des modèles de prévision de la cessation de paiement, puisque l’ouverture d’une procédure judiciaire est souvent conditionnée à la survenue de cette occurrence ou à son anticipation par l’entreprise elle-même ou par un créancier. Ils ne permettent donc pas une réelle appréhension d’autres types de défaillance que l’on pourrait légitimement envisager de prévoir. Cette définition pose aussi la question du moment où la défaillance s’est révélée. Une date de jugement de liquidation, par exemple, ne désigne pas le moment, ou l’intervalle de temps précis au cours duquel l’entreprise s’est effectivement trouvée dans un état problématique, qui peut être bien antérieure à la date officielle. Elle pose enfin la question de la fonction même de tels modèles : doivent-ils permettre de prévoir uniquement la survenue du stade ultime de la défaillance dans une logique de prévision de ce risque, ou plutôt celle des signes annonciateurs d’une telle issue dans une optique de prévention de ce même risque ?

1.1.1 MODÈLES DE CESSATION DE PAIEMENT

La première critique a été émise à plusieurs reprises. Les modèles de défaillance étant d’abord des modèles de faillite (*bankruptcy*), quelques auteurs ont tenté de prévoir d’autres types d’issues qui participent aussi de la défaillance financière. Le tableau 2 ci-après répertorie, parmi les 190 études que nous avons examinées, celles dont l’objet consistait à modéliser des règles de décision capable de prévoir des situations autres que la faillite ou la non faillite,

comme le non remboursement d'une dette, le non paiement de dividendes, la détresse financière... Pour chaque étude, nous donnons les catégories de prévision prises en compte.

Tableau 2 : Études dont l'objectif consiste à tenter de discriminer d'autres catégories d'entreprises que les catégories habituelles : « saine » versus « défailante » ou « en faillite »

Études	États pris en compte pour les prévisions	Études	États pris en compte pour les prévisions
AGARWAL et al. (2001)	<ul style="list-style-type: none"> • Non défailance • Réduction de dividendes • Défaut de paiement • Faillite 	KUMAR et al. (1997)	<ul style="list-style-type: none"> • Liquidation • Réorganisation
ALAM et al. (2000)	<ul style="list-style-type: none"> • Sur-performance • Non défailance • Faillite 	KURUPPU et al. (2003)	<ul style="list-style-type: none"> • Mauvaise santé • Liquidation
ALTMAN (1994)	<ul style="list-style-type: none"> • Non défailance • Vulnérabilité • Défaut de paiement 	LAU (1987)	<ul style="list-style-type: none"> • Non défailance • Réduction de dividendes • Défaut de paiement • Mise sous protection des lois sur la faillite • Faillite
ANANDARAJAN et al. (2001)	<ul style="list-style-type: none"> • Défaillance • Faillite 	PEEL et PEEL (1987)	<ul style="list-style-type: none"> • Non défailance et profits • Non défailance et pertes • Défaillance
BARANOFF et al. (2000)	<ul style="list-style-type: none"> • Non défailance • Détresse 	PLATT et PLATT (2002)	<ul style="list-style-type: none"> • Non défailance • Détresse
BARNIV et al. (1997)	<ul style="list-style-type: none"> • Acquisition • Restructuration • Liquidation 	RAHMAN et al. (2004)	<ul style="list-style-type: none"> • Non défailance • Problématique
BARNIV et al. (2002)	<ul style="list-style-type: none"> • Acquisition • Restructuration • Liquidation 	SINKEY (1975)	<ul style="list-style-type: none"> • Non problématique • Problématique
BOSE et PAL (2006)	<ul style="list-style-type: none"> • Non défailance • Mauvaise santé financière 	SUN et al. (2005)	<ul style="list-style-type: none"> • Mauvaise santé • Faillite
GILBERT et al. (1990)	<ul style="list-style-type: none"> • Non défailance • Défaillance • Faillite 	WILSON et al. (1995)	<ul style="list-style-type: none"> • Survie • Problématique • Défaillance
GOSS et RAMCHANDANI (1995)	<ul style="list-style-type: none"> • Solvabilité • Insolvabilité 	ZURADA et al. (1998)	<ul style="list-style-type: none"> • Non défailance • Réduction de dividendes • Défaut de paiement • Faillite
JONES et HENSHER (2004)	<ul style="list-style-type: none"> • Non défailance • Insolvabilité • Faillite 		

L'examen de ce tableau révèle différentes stratégies de discrimination. Certaines études tentent de discerner plus de deux états représentatifs de la santé financière des firmes étudiées. On retrouve soit des catégories purement juridiques, c'est-à-dire des catégories correspondant à l'issue d'une procédure judiciaire comme la « liquidation », la « réorganisation », la « restructuration », « l'acquisition » par une autre firme ou la « mise sous protection des lois sur la faillite », soit des catégories financières, désignant des événements affectant l'activité économique-financière objective des firmes, comme la « réduction de dividendes » ou « le défaut de paiement ». D'autres études s'attachent plutôt à tenter de repérer des groupes constitués à partir d'un jugement formulé à l'encontre de la situation financière de l'entreprise,

comme la « détresse », « l'insolvabilité », la « vulnérabilité », la « mauvaise santé », la « sur performance » ou l'aspect « problématique » des comptes¹. En définitive, la proportion d'études qui a expérimenté d'autres voies de catégorisation que les voies traditionnelles demeure relativement faible ; seules 21 sur les 190 analysées ont procédé ainsi. Et les résultats de ces expérimentations ne sont pas probants car il est extrêmement difficile de séparer différentes variétés de défaillances ou tout simplement différents profils de structures financières jugés a priori comme étant plus ou moins sains.

Ainsi, les travaux tentant de prédire l'issue d'une procédure judiciaire parmi plusieurs possibles (liquidation, acquisition, restructuration), comme ceux de Barniv et al. (1997), Barniv et al. (2002) ou Kumar et al. (1997), donnent des taux de prévision correcte variant uniquement entre moins de 50% et 70%. Ceux qui ont analysé la possibilité de discriminer selon le type d'événement pouvant affecter l'activité d'une entreprise (réduction de dividendes, défaut de paiement, faillite), comme Lau (1987), Zurada et al. (1998), parviennent à bien isoler les entreprises saines de toutes les autres, mais ne peuvent pas faire de même lorsqu'il s'agit de séparer une firme ayant renoncé à distribuer des dividendes, d'une firme ayant connu un incident de paiement ou une faillite, les taux de classement étant particulièrement mauvais. En revanche, Agarwal et al. (2001) obtiennent, pour 4 des 5 catégories envisagées, des taux légèrement supérieurs à 80%, et Anandarajan et al. (2001) qui ont cherché à discriminer entre des entreprises défaillantes juridiquement et des entreprises déclarées en faillite, parviennent à des taux supérieurs à 93 % pour les deux groupes. D'autres encore, qui se sont consacrés à l'étude de la séparabilité d'entreprises jugées *a priori* financièrement saines ou non, à l'image de Bose et Pal (2006), Goss et Ramchandani (1995), Sinkey (1975), obtiennent des taux de prévision correcte oscillant entre 65% et 75%. Et lorsqu'il s'agit de repérer parmi un ensemble d'entreprises saines, celles qui font des profits de celles qui n'en font pas, comme l'ont fait Peel et Peel (1987), les résultats montrent que de telles classes ne sont pas séparables. Enfin, l'expérience de Platt et Platt (2002) montre qu'il est possible de délimiter la frontière entre d'un côté des firmes saines et de l'autre, des firmes ayant connu soit plusieurs années de pertes, soit des diminutions dans la distribution de dividendes,

¹ Quelques études n'ont pas été retenues pour apparaître dans le tableau 2 dès lors qu'une incertitude entachait la réelle composition des groupes quant à la définition de la défaillance ; parmi elles, certaines ne s'attardent pas sur le sens retenu pour ce mot, donc ne donnent pas la définition employée, d'autres regroupent les entreprises en « faillite », donc juridiquement défaillantes, avec d'autres n'ayant pas payé de dividendes ou honoré une dette, mais sans les dénombrer ni expliciter la façon dont elles ont été repérées, d'autres encore utilisent des catégories autres que « défaillantes-non défaillantes », mais là aussi sans en donner une définition précise.

soit une restructuration majeure, avec un taux de succès de 98%, qui est à relativiser à la vue de la petitesse de l'échantillon, ne comportant que 62 firmes saines et 24 en « détresse ».

Tous ces résultats démontrent bien que les modèles de faillite, qui s'appuient sur une définition juridique et une séparation des entreprises en deux classes, demeurent en dernier ressort les plus fiables, car dans l'ensemble, ils obtiennent de bien meilleurs résultats que ceux cités précédemment, comme indiqué dans le tableau 3 ci-après², et que tenter de cerner d'autres formes de défaillance que celles actées par des procédures judiciaires, est fortement aléatoire. Ceci tient à la nature même de l'écart pouvant exister entre les profils des différents groupes d'entreprises pris en compte et à la façon dont celui-ci se creuse à mesure que l'échéance « fatale » se rapproche. En effet, alors que les valeurs prises par les indicateurs financiers (ratios) habituellement utilisés pour mesurer ces profils demeurent relativement stables au cours du temps lorsqu'ils décrivent des entreprises saines, ces mêmes indicateurs voient leurs valeurs diverger des précédentes d'autant plus que l'horizon de la procédure judiciaire se rapproche, et ce, quelle qu'en soit l'issue. Van Wymeersch (1996), en établissant le profil financier d'entreprises belges cheminant vers la faillite sur une période de cinq ans a, par exemple, constaté une diminution constante de la rentabilité, que les entreprises tentaient de compenser par certains artifices comptables (frais portés à l'actif, décalage des produits et des charges via les comptes de régularisation, recours aux produits exceptionnels, et sans doute, surévaluation des stocks), une augmentation du financement par recours à des capitaux de tiers avec une croissance importante des dettes à long terme et, juste avant la faillite, une augmentation des dettes fiscales, sociales et salariales, et une détérioration nette de la trésorerie ainsi qu'une croissance régulière du pourcentage de la valeur ajoutée absorbée par les frais fixes. Dans le même temps, tous ces paramètres demeuraient pour le moins stables chez les entreprises ne rencontrant pas de difficultés. Dès lors, il n'est pas étonnant que l'on retrouve la conséquence de ceci dans toutes les études qui s'attachent à construire des modèles de prévision à l'aide de différentes séries de données mesurées entre un à cinq ans avant la date « officielle » de défaillance ; on constate une augmentation de la précision à mesure que les données prélevées le sont dans un intervalle proche de la période de défaillance. En revanche, les écarts entre des sous-catégories d'entreprises défaillantes, voire

² Le tableau 3 donne, à titre de comparaison, les résultats qui ont été obtenus avec des réseaux de neurones par les études que nous avons analysées et qui s'appuient sur une définition juridique de la défaillance et un découpage des échantillons en deux groupes. Ne sont donnés ici que les meilleurs résultats, étant entendu que la plupart du temps, ces études ont donné lieu à de multiples expérimentations, en faisant varier toute une série de paramètres. Nous n'avons pas indiqué les résultats obtenus par les autres études ne faisant pas appel aux réseaux, car ces dernières en comportent tellement qu'il est illusoire de tenter d'en extraire un qui servirait de point de référence.

d'entreprises saines, ne sont plus suffisants pour émerger et donc être appréhendés, car noyés dans la variabilité intrinsèque des états de santé ou des profils financiers possibles.

Tableau 3 : Résultats des prévisions obtenus grâce à un réseau de neurones sur des échantillons d'entreprises constitués à partir d'une définition juridique de la défaillance

Études	Meilleurs résultats obtenus avec le réseau et les données de la période la plus récente avant la défaillance		
	Pourcentage de prévision correcte des entreprises saines	Pourcentage de prévision correcte des entreprises défaillantes	Pourcentage total de prévision correcte
AGARWAL (1999)	99,00%	93,70%	96,38%
ANANDARAJAN et al. (2001)	93,75%	95,45%	94,50%
ATIYA (2001)	90,07%	78,13%	85,50%
BACK et al. (1994)	100,00%	98,00%	99,00%
BACK et al. (1996 A)	100,00%	94,25%	97,30%
BACK et al. (1996 B)	100,00%	94,40%	97,30%
BACK et al. (1997)	74,10%	81,20%	77,70%
BAEK et CHO (2003)	80,44%	50,60%	79,39%
BELL et al. (1990)	97,70%	61,80%	94,00%
BERG (2005)	?	?	69,50%
BORITZ et KENNEDY (1995)	84,03%	74,27%	?
BRABAZON et KEENAN (2004)	82,67%	78,67%	80,67%
BROCKETT et al. (1994)	94,50%	73,30%	89,30%
CHARALAMBOUS et al. (2000)	84,21%	84,21%	84,21%
CHARITOU et al. (2004)	76,19%	90,48%	83,33%
COATS et FANT (1993)	97,90%	89,40%	95,00%
DORSEY et al. (1995)	?	?	?
ETHERIDGE et SRIRAM (1997)	92,71%	100,00%	92,85%
FAN et PALANISWAMI (2000)	?	?	66,11%
FANNING et COGGER (1994)	76,00%	94,00%	85,00%
HUANG et al. (1994)	?	?	?
JO et al. (1997)	?	?	83,79%
KIM et al. (2003)	?	?	80,50%
KIM et HAN (2003)	?	?	88,40%
KIM et MCLEOD (1999)	?	?	78,70%
KIVILUOTO (1998)	88,30%	69,50%	86,50%
KOTSIANTIS et al. (2005)	?	?	71,17%
LACHER et al. (1995)	97,20%	91,50%	94,70%
LAITINEN et KANKAANPAA (1999)	89,50%	84,20%	86,80%
LEE et al. (1996)	?	?	80,48%
LEE (2004)	83,33%	83,33%	83,33%
LEE et al. (2005)	76,19%	95,24%	85,22%
LESHNO et SPECTOR (1996)	91,00%	90,90%	90,05%
LI et GUPTA (2002)	?	?	78,15%

McKee et GREENSTEIN (2000)	86,00%	75,00%	86,00%
MIN et LEE (2004)	?	?	75,96%
MIN et LEE (2005)	85,71%	79,36%	82,54%
MIN et al. (2006)	?	?	71,72%
ODOM et SHARDA (1990)	?	?	81,81%
PENDHARKAR (2005)	?	?	?
PIRAMUTHU et al. (1998)	92,70%	85,40%	89,10%
POMPE et BILDERBEEK (2005 A)	?	?	81,00%
POMPE et BILDERBEEK (2005 B)	?	?	80,00%
POMPE et FEELDERS (1997)	?	?	72,60%
SALCHENBERGER et al. (1992)	90,00%	96,00%	92,50%
SEN et al. (2004)	?	?	?
SERRANO-CINCA (1997)	?	?	93,94%
SEXTON et al. (2003)	97,40%	95,65%	?
SHIN et al. (2005)	?	?	74,10%
TAM et KIANG (1990)	85,00%	89,00%	87,00%
TAM et KIANG (1992)	88,60%	81,80%	85,20%
TAN et DIHARDJO (2001)	92,35%	64,70%	92,23%
TANG et CHI (2005)	95,76%	62,50%	90,70%
TUNG et al. (2004)	99,31%	54,54%	73,12%
TYREE et LONG (1996)	100,00%	94,55%	97,95%
VIEIRA et al. (2003)	?	?	90,00%
WALLRAFEN et al. (1996)	?	?	?
WEST et al. (2005)	?	?	87,27%
WILSON et SHARDA (1994)	98,00%	97,50%	97,75%
WU et al. (2007)	99,40%	83,30%	
YANG et al. (1999)	90,00%	63,00%	84,00%
YANG et HARRISON (2002)	95,50%	92,37%	95,27%
YIM et MITCHELL (2002)	94,00%	80,00%	90,96%
YIM et MITCHELL (2005)	100,00%	89,00%	96,55%
ZHANG et al. (1999)	84,55%	85,45%	85,00%

Les colonnes 2, 3 et 4 du tableau indiquent respectivement le pourcentage de prévision correcte des entreprises saines, puis des entreprises défailtantes et enfin le pourcentage total de bonne prévision. Le « ? » signifie qu'aucun résultat n'est donné par le ou les auteurs.

Une définition juridique de la défaillance présente donc un intérêt statistique évident : elle permet de délimiter une fenêtre de temps au cours de laquelle l'écart entre les profils de deux populations d'entreprises est suffisamment important pour être correctement appréhendé par une méthode de discrimination. Bien sûr, cette définition oriente fortement la structure des modèles selon les dimensions privilégiées par le droit. Balcaen et Ooghe (2006) remarquent à juste titre que de tels modèles reflètent plutôt la structure de liquidité et de solvabilité des

firmes, puisque ces deux critères sont des éléments essentiels d'appréciation de leur potentiel de survie et donc, en dernier ressort, de déclenchement des procédures judiciaires à leur encontre, et estiment que ces deux dimensions ne sauraient épuiser toutes celles qu'il serait judicieux de considérer. En effet, la prise en compte explicite de ces notions dans un grand nombre de législations ne peut qu'induire des modèles ayant une propension à détecter certaines variétés de défaillance. Néanmoins, si l'on en revient à la définition la plus générale de la défaillance, qui désigne une situation où l'entreprise n'est pas capable de faire face à ses engagements, on voit bien que la liquidité et la solvabilité sont les deux notions-clés. Et si elles ne désignent effectivement pas l'ensemble des facteurs explicatifs, elles s'apparentent néanmoins à des variables fondamentales.

Au demeurant, on retrouve la prégnance de cette représentation dans les rares études qui utilisent une modélisation théorique de l'entreprise pour légitimer le choix des variables à inclure dans les modèles, depuis Beaver (1966), en passant par Blum (1974), Gentry et al. (1985, 1987), Aziz et al. (1988), Gentry et al. (2002) ou Laitinen et Laitinen (1998).

1.1.2 MODÈLES ADOSSÉS À UNE DATE DE DÉFAILLANCE

La deuxième critique, celle relative à la datation, porte sur l'assimilation de la date entérinant un événement judiciaire à la date effective de défaillance, alors que la première est forcément postérieure à la seconde. Cette critique est en partie purement formelle, puisque, quelle que soit la nature de la défaillance étudiée, se pose toujours la question relative au moment où l'entreprise quitte son état de fonctionnement « normal » et celle relative à la façon de le découvrir. En fait, ce qui importe, ce n'est pas tant de découvrir le « véritable » moment ou l'espace de temps au cours duquel la firme a failli, mais plutôt d'isoler la période où les données, qui seront employées dans la construction des modèles, doivent être prélevées. Il faut en effet pouvoir disposer d'une image restituant le plus précisément possible les difficultés de l'entreprise qui ne soit pas entachée par les implications de la procédure judiciaire que suit la firme. En filigrane de ceci transparaissent les doutes que certains auteurs ont émis quant à la fiabilité des comptes publiés à l'issue des procédures judiciaires, arguant du fait que ces derniers pouvaient être maquillés de manière à favoriser telle ou telle décision de justice. C'est pourquoi, la quasi-totalité des études adopte au moins un délai d'un an entre la date associée à une procédure et celle où les données sont mesurées. Seules quelques exceptions admettent des délais plus courts ou plus longs, comme indiqué dans le tableau 4 ci-après.

Ce délai moyen d'un an représente, d'une part, un compromis sur une façon d'évaluer l'écart réel entre le moment où l'entreprise s'est trouvée confrontée à de véritables difficultés mettant en péril son activité et celui où ces difficultés ont été officiellement actées, et d'autre part, une reconnaissance implicite du fait qu'il s'agit là de l'horizon indépassable de toute prévision dans ce domaine ; les règles utilisées pour élaborer un modèle dictent aussi ses conditions d'emploi. Ainsi, apprendre à une règle à reconnaître un phénomène à partir de mesures réalisées douze mois avant sa survenue, revient bien à lui imposer un tel délai comme réel horizon de prévision. D'ailleurs, les résultats obtenus sur des périodes qui oscillent entre un et cinq ans, donc où l'on tente de discerner à partir de données recueillies en t , si une entreprise est susceptible de défaillir en $t+1$, $t+2$, $t+3$, $t+4$ ou $t+5$, montrent qu'au delà d'un an, comme le disait Malecot (1991), on ne peut guère plus rien prévoir. L'exactitude des résultats est donc inversement proportionnelle à l'horizon de la prévision et se dégrade fortement, passé cette période d'un an.

Tableau 4 : Études dont la durée minimale prise en compte entre la date de « défaillance » et la date de mesure des données est inférieure ou supérieure à 1 an

Études	Délai le plus court	Études	Délai le plus court
COATS et FANT (1993)	Aucun délai	THEODOSSIOU (1993)	< 1 an
KUMAR et al. (1997)	Aucun délai	KIM et al. (2003)	16 mois
LESHNO et SPECTOR (1996)	Aucun délai	LENSBERG et al. (2006)	16 mois
TUNG et al. (2004)	Aucun délai	KIM et MCLEOD (1999)	< 2 ans
LEE (2004)	3 mois	NAM et JINN (2000)	< 2 ans
TAN et DIHARDJO (2001)	3 mois	POMPE et FEELDERS (1997)	< 2 ans
LANINE et VANDER VENNET (2005)	3 mois	BROCKETT et al. (1994)	2 ans
ATIYA (2001)	6 mois	FOREMAN (2003)	2 ans
PANTALONE et PLATT (1987)	6 mois	LEE et al. (2005)	2 ans
ALAM et al. (2000)	< 1 an	THEODOSSIOU (1991)	2 ans
BARNIV et al. (1997)	< 1 an	VIEIRA et al. (2003)	2 ans
BARNIV et al. (2002)	< 1 an	WEST et al. (2005)	2 ans
BORITZ et KENNEDY (1995)	< 1 an	FITZPATRICK (1932)	3 ans
SUNG et al. (1999)	< 1 an	HÄRDLE et al. (2005)	3 ans

Ici l'on rejoint les conclusions précédentes sur l'instant où l'écart entre les profils de deux populations d'entreprises est maximal : il l'est, ou plutôt devrait l'être, au moment où les firmes défaillantes sont reconnues comme telles par une décision de justice. Mais il l'est aussi vraisemblablement douze mois plus tôt, période qui correspond somme toute de manière

grossière au moment où s'est révélée, en moyenne, la situation économique qui conduira à une telle décision.

Ceci étant, la critique portant alors sur la datation ne tombe pas entièrement. Parce que la date légale de défaillance ne coïncide pas avec la date réelle, parce que certaines procédures aboutissent plus rapidement que d'autres ou parce que certaines firmes choisissent de ne plus publier de comptes selon des délais variables, dès lors qu'elles se sentent menacées juridiquement (Balcaen et Ooghe, 2006), il restera toujours une incertitude sur la définition du moment propice où il conviendrait de mesurer la situation de chaque entreprise.

1.1.3 MODÈLES DE PRÉVISION D'UN ÉTAT CRITIQUE

La troisième critique qui fait écho à une définition légale de la défaillance, ne se focalise pas à proprement parler sur le sens qu'il convient de donner au terme, ou sur celui qu'il serait le plus pertinent de retenir, mais tente plutôt de faire ressortir l'incidence de la définition retenue sur le rôle des modèles. Si l'on conçoit la défaillance comme un état critique, sanctionné par le droit, la fonction d'un modèle revient à mesurer la distance qui sépare la situation particulière d'une firme d'une situation critique de référence. On retrouve bien là, la logique dominante de tous les systèmes de prévision qui emploient essentiellement des fonctions « score » ; prévoir revient à calculer un chiffre qui délimite, pour une observation donnée, la plus ou moins grande proximité vis-à-vis de la défaillance. Cette façon de procéder est avant tout destinée aux créanciers qui souhaitent disposer d'un indicateur de risque de non recouvrement de leur argent.

Mais on peut aussi voir la défaillance comme un processus, en dehors de toute référence légale, et imaginer qu'il serait peut-être judicieux de disposer de modèles qui ne soient pas uniquement des modèles de scoring. Si l'on adopte un autre point de vue, en particulier celui de l'entreprise et de ses dirigeants, et non plus seulement celui de ses partenaires, on pourrait attendre de tels modèles qu'ils ne soient pas simplement des moyens de prévoir le pire, mais de l'éviter, et qu'ils puissent donc jouer un rôle de prévention. Ce n'est donc plus l'occurrence d'une situation qu'il conviendrait d'anticiper, mais plutôt une dynamique de comportement qui permettrait, soit de faire apparaître un processus de dégradation financière problématique qui pourrait se solder par une défaillance, soit de rattacher l'entreprise à des profils de comportements pouvant aussi y conduire (Van Caillie, 2000).

On retrouve cette logique dans quelques travaux comme ceux de Koenig (1985), sur la modélisation du « chemin de la défaillance », de Marco (1989) sur la « spirale de la défaillance », ou de Laitinen (1991), sur les processus de défaillance et les dimensions sous-jacentes, ou encore de Serrano-Cinca (1996) ou de Neophytou et Mar-Molinario (2004), sur la façon de représenter des profils là aussi de défaillance. Mais ils sont encore trop épars et manquent de liens avec les études sur le pilotage de l'entreprise pouvant offrir un cadre général d'analyse et de correction des firmes qui suivent une trajectoire résolument orientée vers de graves difficultés.

Cette dernière critique montre bien le grand déséquilibre qui existe entre deux séries de travaux plus enclins finalement à décrire qu'à expliquer, et illustre les conséquences d'une absence de cadre théorique de la défaillance.

1.2 FACTEURS DE DÉFAILLANCE

Maintenant que la définition courante de la défaillance est posée, examinons la question des facteurs qui conduisent au déclin d'une entreprise. La grande majorité des études auxquelles nous nous référons pose une hypothèse très forte sur le phénomène étudié. Elle considère en effet que toute cause menant à la défaillance peut s'incarner dans les états comptables, d'où seront prélevées les variables explicatives des modèles, donc que l'examen de ces états (qui matérialisent les symptômes du phénomène) est suffisant pour épuiser tous les facteurs conduisant à la défaillance.

Qu'en est-il exactement de ces causes ? Et quelle garantie a-t-on que l'on puisse trouver une juste correspondance entre elles et des données financières ? Quelques études menées entre les années 70 et 90 se sont intéressées aux facteurs de disparition des entreprises et à leur poids relatif, et ont tenté de manière empirique de les isoler. Nous avons retenu ces études plutôt que celles d'inspiration macro ou micro-économique, pour deux raisons. Tout d'abord, elles traitent des variables qui affectent directement l'entreprise et son activité, et non pas celles qui par ricochet, pourraient avoir une influence, ces dernières reflétant des causes qui sont, elles, à rechercher dans des analyses économiques. Ensuite, elles présentent l'avantage de dresser un panorama de critères que l'on peut aisément rapprocher des documents financiers pour étudier comment les premiers peuvent trouver ou non une possible traduction dans les seconds.

Les deux premières études en la matière, réalisées en France, ont été publiées par la Caisse nationale des marchés de l'État, en 1977, puis par Conan et Holder, en 1979. Chacune a permis de

dresser une liste des dix principales causes de défaillance, indiquées dans le tableau 5 ci-après. Parmi les dix premières causes de chaque étude, six ont en commun de renvoyer à la gestion des clients, des stocks, des concours bancaires, du chiffre d'affaires, des coûts et des prix et de l'organisation de l'entreprise. De toutes ces causes, deux sont réellement imprévisibles : le décès ou la maladie du dirigeant ainsi que l'expropriation. Sur les douze restantes, neuf sont des causes inhérentes à l'entreprise et trois proviennent de son environnement : défaillance de clients, suppression de concours bancaires et baisse conjoncturelle du chiffre d'affaires.

Ce panorama, bien que restreint quant au nombre de causes isolées, indique néanmoins que la plus grande partie des facteurs de crise est à la fois prévisible et observable d'un point de vue strictement comptable.

Tableau 5 : Principaux facteurs de défaillance selon la Caisse nationale de marchés de l'État (1977) et Conan et Holder en (1979)

Caisse nationale des marchés de l'Etat (1977)		Conan et Holder (1979)	
Facteurs de défaillance	Occurrence	Facteurs de défaillance	Occurrence
1 Défaillance de clients	21%	1 Baisse conjoncturelle du chiffre d'affaires	21%
2 Stocks trop importants	11%	2 Mauvaise organisation de l'entreprise	11%
3 Suppression de concours bancaires	11%	3 Défaillance de clients	10%
4 Frais de personnel trop élevés	10%	4 Croissance trop rapide	9%
5 Décès ou maladie du dirigeant	10%	5 Stocks trop importants	7%
6 Mauvaise comptabilité, ignorance des prix de revient	9%	6 Obsolescence de produits, erreur de diversification	5%
7 Manoeuvres malhonnêtes ou déloyales	8%	7 Mauvaise comptabilité, ignorance des prix de revient	5%
8 Expropriation	8%	8 Politique d'investissement trop ambitieuse	5%
9 Baisse conjoncturelle du chiffre d'affaires	8%	9 Obsolescence de l'outil de production	4%
10 Mauvaise organisation de l'entreprise	4%	10 Suppression de concours bancaires	3%

Tableaux tirés de Cohen (1997).

Bien après l'étude de Conan et Holder (1979), Blazy et Combier (1997) mèneront une enquête de grande ampleur auprès d'un échantillon de 800 entreprises en cessation de paiement, permettant de préciser et surtout d'étendre la gamme de facteurs de crise isolés précédemment, et chercheront à analyser la relation qui existe entre les causes et l'issue des procédures judiciaires intentées contre les firmes. Le tableau 6 ci-après indique ces facteurs et leur fréquence d'occurrence selon le type de procédure.

Tableau 6 : Principaux facteurs de défaillance selon Blazy et Combiér (1997)

Facteurs de défaillance	Issue des procédures judiciaires :	Continuation	Cession	Liquidation	Total
1 Causes accidentelles		42,70%	31,20%	46,00%	45,40%
• Accidents divers (départ du dirigeant, divorce,...)		18,80%	8,00%	20,00%	19,50%
• Litiges avec des partenaires privés		8,30%	12,80%	16,80%	16,40%
• Litiges avec des partenaires publics (redressement fiscal...)		13,50%	5,60%	9,80%	9,70%
• Malversations internes à l'entreprise		4,20%	0,80%	4,50%	4,40%
• Problèmes sociaux au sein de l'entreprise		4,20%	3,20%	2,90%	3,00%
• Escroquerie dont est victime l'entreprise		2,10%	3,20%	2,90%	2,90%
• Décès du dirigeant		0,00%	0,80%	1,80%	1,70%
• Sinistre		3,10%	2,40%	1,10%	1,20%
2 Problèmes de débouchés		38,50%	44,80%	44,70%	44,50%
• Perte tendancielle de parts de marché		16,70%	24,80%	19,50%	19,60%
• Perte brutale de clients		9,40%	15,20%	12,50%	12,50%
• Défaillance de clients importants		7,30%	6,40%	9,50%	9,40%
• Mauvaise évaluation du marché		8,30%	9,60%	9,30%	9,30%
• Produits inadaptés		1,00%	4,80%	3,60%	3,60%
• Produits obsolètes		1,00%	1,60%	1,60%	1,60%
• Prix de vente trop élevés		1,00%	1,60%	0,90%	0,90%
3 Difficultés financières		42,70%	61,60%	41,70%	42,40%
• Déficit de ressources propres		26,00%	40,00%	19,70%	20,60%
• Défaut de paiement des clients		10,40%	18,40%	15,90%	15,80%
• Refus de prêt à l'entreprise		4,20%	8,80%	9,30%	9,10%
• Allongement des délais clients		8,30%	6,40%	5,70%	5,80%
• Taux d'intérêt supportés par l'entreprise trop élevés		4,20%	12,00%	2,90%	3,30%
• Arrêt du soutien par la maison mère		0,00%	4,80%	1,80%	1,90%
• Raccourcissement des délais fournisseurs		0,00%	0,00%	0,90%	0,90%
• Arrêt des subventions à l'entreprise		0,00%	0,00%	0,70%	0,60%
• Spéculation de l'entreprise		0,00%	0,80%	0,50%	0,50%
4 Problèmes d'information et de management		30,20%	32,80%	27,00%	27,30%
• Système comptable déficient		12,50%	14,40%	12,90%	13,00%
• Incompétence		12,50%	14,40%	9,30%	9,60%
• Mécontentement de l'équipe dirigeante		4,20%	6,40%	6,30%	6,30%
• Prélèvements excessifs		7,30%	4,00%	2,30%	2,50%
• Méconnaissance des coûts de revient		6,30%	6,40%	1,60%	1,90%
• Difficultés de transmission de l'entreprise		3,10%	0,80%	0,90%	1,00%
• Mauvaise évaluation des stocks		5,20%	0,80%	0,50%	0,60%
• Provisionnements insuffisants		1,00%	0,00%	0,00%	0,00%

5 Facteurs macro-économiques de fragilité	31,30%	33,60%	25,90%	26,30%
• Diminution de la demande adressée au secteur	19,80%	25,60%	14,70%	15,20%
• Cas de force majeure (guerre du Golfe, catastrophe naturelle...)	10,40%	3,20%	5,70%	5,70%
• Hausse de la concurrence hors les prix	5,20%	8,00%	5,20%	5,30%
• Hausse de la concurrence sur les prix	4,20%	7,20%	5,00%	5,00%
• Politique publique moins favorable au secteur	4,20%	2,40%	1,60%	1,70%
• Période de rationnement des crédits	1,00%	0,00%	0,50%	0,50%
• Niveau général des taux d'intérêt trop élevé	1,00%	0,00%	0,20%	0,20%
• Evolution défavorable du taux de change	2,10%	0,00%	0,00%	0,10%
6 Coûts et structure de production	40,60%	36,80%	21,10%	22,10%
• Coûts d'exploitation trop élevés (hors frais de personnel)	26,00%	18,40%	10,20%	10,90%
• Frais de personnel trop élevés	14,60%	11,20%	5,20%	5,70%
• Capacité de production trop forte, sur investissements	5,20%	8,00%	2,70%	3,00%
• Hausse brutale des coûts d'exploitation	1,00%	3,20%	2,50%	2,50%
• Sous investissements	2,10%	2,40%	1,60%	1,60%
• Perte brutale d'un fournisseur ou refus d'accepter des retards de paiement	0,00%	4,00%	1,10%	1,20%
• Processus de production inadapté ou obsolète	4,20%	3,20%	0,90%	1,10%
• Dépréciation des actifs	0,00%	0,80%	1,10%	1,10%
7 Stratégie de l'entreprise	14,60%	20,00%	15,90%	16,00%
• Echecs de projets importants : investissements, restructuration...	14,60%	18,40%	15,00%	15,10%
• Acceptation consciente de marchés peu rentables (dumping...)	1,00%	1,60%	0,90%	0,90%
Nombre moyen de causes recensées par entreprise	3,20	3,60	2,80	2,90
Nombre d'entreprises ayant une information sur la cause	96	125	441	

La somme de chaque colonne est supérieure à 100, car plusieurs causes ont pu affecter la même entreprise.

L'étude de Blazy et Combiér (1997) est intéressante à plusieurs égards. D'abord, elle étend largement le spectre des causes mis en lumière par celle de la Caisse nationale des marchés de l'État et celle de Conan et Holder, tout en recoupant la quasi-totalité des raisons déjà évoquées (seules la « croissance trop rapide » et la « mauvaise gestion de l'entreprise » ne se retrouvent pas dans cette liste) avec quelques variations dans l'ordre d'importance. Ensuite, et plus fondamentalement, elle permet de faire réellement la part entre les défaillances pour causes accidentelles, donc par nature imprévisibles et celles qui le sont moins, donc pour lesquelles on peut penser qu'elles laisseront une trace dans les comptes. Elle permet aussi de distinguer les facteurs dont la brutalité des effets ne laisse aucun chance de prévision, de ceux qui, *a contrario*, possèdent une telle inertie d'action qu'ils laissent suffisamment de temps pour anticiper leurs conséquences.

La première série de facteurs de défaillance, qui présente le degré de « dangerosité » le plus élevé, est représentée par des événements accidentels et plus particulièrement des événements pouvant affecter le dirigeant ou les relations de l'entreprise avec ses partenaires. Cette série est manifestement la plus problématique, car sa traduction financière est fortement sujette à caution. Il y a vraisemblablement peu de chance que le décès d'un dirigeant, une malversation ou encore des problèmes sociaux puissent transparaître explicitement dans des comptes. Et lorsque ce n'est pas le cas, comme on peut l'envisager lors d'un litige (qui représente près de 60 % des causes de cette catégorie), puisqu'une trace devrait se retrouver dans les provisions, la survenue de l'événement est vraisemblablement trop proche du moment de la défaillance pour que cette information puisse avoir une quelconque utilité.

La deuxième série est relative aux ventes et aux débouchés et fait apparaître trois éléments de crise essentiels : la « perte tendancielle de parts de marché », la « perte brutale de clients » et la « défaillance de clients importants ». Une analyse de l'évolution du chiffre d'affaires, des comptes clients et des provisions pour clients douteux est à même de révéler ces facteurs, mais, comme précédemment, l'utilité de leur diagnostic dépendra de la vitesse à laquelle l'onde de choc de ces événements se répercutera sur l'activité. Or, deux de ces trois causes portent en elles-mêmes un caractère soudain qui laisse présager quelque impossibilité de prévision.

La troisième série a trait aux problèmes financiers qui se révèlent dans trois facteurs importants : un « déficit de ressources propres », un « défaut de paiement de clients » et « un refus de prêt à l'entreprise ». Ici aussi, on peut retrouver, dans les états comptables, les effets induits par ces difficultés, mais peut-être avec une incertitude moins forte quant à la rapidité de traduction de ces effets. Il est vraisemblable que les critères liés au financement présentent une inertie plus élevée que ceux liés aux débouchés, dès lors qu'ils se jouent sur des horizons de temps plus éloignés.

La quatrième série relève de questions liées au système d'information et au management et s'articule autour de la déficience du « système comptable », de « l'incompétence » de l'encadrement et de la « mésentente au sein de l'équipe dirigeante », autant de critères difficilement appréhendables avec précision par des états financiers. Seules des effets induits pourraient être mesurés, mais en l'état, ces facteurs sont indétectables.

La cinquième série laisse la place à des facteurs vraiment indépendants de l'entreprise, donc sur lesquels elle ne peut agir, et qui renvoient à des paramètres macro-économiques comme la

« diminution de la demande adressée au secteur », des « cas de force majeure : guerre... » ou une « hausse de la concurrence ». Leur traduction dans les comptes ne sera ni immédiate, ni circonscrite à certains postes, mais se retrouvera à terme dans l'ensemble des états.

La sixième série renvoie aux coûts et à la structure de production et discerne, comme facteurs de risque, des « coûts d'exploitation » ou des « frais de personnel » trop élevés, une « capacité de production trop forte » ou encore une « hausse brutale des coûts d'exploitation », autant d'éléments qui s'analysent aisément par le biais des soldes intermédiaires de gestion ou au travers des postes de l'actif.

Enfin, la septième et dernière série, met en jeu la stratégie de l'entreprise avec comme motif essentiel, « l'échec de projets importants : investissements, restructuration... », qui peut trouver diverses traductions comptables.

Que retenir de cette classification ? En première approximation, 80 % des facteurs de défaillance sont traçables, donc détectables. Mais ce pourcentage offre une vision optimiste de la réalité, car en fait, ce qui importe, c'est de pouvoir détecter la survenue d'une situation de crise, donc disposer d'un laps de temps suffisant entre celle-ci et le moment où les effets avant coureurs pourront être mesurés. Et l'on voit bien que la réelle difficulté se situe là ; combien de temps dispose-t-on entre les deux ?

Le tableau 7 indiqué ci-après résume cette problématique. Il reprend les dix principales causes appréhendées par Blazy et Combiér (1997) et mentionne, pour chacune, la ou les rubriques comptables à analyser. Il donne une appréciation du délai qui peut séparer le moment où le facteur survient, du moment où l'entreprise se trouve en défaillance.

Les délais indiqués ici ne sont pas des délais objectifs, car ils n'ont pas été mesurés. Ils correspondent à une appréciation subjective du laps de temps moyen qui peut s'écouler entre deux instants, compte tenu de l'inertie que l'on peut associer à chaque processus. Par exemple, si la défaillance est d'abord due au décès du dirigeant, le délai dans ce cas est certainement très court. Et si d'aventure, il ne l'était pas, on pourrait imputer la défaillance à d'autres causes induites par le décès, qui deviendraient alors les causes principales. C'est sur la base de ce type de raisonnement que nous avons défini les différents délais.

On constate bien que les états comptables constituent donc une bonne source d'information mais présentent cependant une double lacune. Tout d'abord, ce ne sont que des moyens

d'approximation de facteurs qui ne sont généralement pas réductibles totalement à des données numériques. Ensuite, ces moyens n'ont pas la capacité de tout restituer et lorsque, d'aventure, ils le peuvent, il leur est parfois impossible de le faire suffisamment à l'avance pour être réellement exploitables.

Tableau 7 : Facteurs de défaillance et modalités de traduction comptable

Critères de défaillance	Rubriques comptables à analyser	Délai de répercussion sur la défaillance
1 Déficit de ressources propres	Capital, trésorerie	Long
2 Perte tendancielle de parts de marché	Chiffre d'affaires, soldes intermédiaires de gestion	Long
3 Accidents divers (départ du dirigeant, divorce,...)		Court
4 Litiges avec des partenaires privés	Provisions	Moyen
5 Défaut de paiement des clients	Créances clients, créances douteuses, provisions	Court - Moyen
6 Diminution de la demande adressée au secteur	Chiffre d'affaires Production	Moyen
7 Echecs de projets importants : investissements, restructuration...	Production vendue, production stockée, frais de R&D	Moyen
8 Système comptable déficient	Tous les comptes	Long
9 Perte brutale de clients	Créances clients, créances douteuses, provisions	Court - Moyen
10 Coûts d'exploitation trop élevés (hors frais de personnels)	BFR, trésorerie	Long

D'autres études similaires à celles présentées précédemment ont aussi été réalisées et une grande partie d'entre elles s'est plutôt focalisée sur les petites et moyennes entreprises car ce sont essentiellement elles qui subissent des défaillances et qu'elles représentent un poids économique non négligeable. Elles donnent donc lieu depuis de nombreuses années à un courant de recherche spécifique qui leur voue une attention particulière. Ainsi Lussier (1995) et Thornhill et Amit (2003) parviennent-ils à recenser ensemble plus de 60 études empiriques qui se sont attachées à étudier les causes de défaillance qui leur sont propres.

Parmi toutes les études quantitatives portant sur les facteurs de défaillance des petites et moyennes entreprises, celle effectuée par Sullivan et al. (1998), pour le compte du Département du commerce des Etats-Unis, est certainement la plus aboutie car elle s'appuie sur une vaste enquête portant sur 3 284 firmes réparties sur l'ensemble du territoire des Etats-Unis et placées sous le coup d'une procédure judiciaire. De cette enquête il ressort les causes résumées dans le tableau 8 présenté ci-après, qui met particulièrement en avant la prégnance des facteurs

environnementaux, des contraintes financières et des problèmes de gestion interne sur le devenir des PME. On y retrouve les conditions économiques d'exercice de l'activité et l'environnement concurrentiel, le mode de financement et la structure capitalistique de la firme ainsi que tout ce qui participe de la gestion et de l'organisation interne, du mode de management aux techniques de gestion des clients. On voit ici aussi que les causes de nature accidentelle occupent une place non négligeable (problèmes affectant le dirigeant, cas de force majeures).

Tableau 8 : Principaux facteurs de défaillance des petites entreprises selon Sullivan (1998)

Facteurs de défaillance	Occurrence
1 Conditions d'exercice de l'activité externe à l'entreprise (concurrence...)	38,50%
2 Financement (coût du capital, endettement...)	28,00%
3 Problèmes internes à l'entreprise (erreur de management, perte de clients...)	27,10%
4 Litige avec l'administration fiscale	20,10%
5 Litige avec un partenaire économique (créancier...)	18,10%
6 Problème personnel affectant le dirigeant (accident, divorce, maladie...)	16,90%
7 Cas de force majeure (fraude, accident, catastrophe naturelle...)	9,60%
8 Autres	6,40%

Le total est supérieur à 100, car plusieurs causes ont pu affecter la même entreprise.

On trouve une hiérarchie similaire mais plus détaillée dans l'étude réalisée par Bradley (2004) qui s'appuie sur tout un ensemble d'enquêtes conduites par le Small Business Center de l'université d'Arkansas. Cette hiérarchie est présentée le tableau 9 ci-après.

Tableau 9 : Principaux facteurs de défaillance des petites entreprises selon Bradley (2004)

1 Sous capitalisation
2 Manque de prévision
3 Coût de la fiscalité et difficulté de suivi de l'évolution des règles fiscales
4 Problème personnel affectant le dirigeant
5 Méconnaissance de la gestion de la part du dirigeant
6 Liquidité trop faible
7 Départ d'une personne au rôle-clé
8 Technologie obsolète
9 Mauvaise implantation
10 Catastrophe naturelle
11 Système d'information comptable défectueux ou inexistant
12 Manque d'information et de conseil

Ces résultats confortent bien l'idée selon laquelle les motifs de défaillance peuvent s'incarner dans des éléments objectivables qui ne sont pas tous traduisibles dans des états comptables et financiers.

Quelles conséquences peut-on tirer de ceci ? Tout d'abord, que le choix d'une définition juridique de la défaillance, même si elle n'épuise toutes les formes qui y sont associées, est le meilleur garant de la séparabilité statistique de deux populations que tout modèle de prévision cherche à discriminer. En ce sens, cette définition présente le caractère le plus opérationnel compte tenu de la tâche envisagée. Ensuite, qu'un modèle peut reposer sur une vision financière de la firme, mais qu'il gagnerait à être complété par des éléments de nature extra-comptable, indépendamment de toute considération liée à leur nature mesurable, afin de garantir une pleine prise en compte de la multidimensionnalité du phénomène.

Maintenant, se posent deux questions. Puisque l'on est en mesure de matérialiser un horizon de la défaillance, et de repérer ses traces dans différents types d'indicateurs, quelles variables faut-il choisir et comment structurer un échantillon pour pouvoir construire un modèle ?

1.3 VARIABLES INCARNANT LES FACTEURS DE DÉFAILLANCE

La multiplicité des causes dont nous avons rendu compte n'a échappé à personne, pas plus que les limites d'une approche strictement comptable ou financière permettant d'en rendre compte. C'est pourquoi on trouve dans la littérature de multiples études ayant testé les capacités prédictives de données diverses et variées pouvant, de prêt ou de loin, jouer un quelconque rôle dans l'explication de la défaillance. Elles peuvent être classées en trois grands ensembles.

Le premier, bien évidemment, décrit l'entreprise en elle-même, sous toutes ses formes, et inclus aussi bien des variables financières issues des bilans ou comptes de résultats, que des variables relatives à sa structure, son organisation, son management, ses dirigeants, sa stratégie, ses produits ou couples produits-marchés...

Le deuxième renvoie à l'environnement de la firme et notamment à son environnement économique au travers, soit d'indicateurs généraux (inflation, taux d'intérêt, taux de croissance...), soit d'indicateurs décrivant un secteur ou un marché.

Le troisième, qui n'est pas sous-tendu explicitement par l'analyse que nous avons proposée précédemment, a trait aux marchés financiers et aux informations qu'ils véhiculent sur leurs

anticipations de risque. Il trouve sa légitimité dans une hypothèse d'efficience qui veut que le prix d'un titre et son rendement incorporent toute donnée comptable et non comptable relatives à l'entreprise dont il est issu. Cet ensemble représente ainsi, une sorte de synthèse objective des deux autres qui, au regard de certains auteurs, pourrait être avantageusement employée pour les compléter voire les remplacer.

Toutes les variables explicatives des modèles de défaillance dérivent de ces trois sources. On les retrouve sous plusieurs formes, résumées dans le tableau 10 ci-après, par ordre décroissant de fréquence d'utilisation.

Tableau 10 : Typologie des variables explicatives utilisées par les modèles de défaillance

Variables	Fréquence d'utilisation dans les 190 études analysées
1 Ratio financier (rapport entre deux variables financières)	93%
2 Variable statistique (calcul opéré sur un ratio ou une variable financière ou non : moyenne, écart-type, variance, logarithme, score ACP...)	28%
3 Variable de variation (évolution dans le temps d'un ratio ou d'une variable financière)	14%
4 Variable non financière (toute caractéristique de l'entreprise ou de son environnement autre que celle décrite par une variable financière)	13%
5 Variable de marché (ratio ou non incluant le cours, le rendement d'un titre...)	6%
6 Variable financière (ligne d'un bilan, d'un compte de résultat ou d'un état financier)	5%

Le total est supérieur à 100, car plusieurs types de variables ont pu être employés simultanément.

Le tableau 11, donne le détail à partir duquel les valeurs du tableau 10 ont été calculées. Il précise, pour chacune des 190 études analysées, le ou les types de variables employés dans les modèles. Seuls ceux retenus pour créer les modèles, et souvent sélectionnés parmi un ensemble plus vaste, y ont été reportés. Ce tableau rend donc compte principalement du pouvoir prédictif de chaque type.

Tableau 11 : Types de variables employés pour construire les modèles de prévision

Études	Types de variables	Études	Types de variables	Études	Types de variables
AGARWAL (1999)	r	POMPE et FEELDERS (1997)	r	BARANOFF et al. (2000)	r – vs
ALAM et al. (2000)	r	RAHIMIAN et al. (1993)	r	BARNIV et RAVEH (1989)	r – vs
ALTMAN (1968)	r	RAHMAN et al. (2004)	r	DAMBOLENA et KHOURY (1980)	r – vs
ALTMAN (1994)	r	RAMSER et FOSTER (1931)	r	DORSEY et al. (1995)	r – vs
ATIYA (2001)	r	SALCEDO-SANZ et al. (2004)	r	ELMER et BOROWSKI (1988)	r – vs
BACK et al. (1996 A)	r	SALCEDO-SANZ et al. (2005)	r	FLAGG et al. (1991)	r – vs

BACK et al. (1996 B)	r	SALCHENBERGER et al. (1992)	r	FRYDMAN et al. (1985)	r – vs
BACK et al. (1997)	r	SARKAR et SRIRAM (2001)	r	HER et CHOE (1999)	r – vs
BAEK et CHO (2003)	r	SEN et al. (2004)	r	HUNTER et ISACHENKOVA (2001)	r – vs
BEAVER (1966)	r	SERRANO-CINCA (1997)	r	JONES et HENSHER (2004)	r – vs
BELL et al. (1990)	r	SEXTON et al. (2003)	r	KANE et al. (1998)	r – vs
BORITZ et KENNEDY (1995)	r	SHARMA et ISELIN (2003)	r	LAITINEN et LAITINEN (1998)	r – vs
BOSE et PAL (2006)	r	SHARMA et MAHAJAN (1980)	r	LANINE et VANDER VENNET (2005)	r – vs
BRABAZON et KEENAN (2004)	r	SHIN et LEE (2003)	r	LI et GUPTA (2002)	r – vs
BRABAZON et O'NEILL (2004)	r	SHIRATA (1998)	r	LIN et PIESSE (2004)	r – vs
CHARALAMBOUS et al. (2000)	r	SINKEY (1975)	r	MIN et LEE (2005)	r – vs
CHARITOU et al. (2004)	r	TAFFLER (1983)	r	PACEY et PHAM (1990)	r – vs
CIELEN et al. (2004)	r	TAM et KIANG (1990)	r	PLATT et PLATT (2002)	r – vs
COATS et FANT (1993)	r	TAM et KIANG (1992)	r	SUN et al. (2005)	r – vs
CYBINSKI (1995)	r	TANG et CHI (2005)	r	WEISS et CAPKUN (2005)	r – vs
DEAKIN (1972)	r	THEODOSSIOU (1991)	r	WU et al. (2007)	r – vs
DIMITRAS et al. (1999)	r	TSAKONAS (2006)	r	LEE et al. (1996)	r – vv
DOUMPOS et ZOPOUNIDIS (1999)	r	TSENG (2005)	r	MIN et LEE (2004)	r – vv
DREZNER et al. (2001)	r	TUNG et al. (2004)	r	PIRAMUTHU et al. (1998)	r – vv
EDMISTER (1972)	r	TYREE et LONG (1996)	r	PLATT et PLATT (1990)	r – vv
FITZPATRICK (1932)	r	VAN GESTEL et al. (2006)	r	SKOGSVIK (1990)	r – vv
GENTRY et al. (1985)	r	VIEIRA et al. (2003)	r	SUNG et al. (1999)	r – vv
GILBERT et al. (1990)	r	WALLRAFEN et al. (1996)	r	YANG et al. (1999)	r – vv
GOMBOLA et al. (1987)	r	WEST et al. (2005)	r	YIM et MITCHELL (2002)	r – vv
GOSS et RAMCHANDANI (1995)	r	WILSON et SHARDA (1994)	r	ANANDARAJAN et al. (2001)	vf – vnf
GRICE et INGRAM (2001)	r	WINAKOR et SMITH (1935)	r	BEYNON et PEEL (2001)	r – vf – vnf
GUPTA et al. (1990)	r	YANG et HARRISON (2002)	r	BARNIV et McDONALD (1992)	r – vf – vs
HÄRDLE et al. (2005)	r	YIM et MITCHELL (2005)	r	BARNIV et al. (1997)	r – vm – vnf
HUA et al. (2007)	r	ZAVGREN (1985)	r	CHAVA et JARROW (2004)	r – vm – vnf
HWANG et al. (1997)	r	ZHANG et al. (1999)	r	FOREMAN (2003)	r – vm – vnf
KARELS et PRAKASH (1987)	r	ZMIJEWSKI (1984)	r	MEYER et PIFER (1970)	r – vs – vv
KEASEY et MCGUINNESS (1990)	r	ZOPOUNIDIS et DOUMPOS (2002)	r	GENTRY et al. (1987)	r – vs – vf
KIM et al. (2003)	r	GENTRY et al. (2002)	vf	BARNIV et al. (2002)	r – vs – vnf
KIM et MCLEOD (1999)	r	KIM et HAN (2003)	vnf	LENSBERG et al. (2006)	r – vs – vnf
KIVILUOTO (1998)	r	LUSSIER (1995)	vnf	PEEL et PEEL (1987)	r – vs – vnf
KUMAR et al. (1997)	r	BOOTH (1983)	vs	RICHARDSON et al. (1998)	r – vs – vnf
KURUPPU et al. (2003)	r	FANNING et COGGER (1994)	vs	BARNIV et HERSHBARGER (1990)	r – vs – vv

LACHER et al. (1995)	r	MENSAH (1984)	vs	BROCKETT et al. (1994)	r – vs – vv
LAITINEN (1991)	r	WANG (2004)	vs	HUANG et al. (1994)	r – vs – vv
LAITINEN et KANKAANPAA (1999)	r	WILCOX (1973)	vs	KOTSIANTIS et al. (2005)	r – vs – vv
LAITINEN et LAITINEN (2000)	r	AZIZ et al. (1988)	vv	LAITINEN et LAITINEN (1995)	r – vs – vv
LEE et al. (2005)	r	BACK et al. (1994)	r – vf	LAU (1987)	r – vs – vv
MCGURR et de VANEY (1998)	r	GODLEWSKI (2003)	r – vf	LEE et URRUTIA (1996)	r – vs – vv
MCKEE et GREENSTEIN (2000)	r	JO et al. (1997)	r – vf	SHIN et al. (2005)	r – vs – vv
MCKEE et LENSBERG (2002)	r	TAN et DIHARDJO (2001)	r – vf	KOLARI et al. (2002)	r – vf – vs – vv
MERWIN (1942)	r	ALTMAN et al. (1995)	r – vs	BERG (2005)	r – vm – vnf – vv
MIN et al. (2006)	r	BEAVER (1968)	r – vm	FAN et PALANISWAMI (2000)	r – vm – vnf – vv
MOYER (1977)	r	THEODOSSIOU (1993)	r – vm	BLUM (1974)	r – vs – vm – vv
NAM et JINN (2000)	r	EL HENNAWY et MORRIS (1983)	r – vnf	SHUMWAY (2001)	r – vs – vm – vv
NANDA et PENDHARKAR (2001)	r	KEASEY et WATSON (1987)	r – vnf	GRICE et DUGAN (2003)	r – vs – vnf – vv
NEOPHYTOU et MAR-MOLINERO (2004)	r	LEE (2004)	r – vnf	LESHNO et SPECTOR (1996)	r – vs – vnf – vv
NORTON et SMITH (1979)	r	LENNOX (1999)	r – vnf	OHLSON (1980)	r – vs – vnf – vv
ODOM et SHARDA (1990)	r	PLATT et al. (1994)	r – vnf	MOSSMAN et al. (1998)	r – vf – vs – vm – vv
PANTALONE et PLATT (1987)	r	TIRAPAT (1999)	r – vnf	WILSON et al. (1995)	r – vs – vnf – vm – vv
PENDHARKAR (2005)	r	WESTGAARD et VAN DER WIJST (2001)	r – vnf	ETHERIDGE et SRIRAM (1997)	?
PHILOSOPHOV et PHILOSOPHOV (2002)	r	ZURADA et al. (1998)	r – vnf	LINDSAY et CAMPBELL (1996)	?
PINCHES et TRIESCHMANN (1974)	r	AGARWAL et al. (2001)	r – vs	VARETTO (1998)	?
POMPE et BILDERBEEK (2005 A)	r	ALTMAN (1973)	r – vs		
POMPE et BILDERBEEK (2005 B)	r	ALTMAN et al. (1977)	r – vs		

r : ratio financier
 vf : variable financière
 vm : variable de marché

vnf : variable non financière
 vs : variable statistique
 vv : variable de variation

Le premier type de variables, et historiquement le plus employé, est bien entendu le ratio, qui exprime le rapport entre deux valeurs comptables représentant toute ligne que l'on peut trouver dans un bilan, un compte de résultat ou un état financier traditionnel. Les ratios tirent leur légitimité de l'absolue nécessité, dès lors que l'on utilise simultanément des variables financières appartenant à différentes entreprises, de contrôler l'effet de taille (Salmi et Martikainen, 1994). En effet, toute donnée financière ne peut s'interpréter que par rapport à une évaluation de

« l'importance » d'une entreprise et c'est encore plus vrai si l'on tente des analyses comparatives. Quel sens donner à l'égalité de deux chiffres d'affaires ou de deux résultats, par exemple, si l'on n'a aucun renseignement sur le nombre d'employés ou la structure d'actifs qui ont permis de les obtenir ? Les ratios participent ainsi d'une stratégie de normalisation des données rendant possible des opérations de catégorisation ou de comparaison impliquées dans tout processus de prévision. Mais, l'hypothèse sous-jacente aux ratios, qui réside dans l'existence d'une relation de proportionnalité entre un numérateur et un dénominateur n'est pas toujours respectée (Lev et Sunder, 1979) ; ils ne permettent donc de contrôler l'effet de taille que de manière imparfaite. Gupta (1969), notamment, a montré que lorsque la taille croît, les ratios de profitabilité (Marge / Ventes) et de liquidité ((Actif Circulant - Stocks) / Dette Court Terme, Actif Circulant / Dette Court Terme) augmentent alors que les ratios d'activité (Rotation des Actifs) et de levier (Ratios sur Dettes ou Actifs) diminuent. Horrigan (1983) faisait remarquer que l'absence de contrôle total de l'effet de taille est plutôt une bonne chose, arguant du fait que celle-ci est aussi une variable explicative de certains comportements. Or, tout modèle de prévision de défaillance étant construit sur la base d'échantillons d'entreprises, devrait ne pas être contraint par la sur- ou sous-représentation de telle ou telle catégorie (petite *versus* grande entreprise). Malgré cet écueil, les ratios demeurent la forme privilégiée de mesure de la santé financière des firmes. On les retrouve en effet, dans plus de 93 % des modèles décrits dans les 190 études que nous avons analysées, les quelques 7 % restant utilisant d'autres formes de variables. Au demeurant, plus de 53 % de ces études incluent uniquement des ratios dans leurs modèles et presque 78 % incluent des ratios utilisés soit seuls, soit en conjonction avec un autre type de variable.

À côté des ratios, on trouve cinq autres formes qui jouent des rôles secondaires, sans être totalement négligeables. Parmi celles-ci, les « variables statistiques » occupent la deuxième place de la liste. Elles désignent toutes les transformations appliquées à des ratios ou à des variables financières à l'aide de fonctions statistiques ou mathématiques : moyenne, écart-type, variance... logarithme. On y retrouve aussi les variables représentant des scores issus d'une analyse factorielle. Souvent, ces opérations participent d'une logique de normalisation des données, la stratégie dominante consistant à appliquer des transformations logarithmiques à des « variables financières ». D'ailleurs, depuis Altman (1968), le logarithme de la variable « Total Actifs » fait partie de ces variables qui présentent un réel pouvoir discriminant.

Les « variables de variation » arrivent à la troisième place. Elles correspondent à des évolutions, généralement annuelles, permettant de tenir compte, non pas d'un ratio ou d'une variable en tant que telle, mais des changements qui ont pu l'affecter. Elles incarnent plutôt une vision dynamique de la défaillance qui fait l'hypothèse qu'il existe un ou des processus qui y conduisent et que ce sont les tendances et les variations qui sont d'abord en mesure de révéler la position d'une entreprise dans le cours de l'un de ces processus. Cette représentation ne s'oppose pas à une vision statique de la défaillance, mais doit être vue comme complémentaire, l'une permettant d'apprécier un déséquilibre, l'autre le sens de son évolution.

Les « variables non financières » se situent à la quatrième place, incluant des caractéristiques quantitatives ou qualitatives, autres que d'origine comptable, propres aux entreprises³ ou à leur environnement⁴. Elles participent de cette volonté d'élargir le spectre des dimensions de la défaillance à celles qui échapperaient aux modèles purement comptables et qui, comme nous l'avons vu précédemment, jouent un rôle non négligeable.

À la cinquième place, on trouve les « variables de marché », reflétant l'évaluation que les marchés financiers font des entreprises au travers du cours et du rendement de leurs titres.

Enfin, à la sixième et dernière place, se situent les « variables financières », et notamment celles souvent employées dans le calcul des ratios, mais parfois utilisées en l'état, et sans transformation, comme notamment des éléments de liquidité dérivant de la notion de *Cash Flow*, d'actifs ou de chiffre d'affaires.

Les tableaux 10 et 11, p. 33, illustrent l'absolue domination des ratios parmi les variables possédant le meilleur pouvoir de prévision, puisqu'on les retrouve dans plus de neuf modèles sur dix. Est-ce à dire que les autres n'ont pas une telle qualité ? La réponse ne peut qu'être nuancée, car le choix des variables initiales semble dépendre d'une multitude de considérations.

Lorsque l'on regarde avec attention les arguments avancés par les uns et les autres pour expliquer comment le choix initial a été opéré, on constate que ce choix découle souvent de choix antérieurs, et donc repose sur ce que les auteurs de référence ont fait. Les tableaux 12 et 13 ci-après sont à cet égard assez symptomatiques de cette pratique. Et comme les choix des tous premiers

³ Caractéristiques des dirigeants (expérience du management, niveau d'études, âge, motivation, niveau social, capacité de leadership...), réactivité du management en cas de problème, existence d'une stratégie à long terme, nombre de partenaires, concentration des clients, dépendance vis-à-vis des fournisseurs, niveau d'exportation, proximité géographique des principaux concurrents, relation avec les banques, niveau de diversification, avis des auditeurs sur la firme... part de marché, âge de la firme, taille, secteur d'intervention...

⁴ Inflation, taux d'intérêt, cycle économique, disponibilité du crédit, profitabilité du secteur, potentiel du marché, intensité concurrentielle...

auteurs (Ramser et Foster, 1931 ; Fitzpatrick, 1932 ; Winakor et Smith, 1935 ; Merwin, 1942 ; Beaver, 1966 ; Altman, 1968 ; Deakin, 1972 ; Edmister, 1972 ; Blum, 1974 ; Altman, 1977 ; Taffler, 1983) étaient avant tout structurés autour des ratios, mécaniquement, la pratique s’est répétée.

Tableau 12 : Modalités de sélection des variables employées par les études faisant intervenir un réseau de neurones dans l’élaboration des modèles

Études faisant intervenir les réseaux de neurones	Modalités de sélection des variables initiales ⁵	Modalités de sélection des variables employées pour l’élaboration des modèles neuronaux ⁶
BACK et al. (1994)	Variables utilisées dans des études précédentes	Algorithme génétique
BACK et al. (1996 A)	Variables utilisées par Ramser et Foster (1931), Fitzpatrick (1932), Winakor et Smith (1935), Merwin (1942), Beaver (1966), Altman (1968), Deakin (1972), Edmister (1972), Blum (1974), Altman (1977) et El Hennawy et Morris (1983)	Algorithme génétique
BACK et al. (1996 B)	Variables utilisées par Ramser et Foster (1931), Fitzpatrick (1932), Winakor et Smith (1935), Merwin (1942), Beaver (1966), Altman (1968), Deakin (1972), Edmister (1972), Blum (1974), Altman (1977), Taffler (1982) et El Hennawy et Morris (1983)	Algorithme génétique
BACK et al. (1997)	Variables utilisées par Ramser et Foster (1931), Fitzpatrick (1932), Winakor et Smith (1935), Merwin (1942), Beaver (1966), Altman (1968), Deakin (1972), Edmister (1972), Blum (1974), Altman (1977) et El Hennawy et Morris (1983)	Algorithme génétique
BRABAZON et KEENAN (2004)	Variables utilisées par Beaver (1968), Altman (1968) et (1977), Dambolena et Khoury (1980), Ohlson (1980), Serrano-Cinca (1996), Kahya (1996), Back et al. (1996), Sung et al. (1999) et Moody's (2000)	Algorithme génétique
KIM et al. (2003)	Variables utilisées habituellement dans la littérature	Algorithme génétique
KIVILUOTO (1998)	?	Algorithme génétique
SEXTON et al. (2003)	Variables utilisées habituellement par l’analyse financière	Algorithme génétique
WALLRAFEN et al. (1996)	Variables appartenant à chacune des catégories suivantes : liquidité, profitabilité, capital, solidité, endettement et capacité de remboursement	Algorithme génétique
BOSE et PAL (2006)	Variables utilisées dans l’étude de Dimitras et al. (1996) et nouveaux ratios	Méthode adaptée au réseau de neurones utilisé, mais non décrite (!)
CHARALAMBOUS et al. (2000)	Variables utilisées dans des études précédentes	Méthode adaptée au réseau de neurones utilisé
TYREE et LONG (1996)	?	Méthode adaptée au réseau de neurones utilisé (mesure de sensibilité) et recherche stepwise sur un critère optimisé pour l’analyse discriminante

⁵ Certains articles correspondant aux études recensées dans cette colonne ne sont pas cités dans la bibliographie. Le lecteur intéressé trouvera les références des dites études dans les articles des auteurs cités dans la première colonne de ce tableau.

⁶ Le tableau 12 indique le mode de sélection des variables pour chaque étude que nous avons recensée et qui utilise un réseau de neurones. Dans la plupart des cas, cette sélection est effectuée en deux temps. Tout d’abord, un premier ensemble, comportant un nombre assez conséquent de variables, est choisi, puis une sélection plus fine est opérée afin de trouver un ou plusieurs sous-ensembles susceptibles ensuite d’être utilisés pour construire les modèles. Le premier ensemble est généralement retenu parce que ses composantes ont déjà été employées dans des études précédentes. Lorsque l’on remonte dans la chaîne de filiation de ces études, on constate bien que les variables en question sont considérées comme étant de bons prédicteurs de la santé financière d’une firme soit parce que la littérature financière les décrits comme tels, soit parce que cette qualité a été démontrée par une étude expérimentale *ad hoc*. Dans tous les cas de figure, les critères d’appréciation ne présentent aucun caractère d’optimalité vis-à-vis d’un réseau de neurones. Quant aux sous-ensembles qui ont servi à établir les modèles, ils sont issus là aussi dans la plupart des cas, de méthodes qui ne tiennent pas compte des particularités des réseaux. En définitive, le choix est effectué directement en puisant dans la littérature ou au travers de procédures associées, dans les progiciels statistiques, à l’analyse discriminante et à la régression logistique. Il est particulièrement intéressant de constater que certains auteurs disent avoir choisi leurs variables avec telle ou telle technique de classification alors qu’ils les ont en fait choisies avec des procédures associées à ces techniques par des concepteurs de progiciels.

SEN et al. (2004)	Variables utilisées dans des études précédentes	Test sur les poids du réseau
BROCKETT et al. (1994)	Variables utilisées dans des études précédentes	Analyse canonique, test de colinéarité et recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique
MIN et LEE (2005)	?	Analyse en composantes principales, test t et recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique
TANG et CHI (2005)	Variables utilisées dans la littérature financière et ratios choisis par des experts	Analyse en composantes principales et recherche stepwise sur un critère optimisé pour l'analyse discriminante
CHARITOU et al. (2004)	Variables utilisées dans des études précédentes	Recherche forward et backward sur un critère optimisé pour la régression logistique et tests sur les coefficients des fonctions
LEE et al. (1996)	Variables utilisées dans des études précédentes	Recherche stepwise sur un critère optimisé pour l'analyse discriminante, Lambda de Wilks et emploi d'une mesure d'entropie
MIN et al. (2006)	Variables utilisées dans des études précédentes	Recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique et test t
KUMAR et al. (1997)	Variables utilisées par Altman (1968), Karels et Prakash (1987) et Coats et Fant (1993)	Recherche stepwise sur un critère optimisé pour l'analyse discriminante
POMPE et BILDERBEEK (2005 B)	Variables utilisées par Pompe et Bilderbeek (2005 A)	Recherche stepwise sur un critère optimisé pour l'analyse discriminante : Lambda de Wilks
YANG et HARRISON (2002)	Variables utilisées habituellement par l'analyse financière	Recherche stepwise sur un critère optimisé pour l'analyse discriminante : Lambda de Wilks
YIM et MITCHELL (2002)	Variables utilisées dans des études précédentes	Recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique
YIM et MITCHELL (2005)	Variables appartenant à chacune des catégories suivantes : liquidité, profitabilité, structure, actif	Recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique
MIN et LEE (2004)	?	Recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique et test de différence de moyenne
SHIN et al. (2005)	?	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et test t
POMPE et BILDERBEEK (2005 A)	Variables appartenant aux catégories suivantes : activité, profitabilité, liquidité, solvabilité	Recherche stepwise sur un critère optimisé pour l'analyse discriminante, test F, tests de multicollinéarité et analyse en composantes principales
WILSON et al. (1995)	Variables utilisées habituellement dans la littérature et employées dans des études précédentes	Recherche stepwise sur un critère optimisé pour la régression logistique
BELL et al. (1990)	Variables utilisées dans des études précédentes	Analyse univariée
BARNIV et al. (1997)	Variables employées par les tribunaux	Analyse univariée, test de corrélation et test de significativité des coefficients des variables
PIRAMUTHU et al. (1998)	?	Arbre de classification et règles symboliques
FAN et PALANISWAMI (2000)	Variables utilisées par Altman (1968), Ohlson (1980) et Lincoln (1982) et employées habituellement dans la littérature	Calcul d'une distance entre les centres des groupes
ATIYA (2001)	?	Capacité prédictive des variables et tests de corrélation
KIM et MCLEOD (1999)	Variables calculées à partir d'une AFC utilisées conjointement avec des ratios traditionnellement utilisés dans la littérature financière.	Expert
POMPE et FEELDERS (1997)	Ratios financiers traditionnels validés par des experts comme étant de bons prédicteurs de défaillance	Expert
LESHNO et SPECTOR (1996)	Variables utilisées dans des études précédentes dont les ratios d'Altman (1968)	Ratios d'Altman (1968), test de corrélation et jugement
SALCHENBERGER et al. (1992)	Variables utilisées par l'analyse financière	Régression multiple
KOTSIANTIS et al. (2005)	Variables utilisées habituellement dans la littérature	Relief

JO et al. (1997)	Variables utilisées par Beaver (1966), Altman (1968), Pogue et Solofski (1969), West (1970), Deakin (1972), Edmister (1972), Pinches (1973), Blum (1974), Elam (1975), Libby (1975), Kaplan et Urwitz (1979), Ohlson (1980), Dambolena et al. (1980), Frydman et al. (1985), Srinivasan et al. (1987), Odom et Sharda (1990), Lee et Oh (1990), Surkan et Singleton (1990), Raghupathi et al. (1991) et Horrigan (1996)	Test t
TUNG et al. (2004)	Variables utilisées dans des études précédentes	Test de corrélation
BERG (2005)	Variables utilisées par Bernhardsen (2001)	Test de significativité des coefficients des variables
AGARWAL (1999)	Variables utilisées simultanément par Altman (1968), Wilson et Sharda (1994) et Agarwal (1996)	Pas de sélection de sous-ensembles
AGARWAL et al. (2001)	Variables utilisées par Ward (1994)	Pas de sélection de sous-ensembles
ALAM et al. (2000)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
ANANDARAJAN et al. (2001)	Variables utilisées dans la littérature financière	Pas de sélection de sous-ensembles
BAEK et CHO (2003)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
BORITZ et KENNEDY (1995)	Variables utilisées par Altman (1968) et Ohlson (1980)	Pas de sélection de sous-ensembles
COATS et FANT (1993)	Variables de l'étude d'Altman (1968) afin de pouvoir effectuer une analyse comparative	Pas de sélection de sous-ensembles
DORSEY et al. (1995)	Variables utilisées par Frydman et al. (1985) et variables choisies à partir de résultats d'études antérieures	Pas de sélection de sous-ensembles
FANNING et COGGER (1994)	Variables utilisées par Wilcox (1973)	Pas de sélection de sous-ensembles
GOSS et RAMCHANDANI (1995)	Variables utilisées par Barniv et Hershberger (1990)	Pas de sélection de sous-ensembles
HUANG et al. (1994)	Variables utilisées par les autorités de régulation des compagnies d'assurance américaines	Pas de sélection de sous-ensembles
KIM et HAN (2003)	Variables qualitatives incarnant un ensemble de facteurs de risque	Pas de sélection de sous-ensembles
LACHER et al. (1995)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
LAITINEN et KANKAANPAA (1999)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
LEE et al. (2005)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
LI et GUPTA (2002)	Variables utilisées par Frydman et al. (1985)	Pas de sélection de sous-ensembles
MCKEE et GREENSTEIN (2000)	Variables utilisées par Hopwood et al. (1989)	Pas de sélection de sous-ensembles
ODOM et SHARDA (1990)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
PENDHARKAR (2005)	Variables utilisées par Altman (1968) et Altman et al. (1977)	Pas de sélection de sous-ensembles
RAHIMIAN et al. (1993)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
SERRANO-CINCA (1997)	Variables utilisées par Pina (1989)	Pas de sélection de sous-ensembles
TAM et KIANG (1990)	Variables appartenant à chacune des catégories suivantes : capital, revenu, liquidité	Pas de sélection de sous-ensembles
TAM et KIANG (1992)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
TAN et DIHARDJO (2001)	Variables utilisées dans les études non publiées de Hall et Byron (1992) et dans celle de Tan (1993)	Pas de sélection de sous-ensembles
VIEIRA et al. (2003)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
WEST et al. (2005)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
WILSON et SHARDA (1994)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
WU et al. (2007)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
YANG et al. (1999)	Variables utilisées par Platt et al. (1994)	Pas de sélection de sous-ensembles
ZHANG et al. (1999)	Variables utilisées Altman (1968) et une variable supplémentaire	Pas de sélection de sous-ensembles

ZURADA et al. (1998)	Variables utilisées par Ward et Foster (1996)	Pas de sélection de sous-ensembles
ALTMAN (1994)	?	?
ETHERIDGE et SRIRAM (1997)	?	?
LEE (2004)	Variables utilisées habituellement par l'analyse financière et présentes dans les études de Beaver (1966), Altman (1968), Deakin (1972), Blum (1974), Altman et al. (1977) et Dimitras et al. (1996) et variables non financières dérivées des modèles permettant de mesurer le capital intellectuel d'une entreprise	?

Tableau 13 : Modalités de sélection des variables employées par les études ne faisant pas intervenir un réseau de neurones dans l'élaboration des modèles

Études n'utilisant pas les réseaux de neurones	Modalités de sélection des variables initiales ⁷	Modalités de sélection des variables employées pour l'élaboration des modèles ⁸
ALTMAN (1973)	Variables provenant de la littérature et d'autres études et variables propres au secteur étudié	Recherche stepwise sur un critère optimisé pour l'analyse discriminante
HÄRDLE et al. (2005)	Variables appartenant aux catégories suivantes : profitabilité, liquidité, levier, activité	Recherche stepwise sur un critère optimisé pour l'analyse discriminante
MEYER et PIFER (1970)	Variables utilisées en partie dans des études précédentes et nouvelles variables	Recherche stepwise sur un critère optimisé pour l'analyse discriminante
FRYDMAN et al. (1985)	Variables utilisées par Altman (1968), Deakin (1972) et Altman et al. (1977)	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et arbre de classification
SHIRATA (1998)	Variables utilisées dans des études précédentes essentiellement	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et arbre de classification
SUNG et al. (1999)	Variables utilisées dans des études précédentes.	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et arbre de classification
EDMISTER (1972)	Variables utilisées dans d'autres études et ayant montré leur capacité prédictive	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et corrélations
EL HENNAWY et MORRIS (1983)	Variables appartenant à l'analyse financière	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et corrélations
BARNIV et RAVEH (1989)	Variables utilisées par Frydman et al. (1985)	Recherche stepwise sur un critère optimisé pour l'analyse discriminante
SINKEY (1975)	Variables utilisées dans des études précédentes portant ou non sur le secteur bancaire	Recherche stepwise sur un critère optimisé pour l'analyse discriminante
ALTMAN et al. (1977)	Variables utilisées dans d'autres études et ayant montré leur capacité prédictive	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et test de corrélation
BOOTH (1983)	?	Recherche stepwise sur un critère optimisé pour l'analyse discriminante : Lambda de Wilks
DAMBOLENA et KHOURY (1980)	Variables appartenant à l'analyse financière	Recherche stepwise sur un critère optimisé pour l'analyse discriminante : Lambda de Wilks
KURUPPU et al. (2003)	Variables utilisées dans des études précédentes et faisant partie de littérature financière	Recherche stepwise sur un critère optimisé pour l'analyse discriminante : Lambda de Wilks
PACEY et PHAM (1990)	Ratios utilisés dans des études précédentes	Recherche stepwise sur un critère optimisé pour l'analyse discriminante : Lambda de Wilks

⁷ Ici aussi, certains articles correspondant aux études recensées dans cette colonne ne sont pas cités dans la bibliographie. Le lecteur intéressé trouvera les références des dites études dans les articles des auteurs cités dans la première colonne de ce tableau.

⁸ Le tableau 13 recense la façon dont les principales études portant sur l'élaboration de modèles de prévision de la défaillance à l'aide de techniques autres que les réseaux, ont effectué la sélection des variables explicatives. Lorsque l'on croise les tableaux 12 et 13, on voit bien la similarité des démarches, l'analyse financière et un socle commun d'études publiées entre les années 60 et les années 80 servant de points de références systématiques. Bien sûr, lorsque l'objectif de l'étude vise à tester l'efficacité d'un modèle existant dans d'autres conditions que celles ayant présidé à son élaboration, l'emploi d'une technique de sélection de variables ne se justifie pas. Mais, la plupart du temps, et notamment avec les études employant un réseau de neurones, l'objectif poursuivi n'est pas de cette nature et s'inscrit plutôt dans une stratégie de recherche du modèle le plus précis possible. Dans ce cas, alors, l'emploi de techniques de sélection adaptées aux réseaux peut parfaitement se légitimer.

SHARMA et MAHAJAN (1980)	Variables utilisées dans des études précédentes et notamment par Beaver (1966), Altman (1968) et Blum (1974)	Recherche stepwise sur un critère optimisé pour l'analyse discriminante : Lambda de Wilks
TAFFLER (1983)	Variables présentes dans la littérature financière	Recherche stepwise sur un critère optimisé pour l'analyse discriminante et capacité prédictive des variables
PEEL et PEEL (1987)	Variables appartenant à l'analyse financière	Recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique
CYBINSKI (1995)	Variables utilisées par McNamara et al. (1988)	Recherche stepwise sur un critère optimisé pour la régression logistique
GILBERT et al. (1990)	Variables utilisées par Altman (1968) et Casey et Bartczak (1985)	Recherche stepwise sur un critère optimisé pour la régression logistique
KEASEY et MCGUINNESS (1990)	Variables appartenant à l'analyse financière et utilisées par Zavgren (1985)	Recherche stepwise sur un critère optimisé pour la régression logistique
KOLARI et al. (2002)	Variables appartenant aux catégories suivantes : taille, profitabilité, capitalisation, risque de crédit, liquidité, passif, diversification	Recherche stepwise sur un critère optimisé pour la régression logistique
LAITINEN et LAITINEN (1995)	?	Recherche stepwise sur un critère optimisé pour la régression logistique
LAITINEN et LAITINEN (1998)	Variables financières traditionnelles et variables de liquidité	Recherche stepwise sur un critère optimisé pour la régression logistique
LIN et PIESSE (2004)	Variables généralement considérées dans la littérature comme expliquant la défaillance	Recherche stepwise sur un critère optimisé pour la régression logistique
RAHMAN et al. (2004)	Variables appartenant à l'analyse financière	Recherche stepwise sur un critère optimisé pour la régression logistique
SUN et al. (2005)	Variables utilisées par Altman (1968), Ohlson (1980) et Hopwood (1994) et variables non financières susceptibles d'expliquer la défaillance	Recherche stepwise sur un critère optimisé pour la régression logistique
TSENG (2005)	Variables utilisées par Lin et Piesse (2004)	Recherche stepwise sur un critère optimisé pour la régression logistique
KEASEY et WATSON (1987)	?	Recherche stepwise sur un critère optimisé pour la régression logistique et test de corrélation
VAN GESTEL et al. (2006)	?	Recherche backward sur un critère optimisé pour l'analyse discriminante et la régression logistique
BARANOFF et al. (2000)	Variables utilisées dans des études précédentes	Recherche stepwise par simulation d'un réseau de neurones (?) et jugement
DREZNER et al. (2001)	Variables utilisées par Altman (1968) et Brigham et al. (1999)	Recherche Tabu
SALCEDO-SANZ et al. (2004)	Variables utilisées dans des études précédentes sur les compagnies d'assurance	Recuit simulé et ensemble grossier
BRABAZON et O'NEILL (2004)	Variables financières présentes dans les états financiers des entreprises.	Technique d'évolution grammaticale
GODLEWSKI (2003)	Variables de l'analyse financière	Test de corrélation avec les classes et recherche stepwise sur des critères optimisés pour l'analyse discriminante et la régression logistique
BARNIV et al. (2002)	Variables utilisées par Altman (1968), Ohlson (1980) et Frydman et al. (1985) et autres variables utilisées par les avocats et les juges	Test de corrélations, différence de moyenne et significativité des coefficients des variables
AZIZ et al. (1988)	Variables utilisées par l'étude de Lawson (1971)	Test de différence de moyenne
DOUMPOS et ZOPOUNIDIS (1999)	Variables appartenant à la littérature financière et validées par des experts	Test de différence de moyenne et corrélations
ALTMAN et al. (1995)	Variables appartenant à l'analyse financière et ratios d'Altman (1968)	Test de différence de moyenne, signe des coefficients de la fonction discriminante et capacité prédictive des variables
KARELS et PRAKASH (1987)	Variables utilisées dans des études précédentes	Test de multi-normalité et recherche stepwise sur un critère optimisé pour l'analyse discriminante : distance de Mahalanobis
BARNIV et HERSHBARGER (1990)	Variables pertinentes pour l'analyse financière des compagnies d'assurance	Test de significativité des coefficients des variables
HWANG et al. (1997)	Variables d'analyse financière propres aux banques	Test de significativité des coefficients des variables

JONES et HENSHER (2004)	Variables utilisées par Beaver (1966), Altman (1968), Altman et al. (1977), Ohlson (1980), Zmijewski (1984), Casey et Bartczak (1985), Gentry et al. (1985) et Jones (1987)	Test de significativité des coefficients des variables
PLATT et PLATT (1990)	Variables appartenant à l'analyse financière	Test de significativité des coefficients et capacité prédictive des variables
ALTMAN (1968)	Variables utilisées dans des études précédentes et nouveaux ratios	Test de significativité des fonctions discriminantes obtenues, corrélations, capacité prédictive des variables et jugement
PLATT et PLATT (2002)	Variables utilisées par Platt (1991) et Platt (2000)	Test de significativité et signe des coefficients, test de corrélation et capacité prédictive des variables
PLATT et al. (1994)	Variables utilisées dans des études précédentes et nouveaux ratios	Test de significativité et signe des coefficients, tests de corrélations et capacité prédictive des variables
NAM (2000)	Variables appartenant à l'analyse financière	Test différence de moyenne, recherche stepwise sur un critère optimisé pour la régression logistique et test sur les coefficients des variables
HER et CHOE (1999)	Variables utilisées dans des études précédentes	Test F, test de corrélation et capacité prédictive des variables
HUA et al. (2007)	?	Test F, test de corrélation et recherche stepwise sur un critère optimisé pour l'analyse discriminante
NORTON et SMITH (1979)	Variables utilisées dans des études précédentes	Test F et régression
SALCEDO-SANZ et al. (2005)	Variables utilisées dans des études précédentes sur les compagnies d'assurance	Algorithme génétique
GOMBOLA et al. (1987)	Variables utilisées dans des études précédentes	Analyse en composantes principales
SHARMA et ISELIN (2003)	Variables utilisées dans des études précédentes	Analyse en composantes principales
ZAVGREN (1985)	Variables utilisées dans une étude précédente	Analyse en composantes principales
WANG (2004)	Variables utilisées par Frydman et al. (1985)	Analyse en composantes principales
SKOGSVIK (1990)	?	Analyse en composantes principales, test de différence de moyenne et capacité prédictive des variables
SHIN et LEE (2003)	?	Analyse en composantes principales, test de différence de moyenne et recherche stepwise sur un critère optimisé pour l'analyse discriminante
MENSAH (1984)	Variables utilisées dans des études précédentes	Analyse factorielle des correspondances et test de corrélation
BEAVER (1966)	Variables provenant de la littérature et utilisées dans d'autres études et disposant d'un lien avec la notion de Cash Flow	Cadre théorique et capacité prédictive des variables
LAITINEN et LAITINEN (2000)	Variables sélectionnées selon le cadre conceptuel dit « modèle du Cash Flow » initié par Beaver et selon celui de Scott fondé sur la valeur des actifs détenus par les actionnaires, pour une partie des modèles, et variables utilisées dans des études précédentes, pour l'autre partie : Ramser et Foster (1931), Fitzpatrick (1932), Merwin (1942), Beaver (1966), Altman (1968) et (1977), Deakin (1972), Edmister (1972), Blum (1974) et El Hennawy et Morris (1983)	Cadre théorique et recherche stepwise sur un critère optimisé pour la régression logistique
PANTALONE et PLATT (1987)	Variables utilisées dans des études précédentes ainsi que par les autorités de régulation	Capacité prédictive des variables et signe des coefficients
PINCHES et TRIESCHMANN (1974)	Variables utilisées dans des études précédentes	Capacité prédictive des variables et test de corrélation
GENTRY et al. (1987)	Variables utilisées dans quinze études précédentes et variables issues du modèle de Gentry et al. (1985)	Ratios les plus utilisés dans les études précédentes examinées
WESTGAARD et VAN DER WIJST (2001)	Variables présentes dans la base de données utilisée	Considérations théoriques
THEODOSSIOU (1991)	Variables utilisées dans des études précédentes	C_p de Mallow
PHILOSOPHOV et PHILOSOPHOV (2002)	Variables utilisées dans des études précédentes	Distribution des ratios
MCKEE et LENSBERG (2002)	Variables utilisées par McKee (1999)	Ensemble grossier
DIMITRAS et al. (1999)	Variables utilisées par Dimitras (1995)	Expert

TSAKONAS (2006)	?	Expert
LENSBERG et al. (2006)	Variables utilisées dans des études précédentes	Expert et capacité prédictive des variables
BARNIV et McDONALD (1992)	Variables utilisées dans des études précédentes	?
THEODOSSIOU (1993)	?	?
BEAVER (1968)	Variables utilisées dans l'étude de Beaver (1966)	Pas de sélection de sous-ensembles
BEYNON et PEEL (2001)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
BLUM (1974)	Variables sélectionnées selon le cadre conceptuel dit « modèle du Cash Flow » initié par Beaver	Pas de sélection de sous-ensembles
CHAVA et JARROW (2004)	Variables utilisées par Altman (1968), Zmijewski (1984) et Shumway (2001)	Pas de sélection de sous-ensembles
CIELEN et al. (2004)	Variables utilisées par Foster (1986) et Ooghe et Van Vijmeersch (1994)	Pas de sélection de sous-ensembles
DEAKIN (1972)	Meilleurs Variables parmi les trente testées par Beaver (1966)	Pas de sélection de sous-ensembles
ELMER et BOROWSKI (1988)	Variables utilisées par Altman et al. (1977) et Barth et al. (1985)	Pas de sélection de sous-ensembles
FITZPATRICK (1932)	Variables appartenant à l'analyse financière	Pas de sélection de sous-ensembles
FLAGG et al. (1991)	Variables utilisées par Gentry et al. (1985)	Pas de sélection de sous-ensembles
FOREMAN (2003)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
GENTRY et al. (1985)	Variables dérivées du modèle de Helfert (1982)	Pas de sélection de sous-ensembles
GENTRY et al. (2002)	Variables utilisées par Gentry et al. (1987)	Pas de sélection de sous-ensembles
GRICE et INGRAM (2001)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
GRICE et DUGAN (2003)	Variables utilisées par Ohlson (1980) et Zmijewski (1984)	Pas de sélection de sous-ensembles
GUPTA et al. (1990)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
KANE et al. (1998)	Variables utilisées par Hopwood et al. (1994)	Pas de sélection de sous-ensembles
LANINE et VANDER VENNET (2005)	?	Pas de sélection de sous-ensembles
LAU (1987)	Variables sélectionnées selon un cadre conceptuel dit de « flexibilité financière »	Pas de sélection de sous-ensembles
LEE et URRUTIA (1996)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
LENNOX (1999)	?	Pas de sélection de sous-ensembles
LINDSAY et CAMPBELL (1996)	?	Pas de sélection de sous-ensembles
LUSSIER (1995)	Variables non financières utilisés dans des études précédentes	Pas de sélection de sous-ensembles
MCGURR et de VANEY (1998)	Ratios utilisés dans les études de Deakin (1977), Ohlson (1980), Gentry et al. (1985), Zavgren (1985) et Gombola et al. (1987)	Pas de sélection de sous-ensembles
MERWIN (1942)	Ratios de l'analyse financière	Pas de sélection de sous-ensembles
MOSSMAN et al. (1998)	Variables utilisées par Altman (1968), Aharony et al. (1980), Clark et Weinstein (1983) et Aziz et al. (1988)	Pas de sélection de sous-ensembles
MOYER (1977)	Variables utilisées par Altman (1968)	Pas de sélection de sous-ensembles
NANDA et PENDHARKAR (2001)	Variables utilisées par Altman (1968) et Altman et al. (1977)	Pas de sélection de sous-ensembles
NEOPHYTOU et MAR-MOLINERO (2004)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
OHLSON (1980)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
RAMSER et FOSTER (1931)	Variables appartenant à l'analyse financière	Pas de sélection de sous-ensembles
RICHARDSON et al. (1998)	Variables utilisées par Hopwood et al. (1994)	Pas de sélection de sous-ensembles
SARKAR et SRIRAM (2001)	Variables utilisées par Sarkar et al. (1996)	Pas de sélection de sous-ensembles

TIRAPAT et NITTAYAGASETAWAT (1999)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
VARETTO (1998)	?	Pas de sélection de sous-ensembles
WEISS et CAPKUN (2005)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
WILCOX (1973)	Variables dérivées d'un modèle théorique de la défaillance	Pas de sélection de sous-ensembles
WINAKOR et SMITH (1935)	Variables appartenant à l'analyse financière	Pas de sélection de sous-ensembles
ZMIJEWSKI (1984)	Variables utilisées dans des études précédentes	Pas de sélection de sous-ensembles
ZOPOUNIDIS et DOUMPOS (2002)	?	Pas de sélection de sous-ensembles
HUNTER et ISACHENKOVA (2001)	Variables utilisées dans la littérature britannique	Pas de sélection de sous-ensembles
LAITINEN (1991)	Variables déduites d'un cadre théorique de la défaillance	Pas de sélection de sous-ensembles
SHUMWAY (2001)	Variables utilisées par Altman (1968) et de Zmijewski (1984) et variables de marché	Pas de sélection de sous-ensembles

Ce n'est donc pas à l'origine, semble-t-il, la question de la performance de tel ou tel type de variables qui a imposé les ratios, mais plutôt un phénomène de mimétisme renforcé, plus tard, par d'autres paramètres. Parmi ceux-ci, l'analyse financière et les études théoriques sur les ratios ont fondé une légitimité mathématique nécessaire qui venait compléter astucieusement celle qu'ils avaient déjà acquise d'un point de vue empirique. À ceci s'est ajouté leur caractère économique⁹, car les données dont ils sont issus sont simples à obtenir et à vérifier, alors que les données de nature extra-comptable ne concernent parfois que certaines catégories d'entreprises, comme les données boursières qui n'existent que pour les sociétés cotées sur une place financière, ou tout simplement sont impossibles à trouver ou très difficilement. C'est également le cas des données qualitatives qui décrivent le management, l'organisation et l'environnement des entreprises et dont l'influence sur leur pérennité a souvent été soulignée. D'ailleurs, l'étude de Sullivan (1998) qui a porté sur 3 284 firmes, n'a pu s'appuyer que sur un quart de cet échantillon lorsqu'il s'est agi de recueillir de telles informations, les représentants des trois quarts des entreprises restantes étant injoignables, car introuvables.

Trop peu d'études comparatives sur le pouvoir de prévision des unes et des autres ont été menées pour que l'on puisse imputer une quelconque influence de ce critère sur le taux respectif d'utilisation des différents types de variables. Qu'on en juge. Back et al. (1994) ont montré qu'un modèle constitué uniquement de ratios pouvait donner de meilleurs résultats

⁹ La Banque de France a réalisé en 2003 un « score sur variables qualitatives » venant compléter sa batterie traditionnelle de scores quantitatifs (Lelogeais, 2003). Mais il s'agit d'une étude exceptionnelle, car, comme le fait remarquer Bardos (2005), p. 66, une approche « qualitative est propice à l'approfondissement de l'analyse ou à la détection d'autres facteurs de risques. » Mais, « pour intervenir dans la construction d'un outil statistique, elle nécessiterait de recourir à des enquêtes coûteuses et difficiles à mener à grande échelle. » Et de conclure : « C'est pourquoi, dans la construction des scores, la Banque de France a privilégié la seule approche comptable, quantifiable et largement applicable, malgré les limites qu'elle présente. »

qu'un modèle composé uniquement de variables financières traditionnelles (actifs, dettes, résultats...). Mossman et al. (1998) ont étudié les différences obtenues entre des modèles bâtis avec des ratios et d'autres élaborés avec des variables de marché, et ont montré une légère supériorité des premiers. Keasey et Watson (1987) ont comparé les résultats issus de trois modèles : un constitué uniquement de ratios, un autre composé seulement de variables non financières, et un dernier élaboré à partir de l'ensemble des variables précédentes. Dans cette hypothèse, le dernier modèle a présenté de bien meilleures prévisions que les deux autres, et celui composé de ratios s'est montré légèrement plus précis que celui s'appuyant sur des variables non financières. Lussier (1995) s'est attaché à construire un modèle à partir de variables qualitatives caractérisant aussi bien les dirigeants que l'activité des entreprises de son échantillon, sans que le modèle produit ne donne vraiment satisfaction, le taux de bons classements des entreprises défaillantes avoisinant les 73 % et celui des entreprises saines, les 65 %. Atiya (2001) a opposé un modèle construit avec des ratios, à un modèle mixant des ratios avec des variables de marché, montrant une légère supériorité du second sur le premier. Tirapat et Nittayagasetwat (1999) ont étudié l'apport de variables macro-économiques à un modèle constitué de ratios, permettant de tenir compte de la sensibilité des entreprises aux variations de facteurs environnementaux, mais n'ont réussi à obtenir un taux de bonnes prévisions qu'à peine supérieur à 70 %. D'autres études encore se sont penchées sur la plus ou moins grande propension de ces familles de variables à prédire correctement la défaillance, sans qu'aucun résultat tranché n'apparaisse. Quant aux « variables de variation », Pérez (2002), au travers d'une étude conduite sur de nombreux échantillons de PME, avec plusieurs séries de données décrivant des variations de comptes sur plusieurs années, a obtenu des résultats nettement moins bons qu'avec des valeurs absolues. Ses conclusions vont dans le même sens que celles obtenues par Pompe et Bilderbeek (2005 A), qui considèrent que les valeurs absolues des ratios sont de meilleurs prédicteurs que leurs variations dans le temps.

On dispose donc de très peu de points de repère pour expliquer avec certitude la prédominance des ratios, même si on en discerne les contours. Les ratios étant capables de donner d'excellentes prévisions, le gain marginal d'autres types de variables, compte tenu de leur coût d'acquisition, les met de fait hors jeu. C'est aussi pourquoi la plus grande partie des efforts de recherche s'est portée, et se porte toujours, sur d'autres questions ; notamment celle des méthodes et de leur efficacité relative. L'examen des deux tableaux suivants présentant les problématiques des principaux travaux de recherche sur les modèles de défaillance,

employant (tableau 14) ou non (tableau 15) des réseaux de neurones, confirme bien ceci ; la question des variables, dès lors qu'il ne s'agit plus de ratios, est réellement périphérique.

Tableau 14 : Problématiques des études faisant intervenir un réseau de neurones dans l'élaboration des modèles

Études	Problématiques
AGARWAL (1999)	Montrer l'efficacité d'une nouvelle méthode (réseau abductif) pour construire un modèle de prévision de la défaillance en la comparant à celle d'un réseau de neurones construit pour faire de même
AGARWAL et al. (2001)	Construire un modèle de prévision capable de discerner plusieurs niveaux de défaillance à l'aide de plusieurs méthodes
ALAM et al. (2000)	Comparer la capacité d'une méthode de clustering floue à regrouper correctement des banques défaillantes réparties dans un échantillon de banques financièrement saines avec celle d'une carte de Kohonen confrontée au même problème
ALTMAN et al. (1994)	Comparer la performance de modèles de prévision construits à l'aide d'une analyse discriminante, d'une régression logistique et d'un réseau de neurones
ANANDARAJAN et al. (2001)	Comparer l'efficacité d'un réseau de neurones entraîné avec un algorithme génétique avec celle d'un réseau entraîné avec une méthode de descente de gradient pour construire un modèle susceptible de prévoir la faillite d'entreprises en mauvaise santé financière
ATIYA (2001)	Construire un modèle de prévision incluant de nouveaux indicateurs dérivés du prix des actions des entreprises considérées
BACK et al. (1994)	Etudier la capacité de prédiction des réseaux de neurones ; comparer les capacités prédictives des ratios par rapport à des variables financières ; étudier la performance des modèles créés avec des données calculées 1, 2 et 3 ans avant la défaillance ; analyser l'intérêt d'un algorithme génétique pour la sélection des variables
BACK et al. (1996 A)	Analyser l'influence des méthodes d'analyse sur la sélection des variables : les différences entre les méthodes conduisent-elles à des choix de variables différents et à des différences significatives dans la précision des résultats ?
BACK et al. (1996 B)	Tester l'influence de la méthode employée pour sélectionner les variables sur la performance d'un réseau de neurones
BACK et al. (1997)	Étudier l'effet associé à la taille de l'échantillon sur le pouvoir de prévision de différentes méthodes
BAEK et CHO (2003)	Proposer une méthode d'apprentissage d'un réseau ne tenant compte que des entreprises saines afin de pallier les problèmes d'échantillonnage rencontrés dans la plupart des études sur la défaillance dus à l'asymétrie des populations étudiées
BARNIV et al. (1997)	Construire un modèle capable de prévoir l'issue d'une procédure judiciaire de faillite
BELL et al. (1990)	Examiner l'utilité des données financières annuelles ainsi que celle des techniques statistiques alternatives pour modéliser la décision de fermeture d'établissements bancaires pouvant émaner des autorités de régulation
BERG (2005)	Tester l'efficacité de modèles additifs généralisés par rapport à l'analyse discriminante, aux modèles linéaires et aux réseaux de neurones
BORITZ et KENNEDY (1995)	Comparer l'efficacité de différents réseaux de neurones pour prévoir la défaillance
BOSE et PAL (2006)	Trouver les facteurs explicatifs de la défaillance des « dotcom » au début des années 2000 ainsi que les ratios financiers qui permettent de les mettre en exergue
BRABAZON et KEENAN (2004)	Tester l'efficacité d'un modèle construit à l'aide d'un réseau de neurones paramétré grâce à un algorithme génétique pour prévoir la défaillance
BROCKETT et al. (1994)	Construire un système de prévision de l'insolvabilité adapté aux compagnies d'assurance à destination des autorités de régulation du secteur
CHARALAMBOUS et al. (2000)	Appliquer un algorithme de sélection de variables pour construire le modèle de prévision de la défaillance le plus pertinent possible
CHARITOU et al. (2004)	Développer un modèle fiable de prévision de la défaillance adapté aux entreprises industrielles britanniques incorporant des indicateurs de Cash Flow
COATS et FANT (1993)	Montrer qu'un réseau de neurones peut être utilisé pour détecter prématurément les signes de défaillance et qu'il peut donner de meilleurs résultats que l'analyse discriminante
DORSEY et al. (1995)	Tester les capacités de prévision d'un réseau de neurones sur de grandes entreprises
ETHERIDGE et SRIRAM (1997)	Tester la capacité des réseaux de neurones à construire des modèles de prévision plus précis que ceux pouvant être élaborés avec des méthodes traditionnelles

FAN et PALANISWAMI (2000)	Tester l'efficacité d'un séparateur à vaste marge couplé à une méthode de sélection de variables de type « filter » pour construire un modèle de prévision de la défaillance
FANNING et COGGER (1994)	Examiner les potentialités d'un nouveau réseau de neurones, notamment par rapport à un réseau à rétropropagation traditionnel, pour prévoir la défaillance d'entreprises
GOSS et RAMCHANDANI (1995)	Dépasser les limites des modèles paramétriques grâce aux réseaux de neurones ; tester l'efficacité des réseaux dans la prévision de défaillance des compagnies d'assurance
HUANG et al. (1994)	Étudier le pouvoir de prévision de certaines variables employées pour le contrôle des compagnies d'assurance en faisant varier la forme de la relation qui les unit
JO et al. (1997)	Comparer la précision de trois méthodes différentes pour prévoir la défaillance d'entreprise
KIM et al. (2003)	Comparer le pouvoir prédictif de données comptables avec celui de données de marché grâce à un réseau de neurones
KIM et HAN (2003)	Comparer l'efficacité d'un modèle construit à l'aide d'un algorithme génétique utilisé pour représenter les règles de décision d'un expert avec celle de modèles construits avec un réseau de neurones et une méthode d'induction de règles
KIM et MCLEOD (1999)	Comparer le pouvoir prédictif de modèles linéaires et non linéaires avec les capacités prédictives d'experts
KIVILUOTO (1998)	Comparer l'efficacité d'un modèle de prévision construit à l'aide d'une carte de Kohonen avec celle de modèles issus de méthodes statistiques traditionnelles
KOTSIANTIS et al. (2005)	Comparer l'efficacité de différents modèles construits à partir de méthodes du type « Machine Learning » pour prévoir la défaillance d'entreprises
KUMAR et al. (1997)	Comparer le pouvoir prédictif de modèles issus de différentes méthodes et montrer la capacité des algorithmes génétiques à produire des modèles performants
LACHER et al. (1995)	Étudier l'efficacité d'un réseau de neurones pour prévoir le degré de santé financière d'une entreprise
LAITINEN et KANKAANPAA (1999)	Comparer les résultats issus de plusieurs méthodes de création de modèles afin d'étudier s'ils diffèrent significativement les uns des autres
LEE et al. (1996)	Tester l'efficacité de l'emploi combiné de plusieurs méthodes pour construire des modèles de prévision
LEE (2004)	Tester l'efficacité de variables mesurant le capital intellectuel d'une entreprise afin d'améliorer le pouvoir prédictif d'un modèle de prévision de la défaillance incorporant des ratios financiers traditionnels
LEE et al. (2005)	Comparer l'efficacité d'un réseau de neurones utilisant une technique d'apprentissage supervisé avec celle d'un réseau employant une technique d'apprentissage non supervisé pour prévoir la défaillance
LESHNO et SPECTOR (1996)	Tester l'efficacité de plusieurs modèles créés avec un réseau de neurones pour étudier l'influence de certains paramètres d'apprentissage
LI et GUPTA (2002)	Réexaminer le rôle des réseaux de neurones dans le domaine de la construction de modèles de prévision et étudier l'apport des modèles Logit et Probit avec hétéroscédasticité dans ce domaine
MCKEE et GREENSTEIN (2000)	Tester l'efficacité d'un modèle construit avec un arbre de classification par rapport à celle de modèles élaborés avec un réseau de neurones et une régression logistique
MIN et LEE (2004)	Tester l'efficacité d'un modèle de prévision de la défaillance construit à l'aide d'un séparateur à vaste marge par rapport à un modèle construit avec un réseau de neurones
MIN et LEE (2005)	Tester l'application d'un séparateur à vaste marge pour construire un modèle de prévision de la défaillance
MIN et al. (2006)	Tester l'efficacité d'un séparateur à vaste marge associé à un algorithme génétique servant à sélectionner les variables et les paramètres du séparateur pour construire un modèle de prévision de la défaillance
ODOM et SHARDA (1990)	Tester l'efficacité d'un réseau de neurones pour construire un modèle de prévision de la défaillance
PENDHARKAR (2005)	Tester l'efficacité d'un réseau de neurones disposant d'un seuil variable de séparation des classes pour construire un modèle de prévision
PIRAMUTHU et al. (1998)	Améliorer la capacité d'apprentissage et de prévision d'un réseau de neurones grâce à une méthode permettant de transformer les données initiales en des représentations plus faciles à apprendre par un réseau
POMPE et BILDERBEEK (2005 A)	Étudier l'hypothèse selon laquelle, au cours du processus qui conduit une entreprise à la défaillance, les ratios d'activité et de profitabilité sont les premiers ratios à se dégrader, suivis par les ratios de solvabilité et enfin par les ratios de liquidité, et analyser l'influence de l'âge de l'entreprise sur sa probabilité de défaillance
POMPE et BILDERBEEK (2005 B)	Étudier l'influence de la période de mesure des données sur la précision d'un modèle de prévision de la défaillance ainsi que l'influence de la conjoncture économique sur la performance d'un modèle
POMPE et FEELDERS (1997)	Tester l'efficacité d'un réseau par rapport aux méthodes traditionnelles sur un échantillon de taille conséquente
RAHIMIAN et al. (1993)	Répliquer l'étude d'Odom et Sharda (1990) avec différents types de réseau de neurones
SALCHENBERGER et al. (1992)	Construire un modèle de prévision avec un réseau de neurones

SEN et al. (2004)	Tester une méthode simple de sélection de variables pour améliorer la capacité de prévision d'un modèle de défaillance construit à l'aide d'un réseau de neurones
SERRANO-CINCA (1997)	Montrer les forces et les faiblesses des réseaux de neurones en matière de prévision de la défaillance
SEXTON et al. (2003)	Construire un modèle de prévision de la défaillance avec un réseau de neurones en utilisant un algorithme génétique permettant d'éliminer du réseau les poids inutiles
SHIN et al. (2005)	Étudier l'efficacité d'un séparateur à vaste marge pour construire un modèle de prévision de la défaillance
TAM et KIANG (1990)	Présenter une façon de construire un modèle de prévision de la défaillance bancaire à l'aide d'un réseau de neurones
TAM et KIANG (1992)	Construire un modèle de prévision de la défaillance bancaire tenant compte du coût d'une mauvaise classification
TAN et DIHARDJO (2001)	Étudier la viabilité d'un réseau de neurones pour développer un système de prévision de la défaillance d'établissements de crédits
TANG et CHI (2005)	Proposer une méthode originale de sélection de variables explicatives employant conjointement des experts et des techniques statistiques et tester son efficacité pour construire un modèle de prévision à l'aide d'un réseau de neurones
TUNG et al. (2004)	Proposer l'utilisation d'un nouveau réseau de neurones capable de rendre ses règles de décision internes intelligibles afin de construire un modèle de prévision de la défaillance bancaire
TYREE et LONG (1996)	Tester l'efficacité d'un réseau de neurones probabiliste pour construire un modèle de prévision de la défaillance
VIEIRA et al. (2003)	Tester l'efficacité d'un algorithme d'apprentissage récent particulièrement bien adapté aux problèmes multidimensionnels
WALLRAFEN et al. (1996)	Étudier l'efficacité d'un réseau de neurones paramétré à l'aide d'un algorithme génétique pour élaborer un modèle de prévision de la défaillance sachant plus particulièrement classer correctement les entreprises défaillantes
WEST et al. (2005)	Étudier le potentiel d'une technique consistant à utiliser un ensemble de réseaux de neurones entraînés à l'aide de différentes techniques (validation croisée, bagging, boosting) pour construire un système de prévision de la défaillance
WILSON et al. (1995)	Tenter de prévoir trois états (défaillance, non défaillance et acquisition) après une situation de détresse financière
WILSON et SHARDA (1994)	Comparer l'efficacité d'un modèle de prévision de la défaillance élaboré à l'aide de l'analyse discriminante avec celle d'un modèle construit avec un réseau de neurones et étudier l'effet de la composition des échantillons, et plus précisément du pourcentage de répartition des observations entre les échantillons d'apprentissage et de test, sur la précision des résultats
WU et al. (2007)	Tester l'efficacité d'un séparateur à vaste marge paramétré à l'aide d'un algorithme génétique pour prévoir la défaillance
YANG et al. (1999)	Comparer l'analyse discriminante avec plusieurs techniques de construction de réseaux pour déterminer la meilleure méthode d'élaboration d'un modèle de prévision de faillite dans le secteur pétrolier
YANG et HARRISON (2002)	Tester l'utilisation d'un réseau de neurones probabiliste permettant de construire un modèle de prévision de la défaillance et d'interpréter les symptômes de cette défaillance
YIM et MITCHELL (2002)	Étudier les capacités de prévision des réseaux de neurones par rapport aux méthodes traditionnelles pour prévoir la défaillance d'entreprises australiennes
YIM et MITCHELL (2005)	Tester l'efficacité d'un réseau de neurones hybride pour construire un modèle de prévision
ZHANG et al. (1999)	Tester la sensibilité d'un réseau de neurones aux fluctuations d'échantillonnage
ZURADA et al. (1998)	Mesurer l'incidence du nombre de niveaux de défaillance (absence de paiement de dividendes, défaut de remboursement de prêt, faillite ou liquidation) à prévoir sur la capacité prédictive des modèles construits avec un réseau de neurones et une régression logistique

Tableau 15 : Problématiques des études ne faisant pas intervenir de réseau de neurones dans l'élaboration des modèles

Études	Problématiques
ALTMAN (1968)	Montrer que les ratios peuvent être considérés comme de bons éléments d'analyse de la performance des entreprises, la prévision de la défaillance étant utilisée à titre d'illustration.
ALTMAN (1973)	Construire un modèle de prévision de la défaillance dans le domaine ferroviaire
ALTMAN et al. (1977)	Construire un nouveau modèle de prévision de la défaillance en tenant compte des derniers développements dans ce domaine
ALTMAN et al. (1995)	Construire un modèle de prévision de la défaillance pour des entreprises coréennes

AZIZ et al. (1988)	Tester l'efficacité d'un modèle théorique dit « égalité de Lawson », articulé autour des composantes du Cash Flow, pour prévoir la défaillance
BARANOFF et al. (2000)	Comparer le pouvoir prédictif d'un modèle de prévision de la défaillance construit à l'aide d'une régression Spline à celui d'un modèle élaboré à partir d'un Logit et d'un Probit
BARNIV et RAVEH (1989)	Présenter une nouvelle méthode d'analyse discriminante non paramétrique pour prévoir la défaillance d'entreprises
BARNIV et HERSHBARGER (1990)	Trouver les variables permettant de mesurer le degré de solvabilité des compagnies d'assurance et tester l'efficacité de méthodes multivariées alternatives pour surveiller la solvabilité des assureurs
BARNIV et McDONALD (1992)	Étudier les problèmes méthodologiques posés par l'emploi des techniques statistiques traditionnelles et tester les capacités prédictives de méthodes alternatives
BARNIV et al. (2002)	Montrer que l'on peut prévoir l'issue de la faillite – dépôt de bilan – d'une entreprise : acquisition, nouvelle entreprise, liquidation
BEAVER (1966)	Montrer que les ratios financiers peuvent présenter une utilité pour prédire la défaillance d'entreprise
BEAVER (1968)	Étudier dans quelle mesure les variations du cours de l'action d'une entreprise peuvent être employées pour prédire la défaillance et comment associer ce type de d'information à celle que l'on peut tirer des ratios dans la même optique de prévision
BEYNON et PEEL (2001)	Poursuivre les efforts de recherche sur la capacité prédictive des modèles de défaillance conçus à partir de la théorie des ensembles grossiers en étudiant les propriétés d'un modèle construit sur la base d'une extension de cette théorie nommée : ensemble grossier de précision variable
BLUM (1974)	Développer un modèle de prévision de la défaillance en incluant des variables incorporant le temps
BOOTH (1983)	Étudier l'utilisation de mesures de décomposition de certains agrégats financiers dans un modèle de prévision de la défaillance
BRABAZON et O'NEILL (2004)	Déterminer la capacité d'une méthode fondée sur un algorithme d'évolution grammaticale à extraire d'un ensemble de ratios financiers, suffisamment d'informations pour prévoir la défaillance d'entreprises
CHAVA et JARROW (2004)	Montrer la supériorité d'un modèle de survie, en l'occurrence celui de Shumway (2001), par rapport aux modèles d'Altman (1968) et de Zmijewski (1984), pour prévoir la défaillance d'entreprises, et tester l'effet du secteur sur la probabilité de défaillance
CIELEN et al. (2004)	Comparer l'efficacité de trois techniques d'apprentissage différentes pour construire des modèles de prévision de la défaillance
CYBINSKI (1995)	Évaluer l'efficacité d'un modèle construit à l'aide d'une régression logistique et d'un ensemble de variables binaires représentant chacune une catégorie issue de la discrétisation de ratios financiers
DAMBOLENA et KHOURY (1980)	Tester une façon de stabiliser les ratios des entreprises défaillantes et étudier l'incidence de cette technique sur la capacité de prévision d'un modèle
DEAKIN (1972)	Répliquer l'étude de Beaver (1966) à partir des meilleurs ratios puis rechercher la meilleure combinaison linéaire des ratios utilisés
DIMITRAS et al. (1999)	Tester l'efficacité d'un modèle construit à partir de la théorie des ensembles grossiers pour prévoir la défaillance
DOUMPOS et ZOPOUNDIS (1999)	Tester le potentiel d'une méthode particulière de discrimination non paramétrique pour construire un modèle de prévision de la défaillance
DREZNER et al. (2001)	Montrer l'intérêt que peut présenter une méthode de sélection de variables de type Tabu Search pour construire un modèle de prévision
EDMISTER (1972)	Tester l'utilité des ratios financiers pour prévoir la défaillance de petites entreprises
EL HENNAWY et MORRIS (1983)	Développer une série de modèles pour prévoir la défaillance pas simplement un an avant qu'elle ne survienne, mais 2, 3 4 et 5 ans avant, et tester la capacité prédictive de variables autres que des ratios financiers, comme la situation économique ou le secteur d'appartenance de l'entreprise, afin de construire des modèles plus universels que ceux qui l'ont été jusqu'à présent
ELMER et BOROWSKI (1988)	Tester l'efficacité d'un système expert pour construire un modèle de prévision
FITZPATRICK (1932)	Tester l'emploi des ratios pour étudier la santé financière des entreprises
FLAGG et al. (1991)	Élaborer un modèle de prévision capable de prévoir la faillite d'entreprises faisant partie de la classe des entreprises défaillantes
FOREMAN (2003)	Démontrer la capacité prédictive des ratios financiers traditionnels pour déterminer la défaillance d'entreprises appartenant au secteur des télécommunications
FRYDMAN et al. (1985)	Tester une nouvelle méthode – arbre de classification – pour élaborer un modèle de prévision
GENTRY et al. (1985)	Tester l'emploi d'un ensemble de ratios issus d'un modèle de gestion des liquidités pour créer un modèle capable de discerner une entreprise saine d'une entreprise défaillante
GENTRY et al. (1987)	Comparer le pouvoir prédictif d'un modèle de prévision utilisant des ratios choisis de manière empirique avec celui d'un modèle employant des ratios issus d'un cadre théorique de la défaillance d'entreprise

GENTRY et al. (2002)	À partir d'une modélisation théorique de la défaillance fondée sur l'information associée au Cash Flow, proposer un système permettant de mesurer l'incertitude véhiculée par le Cash Flow et prévoir la défaillance
GILBERT et al. (1990)	Construire un modèle permettant de prendre en compte trois groupes d'entreprises – saines, défaillantes, faillites – et isoler les variables susceptibles de distinguer les entreprises saines des entreprises en faillite ainsi que les entreprises défaillantes des entreprises en faillite
GODLEWSKI (2003)	Tester l'efficacité des ratios appartenant aux catégories décrites par la typologie Camel pour prévoir la défaillance bancaire dans les pays émergents
GOMBOLA et al. (1987)	Étudier le rôle du Cash Flow comme variable permettant de prévoir la défaillance d'entreprises
GRICE et INGRAM (2001)	Étudier la capacité du modèle d'Altman (1968) à prévoir la faillite sur des données récentes et sur des données d'entreprises non industrielles ainsi que sa capacité à prévoir d'autres états de défaillance que celui de la faillite
GRICE (2003)	Étudier la capacité des modèles d'Ohlson (1980) et de Zmijewski (1984) à prévoir la faillite sur des données récentes et sur des données d'entreprises non industrielles ainsi que leur capacité à prévoir d'autres états de défaillance que celui de la faillite
GUPTA et al. (1990)	Montrer l'intérêt d'utiliser la programmation linéaire par objectif pour élaborer un modèle de prévision de la défaillance
HÄRDLE et al. (2005)	Montrer l'efficacité d'un séparateur à vaste marge pour construire un modèle de prévision
HER et CHOE (1999)	Comparer le pouvoir prédictif de modèles de prévision de la défaillance élaborés à partir de données financières décrivant des entreprises australiennes et coréennes
HUA et al. (2007)	Tester l'efficacité d'une approche intégrant un séparateur à vaste marge avec une régression logistique pour construire un modèle de prévision de la défaillance
HUNTER et ISACHENKOVA (2001)	Étudier les différences pouvant exister entre un modèle de prévision de la défaillance adapté aux entreprises britanniques et un modèle adapté aux entreprises russes
HWANG et al. (1997)	Construire un modèle de prévision de la défaillance adapté au secteur bancaire et proposer un modèle de calcul de la prime d'assurance couvrant le risque de défaillance
JONES et HENSHER (2004)	Étudier l'efficacité d'un Logit mixte pour construire un modèle de prévision de la défaillance
KANE et al. (1998)	Tester l'efficacité de la transformation de ratios financiers en données ordinales sur les performances d'un modèle de prévision de la défaillance
KARELS et PRAKASH (1987)	Étudier l'incidence du respect de l'hypothèse de multi-normalité d'un ensemble de ratios sur les capacités prédictives d'un modèle issu de l'analyse discriminante
KEASEY et WATSON (1987)	Tester l'efficacité d'un modèle contenant des variables non financières utilisées conjointement ou non avec des ratios financiers traditionnels pour prévoir la défaillance des PME
KEASEY et MCGUINNESS (1990)	Déterminer les variables susceptibles d'expliquer la défaillance financière et étudier l'apport d'une mesure d'entropie de l'information contenue dans une fonction logistique de prévision de la défaillance
KOLARI et al. (2002)	Comparer l'efficacité d'un modèle construit à l'aide d'une régression logistique avec celle d'un modèle élaboré avec une méthode non paramétrique dite de « reconnaissance de traits » pour prévoir la défaillance d'entreprise
KURUPPU et al. (2003)	Comparer l'efficacité d'un modèle permettant de prévoir la liquidation avec celle d'un modèle permettant de prévoir la faillite face à un problème de prévision de la liquidation
LAITINEN (1991)	Identifier la présence de différents processus de défaillance et déterminer les dimensions importantes qui affectent les ratios des entreprises défaillantes
LAITINEN et LAITINEN (1995)	Proposer une méthode permettant d'identifier deux types d'entreprises défaillantes (celles défaillant pour cause de liquidité et celles défaillant pour cause de solidité financière insuffisante), puis d'évaluer la probabilité de chaque firme d'appartenir à l'une de ces catégories et enfin d'utiliser cette probabilité pour estimer la probabilité finale de défaillance
LAITINEN et LAITINEN (1998)	Tester l'efficacité d'un modèle de prévision de la défaillance fondé sur une représentation dynamique de la gestion des liquidités par rapport à un modèle construit à partir d'une représentation traditionnelle de la santé financière d'une entreprise élaborée avec des ratios
LAITINEN et LAITINEN (2000)	Tester l'utilisation d'un développement de Taylor pour résoudre le problème lié à la forme de la fonction employée par les modèles de prévision traditionnels
LANINE et VANDER VENNET (2005)	Développer un modèle de prévision pour analyser les déterminants de la défaillance de banques russes et tester l'efficacité de différentes méthodes
LAU (1987)	Construire un modèle de prévision permettant de déterminer cinq catégories de situations financières : une catégorie correspondant aux entreprises saines et quatre catégories correspondant à quatre niveaux de défaillance (absence de paiement de dividendes, défaut de remboursement de prêt, mise sous protection des lois sur la faillite, faillite ou liquidation)
LEE et URRUTIA (1996)	Comparer la performance d'un modèle de prévision de la défaillance de compagnies d'assurance construit avec une régression logistique avec celle d'un modèle élaboré à partir d'une méthode de survie

LENNOX (1999)	Montrer qu'un modèle de prévision de la défaillance construit avec une régression logistique ou un Probit bien ajusté donne de meilleurs résultats que s'il est construit avec une analyse discriminante et analyser l'influence du secteur d'activité, de la taille de l'entreprise et du cycle économique sur la probabilité de défaillance
LENSBERG et al. (2006)	Construire un modèle de prévision suffisamment simple et riche pour permettre d'améliorer aussi bien la prévision de la défaillance que la connaissance du phénomène sous-jacent
LIN et PIESSE (2004)	Construire un modèle de prévision avec une régression logistique employant une technique d'estimation des coefficients tenant compte de l'absence d'aléa dans le choix des échantillons
LINDSAY et CAMPBELL (1996)	Étudier la capacité d'un modèle dynamique non-linéaire dérivée de la théorie du chaos à prévoir la défaillance d'entreprises
LUSSIER (1995)	Développer un modèle non financier permettant de prévoir la défaillance des petites et moyennes entreprises
MCGURR et de VANEY (1998)	Tester l'efficacité de modèles de prévision élaborés à partir d'échantillons contenant des entreprises émanant de différents secteurs d'activité pour prévoir la défaillance d'entreprises appartenant à un secteur spécifique
MCKEE et LENSBERG (2002)	Tester l'efficacité d'un modèle construit par programmation génétique à partir de variables identifiées à l'aide d'une méthode dérivée de la théorie des ensembles grossiers
MENSAH (1984)	Étudier l'incidence des critères de choix des observations et celle des contraintes liées aux techniques statistiques sur la construction des modèles de prévision de la défaillance
MERWIN (1942)	Étudier la façon dont les petites entreprises obtiennent et utilisent leurs fonds et analyser leur rentabilité et les symptômes de leur défaillance et, dans ce derniers cas, étudier le rôle des ratios financiers pour rendre compte de leur problème d'insolvabilité
MEYER et PIFER (1970)	Trouver une règle de classification susceptible de prévoir la défaillance des banques
MOSSMAN et al. (1998)	Étudier le pouvoir prédictif de différents modèles selon le type de données employé : ratios financiers, ratios de liquidité, variables de marché, variation de variables de marché
MOYER (1977)	Tester l'efficacité du modèle d'Altman (1968) sur des données postérieures à son élaboration
NAM (2000)	Construire un modèle de prévision en période de crise économique
NANDA et PENDHARKAR (2001)	Montrer que la prise en compte du coût d'une mauvaise classification permet d'améliorer la prévision de défaillance
NEOPHYTOU et MAR-MOLINERO (2004)	Étudier l'utilisation d'une méthode d'analyse multidimensionnelle des similarités pour prévoir la défaillance
NORTON et SMITH (1979)	Comparer la performance d'un modèle de prévision fondé sur des indicateurs financiers tenant compte de l'évolution des prix avec celle d'un modèle employant des indicateurs n'en tenant pas compte
OHLSON (1980)	Prévoir la défaillance avec un modèle probabiliste
PACEY et PHAM (1990)	Étudier les erreurs méthodologiques commises habituellement par les études qui ont pour objectif de construire des modèles de prévision et élaborer un modèle selon une méthode permettant d'éviter ces erreurs
PANTALONE et PLATT (1987)	Construire un modèle de prévision de défaillance propre aux établissements de crédit
PEEL et PEEL (1987)	Développer des modèles à partir d'échantillons contenant des entreprises en mauvaise santé financière et classées comme entreprises saines, car d'autres études les ont retirées des données
PHILOSOPHOV et PHILOSOPHOV (2002)	Construire un modèle susceptible de prévoir non seulement la probabilité de défaillance d'un événement mais aussi son horizon en indiquant l'intervalle de temps à l'intérieur duquel elle se produira
PINCHES et TRIESCHMANN (1974)	Examiner l'efficacité des modèles statistiques univariés et multivariés permettant de surveiller la solvabilité des compagnies d'assurance
PLATT et PLATT (1990)	Tester l'efficacité de ratios pondérés par secteur c'est-à-dire normalisés par la moyenne du secteur de l'entreprise afin de réduire l'instabilité des modèles multi-sectoriels
PLATT et PLATT (2002)	Élaborer un modèle de prévision de la détresse financière et non de la faillite et tester les biais pouvant exister lorsque l'on utilise une méthode de stratification pour construire l'échantillon
PLATT et al. (1994)	Réexaminer l'effet de la pondération des ratios de façon à construire des modèles plus précis en éliminant le biais temporel mais cette fois en pondérant chaque variable par un critère qui lui est propre et non plus toutes les variables par le même critère
RAHMAN et al. (2004)	Construire un modèle de prévision de la défaillance adapté aux banques asiatiques et identifier les variables les plus pertinentes pour une telle tâche
RAMSER et FOSTER (1931)	Évaluer dans quelle mesure une analyse financière conduite sur la base de ratios calculés à partir de données issues d'un bilan permet de déterminer les entreprises susceptibles de survivre
RICHARDSON et al. (1998)	Étudier la sensibilité des modèles de prévision de la défaillance employant des variables comptables aux conditions économiques générales et notamment aux effets d'une récession

SALCEDO-SANZ et al. (2004)	Proposer deux techniques de sélection de variables à utiliser conjointement avec un séparateur à vaste marge pour construire un modèle de prévision de la défaillance
SALCEDO-SANZ et al. (2005)	Étudier le potentiel de la programmation génétique pour construire un modèle de prévision de la défaillance de compagnies d'assurance
SARKAR et SRIRAM (2001)	Examiner la capacité des modèles probabilistes à élaborer des systèmes de prévision de la défaillance bancaire et tester l'efficacité d'un modèle bayésien simulant le processus d'analyse d'un auditeur chargé de diagnostiquer la santé financière d'un établissement bancaire afin de prévoir sa probabilité de survie
SHARMA et ISELIN (2003)	Étudier la pertinence de l'information contenue dans les ratios de liquidité par rapport à celle contenue dans les ratios financiers traditionnels pour déterminer le niveau de solvabilité des entreprises
SHARMA et MAHAJAN (1980)	Développer un modèle du processus de la faillite
SHIN et LEE (2003)	Tester l'efficacité d'un algorithme génétique pour élaborer un modèle de prévision
SHIRATA (1998)	Construire un modèle de prévision de la défaillance d'entreprises japonaises quelles que soient leur taille et leur secteur d'activité
SHUMWAY (2001)	Tester l'efficacité d'un modèle de survie utilisant des variables de marché pour prévoir la défaillance par rapport aux modèles d'Altman (1968) et de Zmijewski (1984)
SINKEY (1975)	Élaborer un modèle multicritères d'analyse et de prévision de la défaillance bancaire
SKOGSVIK (1990)	Comparer le pouvoir prédictif d'un modèle intégrant des ratios financiers calculés à partir de données historiques avec celui d'un modèle incorporant des ratios ajustés pour annuler l'effet de l'inflation
SUN et al. (2005)	Améliorer le modèle d'Altman (1968) en construisant un modèle adapté aux entreprises présentant des signes de défaillance
SUNG et al. (1999)	Construire des modèles de prévision de la défaillance adaptés aux situations économiques normales et aux situations de crise
TAFFLER (1983)	Montrer en quoi les états financiers peuvent être de bons instruments de prévision de la défaillance lorsqu'ils sont bien utilisés
THEODOSSIOU (1991)	Développer un modèle de prévision de la défaillance adapté aux entreprises de fabrication grecques
THEODOSSIOU (1993)	Construire un modèle permettant de détecter l'instant où les caractéristiques financières des entreprises rompent significativement avec leur niveau habituel afin de prévoir la défaillance de ces firmes
TIRAPAT (1999)	Étudier la défaillance dans un pays économiquement émergent ainsi que la capacité prédictive des ratios financiers dans ce contexte couplés à des variables macro économiques
TSAKONAS (2006)	Tester l'efficacité d'un réseau de neurones logiques élaboré par programmation génétique grammaticale pour prévoir la défaillance
TSENG (2005)	Tester l'efficacité d'un modèle de prévision permettant de représenter une relation floue entre un ensemble de variables explicatives et la probabilité de défaillance
VAN GESTEL et al. (2006)	Tester l'efficacité d'un séparateur à vaste marge pour construire un modèle de prévision de la défaillance
VARETTO (1998)	Comparer l'efficacité de deux types de modèles de prévision de la défaillance élaborés avec un algorithme génétique, le premier représentant une fonction linéaire des prédicteurs et le second, un ensemble de règles, avec celle d'un modèle issu d'une analyse discriminante
WANG (2004)	Comparer l'efficacité de deux méthodes de sélection de ratios employés pour construire des modèles de prévision de la défaillance
WEISS et CAPKUN (2005)	Tester l'influence du coût d'une mauvaise classification sur la performance d'un modèle de prévision
WESTGAARD et VAN DER WIJST (2001)	Estimer la probabilité de défaillance associée à chaque entreprise cliente d'un établissement bancaire
WILCOX (1973)	Tester un modèle dynamique de la prévision de défaillance fondé sur l'emploi de ratios financiers
WINAKOR et SMITH (1935)	Analyser la structure financière d'entreprises défallantes
ZAVGREN (1985)	Développer un nouveau modèle de prévision et une méthode d'évaluation du risque financier
ZMIJEWSKI (1984)	Étudier les biais d'échantillonnage qui peuvent apparaître lorsque les modèles de prévision sont construits sur des échantillons non aléatoires
ZOPOUNIDIS et DOUMPOS (2002)	Proposer une nouvelle méthode de discrimination

1.4 MODALITÉS DE CONSTITUTION D'UN ÉCHANTILLON

Un certain consensus s'est structuré au fil du temps autour du cadre expérimental général à adopter pour construire des systèmes de prévision. Comme on a pu le voir, la définition de la

défaillance, ses facteurs explicatifs comme les variables qui les sous-tendent font partie de ces éléments sur lesquels il y a peu de divergence d'appréciation. Maintenant, il n'existe pas à proprement parler de ligne directrice quant à la façon de construire des échantillons de données d'où seront extraits les modèles. Doit-on tirer aléatoirement les observations ou les choisir à partir de critères posés *a priori* ? Peut-on créer des modèles universels, quels que soient les attributs des entreprises (pays, taille, secteur...), ou doit-on envisager de les spécialiser, et si oui, selon quels paramètres ? Peut-on employer des données prélevées à différentes périodes où respecter une contrainte de temps ? Quel horizon de prévoyance faut-il adopter et existe-t-il un délai au-delà duquel on ne peut plus rien prévoir raisonnablement ? Voici une série de questions qui n'est pas tranchée définitivement et que nous allons examiner pour tenter de discerner s'il existe des constantes auxquelles il conviendrait de se référer.

1.4.1 ÉCHANTILLONNAGE

La constitution des échantillons est une question qui se pose avec toute méthode d'analyse de données, mais ici, elle se pose en des termes propres aux méthodes reposant sur un apprentissage. Les modèles que l'on cherche doivent être capables de généralisation, donc savoir reconnaître des éléments qui n'ont pas fait partie de leur entraînement. Le corpus qui sert à établir ces modèles doit donc être représentatif de la variété des « formes » auxquelles ils risquent d'être confrontés lors d'une prévision. Le choix des observations est alors tributaire de la structure à modéliser : il s'agit avant tout d'appréhender les relations particulières qui unissent les objets d'un même groupe et qui se démarquent de celles qui existent dans un autre.

Comme les modèles de prévision « standards » sont des modèles dichotomiques, qui traduisent une règle de classification censée reconnaître deux groupes, des firmes défaillantes et des firmes qui ne le sont pas, ils doivent donc être calculés à partir d'informations les décrivant tous les deux le mieux possible.

Pour y parvenir, la technique consiste à construire un échantillon regroupant un certain nombre d'entreprises de chaque classe, n_1 et n_2 (avec $n = n_1 + n_2$), par tirage dans la population de référence de taille N . Il faut donc prélever aléatoirement n firmes parmi N , et comme il est nécessaire de disposer d'un nombre conséquent et non aléatoire d'observations dans chacun de deux groupes, il est recommandé de procéder à un sondage stratifié avec tirage séparé des deux catégories d'entreprises. La question de la répartition des effectifs par

classe se pose alors. On pourrait envisager une répartition identique à celle de la population, d'autant plus volontiers qu'un échantillonnage stratifié à répartition proportionnelle est meilleur qu'un échantillonnage simple sans stratification (Saporta, 1990). S'il en était ainsi, compte tenu de la problématique qui nous intéresse, l'échantillon verrait alors une classe particulièrement sous représentée, puisque le taux annuel moyen de défaillance se situe aux alentours de 1,6 % (taux de 2005), et les informations essentielles décrivant les entreprises financièrement en difficulté se trouveraient noyées parmi les autres. Donc, une répartition proportionnelle est à déconseiller, et il vaut mieux lui préférer une répartition équilibrée, où n_1 et n_2 sont égaux.

La stratification avec une représentation équilibrée des observations de chaque groupe est la méthode traditionnellement employée depuis Altman (1968), mais a néanmoins fait l'objet de quelques critiques. La plus sérieuse, émise notamment par Zmijewski (1984) et Platt et Platt (2002), considère que la non représentativité statistique d'un échantillon pose problème car elle conduit, *in fine*, à une surestimation du taux de classement correct des entreprises défaillantes, surestimation appréhendée par rapport à un modèle calculé sur un échantillon représentatif de la population étudiée. Ce biais n'affecterait pas le taux global de prévision, mais simplement les taux respectifs associés à chaque groupe. L'argument est curieux car, en l'occurrence, la véritable difficulté consiste précisément à obtenir un modèle qui puisse, en général, plutôt maximiser le taux de classement correct des entreprises défaillantes. Le coût d'une erreur de prévision est fortement asymétrique, notamment si l'on se place du point de vue du créancier, comme une banque, qui souhaite, par exemple, apprécier un risque associé à un emprunt qu'un de ses clients contracterait auprès d'elle. Dans une telle situation, elle encourt deux risques, car elle peut refuser à tort d'octroyer ce prêt ou bien accepter à tort de le consentir. D'un côté, elle court un risque lié à la perte d'une affaire, de l'autre un risque lié au non remboursement d'une dette, le premier représentant un coût bien inférieur au second. C'est pourquoi, à performance générale égale, il vaut mieux disposer d'un modèle qui ait plutôt tendance à mieux classer les entreprises défaillantes que les autres, puisque ceci revient implicitement à employer une technique de réduction du coût des erreurs.

Les résultats des études empiriques illustrent assez clairement les incidences d'une absence d'équilibre entre les classes. Le tableau 16 ci-après, qui recense les expérimentations ayant volontairement construit des échantillons de tailles inégales, montre que cette stratégie conduit

presque systématiquement à des modèles donnant des résultats de prévision, extrêmement précis pour une classe, et très mauvais pour l'autre. Seul Godlewski (2003) réussit à obtenir des taux assez similaires, mais néanmoins pas très élevés. Dans l'ensemble, d'ailleurs, les modèles sont incapables de reconnaître correctement les entreprises défailtantes. On retrouve bien ici, le motif qui justifie la stratification avec répartition proportionnelle et qui suggère qu'en cas d'inégalité marquée, les attributs d'une partie des données risquent d'être masqués par ceux de la partie qui domine.

Tableau 16 : Influence de la composition des échantillons d'apprentissage sur les résultats des modèles

Études	Nombre d'entreprises de chaque groupe		% de défailtantes	Taux de classement correct de chaque groupe et total		
	Saines	Défailtantes		Saines	Défailtantes	Total
Études	Saines	Défailtantes		Saines	Défailtantes	Total
BERG (2005)	60 000	600	1,00%	?	?	78,40%
WESTGAARD et VAN DER WIJST (2001)	35 287	954	2,70%	97,89%	77,27%	?
TAN et DIHARDJO (2001)	1 402	46	3,28%	92,35%	64,70%	92,23%
ZMIJEWSKI (1984)	800	40	5,00%	99,50%	70,70%	98,20%
OHLSON (1980)	2 058	105	5,10%	82,60%	87,60%	82,84%
RICHARDSON et al. (1998)	2 000	128	6,40%	45,95%	94,53%	48,87%
GODLEWSKI (2003)	1 624	105	6,47%	81,50%	83,20%	81,70%
KARELS et PRAKASH (1987)	168	14	8,33%	96,00%	54,50%	?
KANE et al. (1998)	913	82	8,98%	94,30%	50,00%	?
BELL et al. (1990)	906	102	11,26%	97,70%	61,00%	94,00%
GRICE et INGRAM (2001)	824	148	17,96%	78,80%	96,00%	83,50%
MCKEE et GREENSTEIN (2000)	288	54	18,75%	96,88%	58,33%	96,48%
TUNG et al. (2004)	511	110	21,53%	99,31%	54,54%	73,12%

Nous n'avons indiqué dans ce tableau que les études employant des échantillons dont la proportion entre les entreprises défailtantes et non défailtantes est significativement différente de 1, et qui mentionnent les taux de prévision respectifs de 2 groupes, ou le taux global.

La deuxième et la troisième colonne du tableau correspondent au nombre d'entreprises de chaque groupe ayant servi au calcul des modèles ; il s'agit donc de la taille des échantillons d'apprentissage.

La troisième colonne précise le rapport entre le nombre d'entreprises saines de l'échantillon et le nombre de défailtantes.

Les trois dernières colonnes indiquent, respectivement, le pourcentage de classement correct des entreprises saines, des entreprises défailtantes, et de toutes les entreprises, mais calculé sur un échantillon de test, différent de l'échantillon d'apprentissage. Nous n'avons reporté que les meilleurs résultats de chaque étude, qui en comptent souvent énormément, et uniquement lorsque la sélection de tels résultats avait un sens, chaque étude pouvant faire appel à plusieurs techniques de classification comme à différentes conditions expérimentales (en faisant varier les prédicteurs, les dates de prélèvement des données, la structure des échantillons et leur composition, le coût d'une erreur de prévision...).

Le tableau 16 peut être rapproché utilement du tableau 3, p. 19, qui indique les résultats obtenus par des modèles neuronaux. En effet, toutes les expérimentations indiquées dans le tableau 3, et que l'on ne retrouve pas dans le 16, et pour lesquelles nous donnons les pourcentages de prévision, ont été conduites à l'aide d'échantillons composés le plus souvent équitablement, ou à tout le moins en respectant des proportions qui ne déséquilibrent pas fondamentalement les

classes. On constate, dans le tableau 3, que l'équilibrage des effectifs tend à réduire de manière drastique l'écart entre les taux. On voit aussi que les modèles ont une plus grande propension à mieux prévoir ce qu'il adviendra des entreprises défaillantes qu'ils ne le font pour les autres. Ainsi, dans le tableau 3, le taux de prévision correcte des firmes en bonne santé est supérieur à celui des défaillantes dans 2 cas sur 3, contre 3 cas sur 4 dans le tableau 16.

Wilson et Sharda (1994) confirment d'ailleurs que la précision des modèles est fonction des effectifs dans chaque classe et surtout montrent que la qualité des résultats est fonction de l'équilibre obtenu dans l'échantillon d'apprentissage, une répartition égale étant de loin préférable à toute autre forme de répartition.

La problématique de la stratification ne se limite pas à la seule prise en compte du critère qui gouverne la taille optimale des groupes. Elle englobe aussi l'analyse d'autres paramètres qui pourraient à eux seuls influencer de manière significative sur les résultats. Qu'advierait-il d'un échantillon stratifié uniquement comme nous venons de le voir ? On aurait vraisemblablement un ensemble dominé essentiellement par les petites entreprises et, qui plus est, des entreprises de création récente. Il suffit pour s'en convaincre de regarder toute statistique en la matière. Par exemple, au premier trimestre 2006, plus de 86 % des défaillances en France ont affecté les entreprises de moins de 5 salariés, et plus de 73 % de ces mêmes défaillances se sont portées sur des entreprises ayant moins de dix ans d'activité¹⁰. Un modèle conçu sur un tel échantillon disposerait donc d'un domaine de validité bien plus restreint qu'on ne pourrait l'imaginer de prime abord, puisqu'au lieu de discriminer les défaillantes des autres, il discriminerait en fait les petites entreprises des grandes, celles de création récente des anciennes, ou pourquoi pas les entreprises industrielles des entreprises de services...

Cette question se pose depuis Beaver (1966), qui indiquait déjà, après analyse, que les ratios financiers ne peuvent pas être comparés directement, si les structures d'actifs des entreprises sont par trop différentes ou s'ils appartiennent à des firmes faisant partie de secteurs différents. Gupta (1969), lui, a aussi signalé que les ratios étaient sensiblement influencés par la taille et le secteur des entreprises, et qu'ils pouvaient l'être aussi par le rythme de croissance des firmes. Par la suite, d'autres travaux ont souligné l'importance de bâtir des modèles propres à certaines catégories d'entreprises. Edmister (1972), en suggérant que tout ratio doit être pondéré par la moyenne du secteur auquel l'entreprise appartient, reconnaît

¹⁰ Source : Les défaillances d'entreprises en France au premier trimestre 2006, Altares, Les études trimestrielles, 7 p.

implicitement que les règles de prévision diffèrent selon ce critère. Altman (1973), lui, considère explicitement que le secteur est la variable-clé et que tout bon modèle ne peut être que sectoriel et Altman et al. (1977) ajoutent à ce premier paramètre, un second représenté par la taille. Tout ceci est confirmé par Lennox (1999) qui montre que la probabilité de défaillance est fortement affectée par la taille et le secteur, et par Serrano-Cinca (2005) qui fait apparaître, en plus de l'effet du secteur, celui du pays.

Les pratiques expérimentales indiquent très clairement qu'il y a un consensus général sur les variables de contrôle dont il faut tenir compte. Le tableau 17 indiqué ci-après dresse un panorama des variables en question et de leur fréquence d'utilisation.

Tableau 17 : Critères de sélection des entreprises utilisés pour constituer les échantillons des études sur la défaillance

Études	Pays	Secteurs ou activités ou types d'entreprises	Critères de sélection
ALAM et al. (2000)	USA	Banques	Taille des actifs
ETHERIDGE et SRIRAM (1997)	USA	Banques	
GODLEWSKI (2003)	ASIE – AME. SUD – EUR. EST	Banques	
HWANG et al. (1997)	USA	Banques	
MEYER et PIFER (1970)	USA	Banques	Taille des actifs, localisation géographique, âge
RAHMAN et al. (2004)	INDO – COR – THA	Banques	
SARKAR et SRIRAM (2001)	USA	Banques	
SERRANO-CINCA (1997)	ESP	Banques	
SEXTON et al. (2003)	USA	Banques	
SINKEY (1975)	USA	Banques	Montant des dépôts, couverture géographique, nombre d'agences, statut auprès de la Réserve fédérale
TAM et KIANG (1990)	USA	Banques	Taille des actifs, nombre de branches, âge, statut
TAM et KIANG (1992)	USA	Banques	Taille des actifs, nombre de branches, âge, statut, période
TUNG et al. (2004)	USA	Banques	Période
KOLARI et al. (2002)	USA	Banques avec des actifs > 250 M\$	Taille des actifs, domiciliation
BELL et al. (1990)	USA	Banques commerciales	Taille des actifs
LANINE et VANDER VENNET (2005)	RUS	Banques commerciales	
SHARMA et MAHAJAN (1980)	USA	Commerce de détail	Secteur, taille des actifs
MCGURR et de VANEY (1998)	USA	Commerces	Période
BARNIV et McDONALD (1992)	USA	Compagnies d'assurance	Taille des actifs, période
BROCKETT et al. (1994)	USA	Compagnies d'assurance	Taille des actifs, activité, domiciliation
LEE et URRUTIA (1996)	USA	Compagnies d'assurance	Taille des actifs, période, domiciliation
PINCHES et TRIESCHMANN (1974)	USA	Compagnies d'assurance	Période

SALCEDO-SANZ et al. (2004)	ESP	Compagnies d'assurance	Chiffre d'affaires
SALCEDO-SANZ et al. (2005)	ESP	Compagnies d'assurance hors assurance-vie	Volume des primes
BARANOFF et al. (2000)	USA	Compagnies d'assurance-vie	
BARNIV et HERSHBARGER (1990)	USA	Compagnies d'assurance-vie	Domiciliation, taille des actifs, période
GOSS et RAMCHANDANI (1995)	USA	Compagnies d'assurance-vie	Période, taille des actifs, localisation géographique
HUANG et al. (1994)	USA	Compagnies d'assurance-vie	
ALTMAN (1973)	USA	Compagnies ferroviaires avec des revenus > 5 M\$	
POMPE et FEELDERS (1997)	BEL	Construction	Effectifs
YANG et HARRISON (2002)	GBR	Construction	
SUN et al. (2005)	USA	Construction, commerce de détail, services	
GENTRY et al. (1987)	USA	Différentes industries	Secteur, taille des actifs, chiffre d'affaires, période
DIMITRAS et al. (1999)	GRE	Différents	Secteur, taille des actifs, effectifs
DOUMPOS et ZOPOUNDIS (1999)	GRE	Différents	Secteur, taille des actifs, effectifs
TANG et CHI (2005)	CHI	Différents	Secteur, taille des actifs
AGARWAL et al. (2001)		Différents secteurs	Secteur, taille des actifs
ALTMAN et al. (1995)	COR	Différents secteurs	Secteur, période
FANNING et COGGER (1994)	USA	Différents secteurs	Secteur, période
GENTRY et al. (1985)	USA	Différents secteurs	Secteur, taille des actifs, chiffre d'affaires
GILBERT et al. (1990)	USA	Différents secteurs	Période
GRICE et DUGAN (2003)	USA	Différents secteurs	
GRICE et INGRAM (2001)	USA	Différents secteurs	
GUPTA et al. (1990)	USA	Différents secteurs	Secteur, taille des actifs
HER et CHOE (1999)	AUS	Différents secteurs	Secteur, taille des actifs, période
JO et al. (1997)	COR	Différents secteurs	Secteur
KEASEY et MCGUINNESS (1990)	GBR	Différents secteurs	Secteur, taille des actifs, période
KOTSIANTIS et al. (2005)	GRE	Différents secteurs	Secteur, taille des actifs, effectifs, période
LAITINEN et LAITINEN (1998)	FIN	Différents secteurs	Secteur, taille des actifs
LAITINEN et LAITINEN (2000)	USA	Différents secteurs	
LAU (1987)	USA	Différents secteurs	Taille des actifs
LENSBERG et al. (2006)	NOR	Différents secteurs	Secteur
LINDSAY et CAMPBELL (1996)	USA	Différents secteurs	Secteur
LUSSIER (1995)	USA	Différents secteurs	Secteur, effectifs, âge, domiciliation
MCKEE et GREENSTEIN (2000)	USA	Différents secteurs	
MOSSMAN et al. (1998)	USA	Différents secteurs	Secteur, taille des actifs
TSAKONAS (2006)	USA	Différents secteurs	Secteur, taille des actifs, période, chiffre d'affaires
WILCOX (1973)	USA	Différents secteurs	Taille des actifs, secteur, période
ZURADA et al. (1998)	USA	Différents secteurs	Secteur
ODOM et SHARDA (1990)	USA	Différents secteurs industriels	Secteur, période
RAHIMIAN et al. (1993)	USA	Différents secteurs industriels	Secteur, période

WILSON et SHARDA (1994)	USA	Différents secteurs industriels	Secteur, période
ZOPOUNIDIS et DOUMPOS (2002)	GRE	Différents secteurs industriels	Secteur, taille des actifs, effectifs
LEE et al. (2005)	COR	Différents secteurs sauf secteur financier	Secteur, taille des actifs
SHIRATA (1998)	JAP	Différents secteurs sauf services financiers et construction	
BOSE et PAL (2006)	USA	Dot Com cotées en bourse	
BERG (2005)	NOR	Entreprises à responsabilité limitée de différents secteurs	
LEE et al. (1996)	COR	Entreprises cotées en bourse	Taille des actifs, capital, effectifs, âge
SEN et al. (2004)	USA	Entreprises cotées en bourse	
LESHNO et SPECTOR (1996)	USA	Entreprises cotées de différents secteurs ayant des actifs > 10 M\$	Secteur, taille des actifs
PENDHARKAR (2005)	USA	Entreprises cotées de différents secteurs sauf secteur financier et immobilier	
BARNIV et al. (1997)	USA	Entreprises cotées de fabrication et de commerce de détails	
RAMSER et FOSTER (1931)	USA	Entreprises cotées en bourse	
BRABAZON et KEENAN (2004)	USA	Entreprises de différents secteurs cotées en bourse	Secteur, chiffre d'affaires
CHAVA et JARROW (2004)	USA	Entreprises de différents secteurs cotées en bourse	
HUA et al. (2007)	CHI	Entreprises de différents secteurs cotées en bourse	
JONES et HENSHER (2004)	AUS	Entreprises de différents secteurs cotées en bourse	
KANE et al. (1998)	USA	Entreprises de différents secteurs cotées en bourse	
LEE (2004)	THA	Entreprises de différents secteurs cotées en bourse	Secteur, taille des actifs, période
LENNOX (1999)	GBR	Entreprises de différents secteurs cotées en bourse	
MCKEE et LENSBERG (2002)	USA	Entreprises de différents secteurs cotées en bourse	Secteur, taille des actifs
NANDA et PENDHARKAR (2001)	USA	Entreprises de différents secteurs cotées en bourse	
NEOPHYTOU et MAR-MOLINERO (2004)	GBR	Entreprises de différents secteurs cotées en bourse	Secteur, taille des actifs
PLATT et PLATT (1990)	USA	Entreprises de différents secteurs cotées en bourse	Activité, taille des actifs, période
RICHARDSON et al. (1998)	USA	Entreprises de différents secteurs cotées en bourse	
SHARMA et ISELIN (2003)	AUS	Entreprises de différents secteurs cotées en bourse	Secteur, taille des actifs
SHUMWAY (2001)	USA	Entreprises de différents secteurs cotées en bourse	
TIRAPAT (1999)	THA	Entreprises de différents secteurs cotées en bourse	
WANG (2004)	USA	Entreprises de différents secteurs cotées en bourse	
WU et al. (2007)	THA	Entreprises de différents secteurs cotées en bourse	Secteur, taille des actifs, produits, capitalisation
YIM et MITCHELL (2002)	AUS	Entreprises de différents secteurs cotées en bourse	Secteur, taille des actifs, période
ZMIJEWSKI (1984)	USA	Entreprises de différents secteurs cotées en bourse	

WEISS et CAPKUN (2005)	USA	Entreprises de différents secteurs cotées en bourse avec des actifs > 100 M\$	
DREZNER et al. (2001)	USA	Entreprises de différents secteurs, sauf secteur financier	Taille des actifs
BRABAZON et O'NEILL (2004)	USA	Entreprises de différents secteurs, sauf secteur financier cotées en bourse	Secteur, chiffre d'affaires
CIELEN et al. (2004)	BEL	Entreprises de différents secteurs, sauf secteur financier et assurance	
KURUPPU et al. (2003)	NZE	Entreprises de différents secteurs, sauf secteur financier, cotées en bourse	
NAM et JINN (2000)	COR	Entreprises de différents secteurs, sauf secteur financier, cotées en bourse	Secteur, taille des actifs
OHLSON (1980)	USA	Entreprises de différents secteurs, sauf secteur financier, cotées en bourse	
HÄRDLE et al. (2005)	USA	Entreprises de différents secteurs ayant une capitalisation > 1 000 M\$	Secteur
TAFFLER (1983)	GBR	Entreprises de fabrication cotées en bourse	Taille des actifs, secteur
ALTMAN et al. (1977)	USA	Entreprises de fabrication et de commerce ayant des actifs > 100 M\$	Secteur, année de défaillance
THEODOSSIOU (1993)	USA	Entreprises de fabrication et de commerce de détail cotées en bourse	
SKOGSVIK (1990)	SUE	Entreprises de fabrication et entreprises minières ayant un effectif > 200 et des actifs > 200 000 M\$	
BARNIV et al. (2002)	USA	Entreprises de fabrication, de commerce et de distribution cotées en bourse	
BLUM (1974)	USA	Entreprises de l'industrie ayant un passif > 1 M\$	Chiffre d'affaires, effectif, période
LIN et PIESSE (2004)	GBR	Entreprises de l'industrie cotées en bourse	
PACEY et PHAM (1990)	AUS	Entreprises de l'industrie cotées en bourse	
TSENG (2005)	GBR	Entreprises de l'industrie cotées en bourse	
YIM et MITCHELL (2005)	BRE	Entreprises de l'industrie cotées en bourse	Taille des actifs, période
WILSON et al. (1995)	GBR	Entreprises de l'industrie cotées en bourse	
SHIN et al. (2005)	COR	Entreprises moyennes de fabrication	
SHIN et LEE (2003)		Entreprises moyennes de fabrication	
MIN et LEE (2005)	COR	Entreprises moyennes de l'Industrie lourde	
MIN et LEE (2004)	COR	Entreprises moyennes du secteur des technologies de l'information	
VIEIRA et al. (2003)	ESP	Entreprises non financières	
WESTGAARD et VAN DER WIJST (2001)	NOR	Entreprises non financières à responsabilité limitée de différents secteurs	
AZIZ et al. (1988)	USA	Entreprises non financières cotées en bourse	Secteur, taille des actifs
PLATT et PLATT (2002)	USA	Equipements automobiles	
ELMER et BOROWSKI (1988)	USA	Etablissements de crédit	Taille des actifs, domiciliation
PANTALONE et PLATT (1987)	USA	Etablissements de crédit	Taille des actifs, taille de la ville de domiciliation

SALCHENBERGER et al. (1992)	USA	Etablissements de crédit	Période, taille, localisation géographique
TYREE et LONG (1996)	GBR – USA	Etablissements de crédit	
TAN et DIHARDJO (2001)	AUS	Établissements de crédit	Taille des actifs
ALTMAN (1968)	USA	Fabrication	Secteur, taille des actifs, période
BACK et al. (1994)	FIN	Fabrication	Taille et période
BACK et al. (1996 A)	FIN	Fabrication	Taille et période
BEYNON et PEEL (2001)	GBR	Fabrication	Période
THEODOSSIOU (1991)	GRE	Fabrication	Période
ZAVGREN (1985)	USA	Fabrication	Taille des actifs, secteur
ZHANG et al. (1999)	USA	Fabrication	Secteur, taille des actifs
FRYDMAN et al. (1985)	USA	Fabrication, commerce	
LI et GUPTA (2002)	USA	Fabrication, commerce	
DAMBOLENA et KHOURY (1980)	USA	Fabrication, commerce de détail	Secteur
BEAVER (1966)	USA	Fabrication, commerce, distribution	Secteur, taille des actifs, période
BEAVER (1968)	USA	Fabrication, commerce, distribution	Secteur, taille des actifs, période
HUNTER et ISACHENKOVA (2001)	RUS	Fabrication, commerce, distribution	
BACK et al. (1997)	USA	Fabrication, commerces, services	
EL HENNAMY (1983)	GBR	Fabrication, construction, distribution	
MENSAH (1984)	USA	Fabrication, mine, commerce, construction	Secteur, taille des actifs
FLAGG et al. (1991)	USA	Fabrication, services	
KIM et HAN (2003)	COR	Fabrication, services	
PLATT et al. (1994)	USA	Gaz et pétrole	Période
YANG et al. (1999)	USA	Gaz et pétrole	Période
NORTON et SMITH (1979)	USA	Grandes entreprises industrielles cotées en bourse	Secteur, taille des actifs, période
BOOTH (1983)	AUS	Industrie	Secteur, taille des actifs, période
CHARITOU et al. (2004)	GBR	Industrie	Secteur, taille des actifs, période
DEAKIN (1972)	USA	Industrie	
GENTRY et al. (2002)	USA	Industrie	Secteur, chiffre d'affaires
GOMBOLA et al. (1987)	USA	Industrie	Secteur, taille des actifs
KARELS et PRAKASH (1987)	USA	Industrie	Secteur
LACHER et al. (1995)	USA	Industrie	Période
PHILOSOPHOV et PHILOSOPHOV (2002)	USA	Industrie	
COATS et FANT (1993)	USA	Industrie et services	Période
MIN et al. (2006)	COR	Industrie légère	
BARNIV et RAVEH (1989)	USA	Industrie, compagnies d'assurance	Période
VAN GESTEL et al. (2006)	BEL – HOL – LUX	Mid Cap ayant des actifs > 10 M€ et un chiffre d'affaires < 250 M€	
SUNG et al. (1999)	COR	Moyennes et grandes entreprises de différents secteurs cotées en bourse	
CYBINSKI (1995)	AUS	Petites entreprises	
FAN et PALANISWAMI (2000)	AUS	PME	

KEASEY et WATSON (1987)	GBR	PME	
LAITINEN (1991)	FIN	PME	Secteur, taille des actifs
LAITINEN et LAITINEN (1995)	FIN	PME	Secteur, effectifs, chiffre d'affaires, période
POMPE et BILDERBEEK (2005 A)	BEL	PME de l'industrie ayant des effectifs < 50 salariés	
POMPE et BILDERBEEK (2005 B)	BEL	PME de l'industrie ayant des effectifs < 50 salariés	
ALTMAN (1994)	ITA	PME de l'industrie ayant un chiffre d'affaires < 60 M\$	
BACK et al. (1996 B)	FIN	PME de différents secteurs	Période
LAITINEN et KANKAANPAA (1999)	FIN	PME de fabrication	Secteur, taille des actifs, période
MERWIN (1942)	USA	PME de fabrication	
KIVILUOTO (1998)	FIN	PME de l'industrie	
VARETTO (1998)	ITA	PME de l'industrie	
FOREMAN (2003)	USA	Télécommunication	
PEEL et PEEL (1987)	GBR	Très grandes entreprises de l'industrie	Bénéfice
CHARALAMBOUS et al. (2000)	USA		Taille des actifs, période
EDMISTER (1972)	USA		Petites entreprises
FITZPATRICK (1932)	USA		
WINAKOR et SMITH (1935)	USA		
AGARWAL (1999)	USA		Taille des actifs
ANANDARAJAN et al. (2001)	USA		Cotation en bourse
ATIYA (2001)	USA		
BAEK et CHO (2003)	COR		
BORITZ et KENNEDY (1995)			Période
DORSEY et al. (1995)	USA		
KIM et al. (2003)			Secteur, taille des actifs
KIM et MCLEOD (1999)	USA		Taille des actifs, activité, période
KUMAR et al. (1997)			
MOYER (1977)	USA		Secteur, taille des actifs
PIRAMUTHU et al. (1998)	BEL		Secteur, taille des actifs, chiffre d'affaires
WALLRAFEN et al. (1996)	ALL		
WEST et al. (2005)			
ALL :	Allemagne	ESP :	Espagne
AME. SUD :	Amérique du Sud	EUR. EST :	Europe de l'Est
ASIE :	Asie	FIN :	Finlande
AUS :	Australie	GBR :	Grande-Bretagne
BEL :	Belgique	GRE :	Grèce
BRE :	Bésil	HOL :	Hollande
CHI :	Chili	INDO :	Indonésie
COR :	Corée	ITA :	Italie
		JAP :	Japon
		LUX :	Luxembourg
		NOR :	Norvège
		NZE :	Nouvelle Zélande
		RUS :	Russie
		SUE :	Suède
		THA :	Thaïlande
		USA :	États-Unis

Ce tableau fait ressortir plusieurs constantes. D'abord, les modèles sont, dans la quasi-totalité des cas, délimités géographiquement, et propres à un seul pays, pour des raisons essentiel-

lement liées aux systèmes juridiques qui n'obéissent pas aux mêmes conceptions et règles régissant la défaillance, mais aussi aux spécificités des systèmes comptables nationaux qui ne facilitent pas l'emploi conjoint de données transnationales. Ensuite, à l'intérieur de ce premier niveau, les modèles sont construits sur la base de deux paramètres supplémentaires : un critère relatif au secteur (commerce, construction, fabrication, industrie, service, ...), à l'activité (banque, compagnies d'assurance...) ou au type d'entreprises (PME, grandes entreprises, entreprises cotées en bourse...), et un autre critère utilisé notamment lorsque le précédent n'est pas suffisant pour garantir une réelle comparabilité des firmes. C'est notamment le cas des échantillons volontairement multi-sectoriels où, pour annuler les effets induits par la disparité des origines des données¹¹, le tirage des entreprises est effectué par paires (une défaillante et une qui ne l'est pas) et ne sont retenues, par exemple, que celles qui sont identiques du point de vue du secteur et de la taille des actifs. C'est aussi le cas des échantillons regroupant certaines catégories de firmes, par exemple des banques ou des entreprises cotées, dès lors que la variété intrinsèque de telles sous-populations ne permet pas de constituer des ensembles homogènes. Là aussi les entreprises sont sélectionnées par paires, selon des critères similaires aux précédents : secteur, taille des actifs, chiffre d'affaires, localisation géographique, domiciliation, effectifs, âge... Enfin, le fait de focaliser une étude sur des banques et des institutions financières, des firmes industrielles ou des firmes cotées induit implicitement un troisième niveau de sélection qui conduit à privilégier les grandes entreprises. Il faut y voir la traduction d'un phénomène simple : ces entreprises publient des comptes facilement accessibles, critère très prisé, surtout si l'on souhaite disposer de données historiques qui s'étalent sur de longues périodes et sont, pour certains auteurs, plus fiables que ne le seraient ceux des PME. Donc, mécaniquement, ces entreprises font l'objet de toutes les attentions. À ceci se rajoute le poids économique de ce type d'organisation et le coût social de toute défaillance, qui donne une légitimité supplémentaire aux efforts de recherche dans leur direction.

Mais certaines procédures comme le « pairage » ne sont quand même pas totalement neutres. Choisir, comme il est de tradition, des paires d'entreprises sur la base de la taille, du secteur ou d'autres paramètres, en ne retenant que les défaillantes qui ont le même profil que les

¹¹ La disparité provient du fait que les secteurs diffèrent selon les facteurs de production, le cycle de vie de produit, les facteurs de compétitivité, les modes de distribution... qui entraînent des différences de structure financière. Ceci suggère que les ratios d'une entreprise reflètent la structure capitalistique d'un secteur ainsi que la structure de revenus et de dépenses du secteur. Ainsi, avec un échantillon composé d'entreprises appartenant à plusieurs secteurs, la moyenne et la variance des ratios dépendent non seulement des différences entre entreprises mais aussi des différences entre secteurs.

autres, peut conduire à des échantillons qui ne sont plus du tout représentatifs de la population qu'ils sont censés incarner. Zmijewski (1984) et Keasey et Watson (1991) ont montré qu'un choix peu judicieux pouvait nuire grandement à la capacité de généralisation des modèles, sans que ceci ne remette en cause l'idée même qui est au principe d'une telle sélection.

Maintenant, si tous les auteurs s'accordent sur la nécessité de spécialiser les modèles, ils ne donnent aucune indication précise sur la façon de délimiter chaque catégorie sous-jacente aux critères qui viennent d'être évoqués. Et l'on voit bien qu'en eux-mêmes ils ne sont pas suffisants. Comment appréhender un secteur ? Y a-t-il un niveau de détail optimal ? Il en est de même pour la taille des firmes. Quel est le facteur qui représente le mieux une taille : un chiffre d'affaires ou un montant d'actifs ? Quel découpage faut-il opérer ? Selon quelles tranches ? Où commencent et où s'arrêtent les grandes entreprises par rapport aux petites ? Quelles sont les frontières qui séparent un échantillon homogène d'un échantillon qui ne l'est pas ?

Ces questions trahissent des difficultés techniques et conceptuelles de catégorisation, car il faut à la fois être en mesure de constituer des échantillons de taille suffisante¹² pour qu'ils soient exploitables d'un point de vue statistique, sans être piégé par l'hétérogénéité qui risque d'en découler et en même temps définir des périmètres *a priori* mais sans pouvoir s'appuyer sur des points de repère qui découleraient de considérations théoriques.

Tout porte à croire qu'il n'y a donc que des catégories empiriques. On le voit bien dans les découpages qui affectent les secteurs ou les métiers du tableau 17, p. 58. Les services, dans certaines situations, sont envisagés dans leur ensemble et dans d'autres, le sont sans les services financiers. Les compagnies d'assurance peuvent être prises en compte dans leur globalité et parfois sans celles qui commercialisent des contrats d'assurance-vie. Les banques peuvent être uniquement des institutions commerciales ou regrouper tous les types d'activités. Les entreprises industrielles sont souvent envisagées à un niveau très agrégé de cette nomenclature, et parfois non¹³. On le voit aussi dans les autres formes de découpage : la taille

¹² Le « pairage » des entreprises représente une stratégie très utilisée par les auteurs qui travaillent sur de petits échantillons : ceci permet d'accroître le nombre d'observations en s'affranchissant des contraintes induites par l'explosion de la variance d'une variable de contrôle.

¹³ Ici aussi, les catégories sont souvent construites pour des motifs de représentativité statistique. Bardos (2001) remarque qu'au sein de l'industrie, il existe une certaine hétérogénéité sectorielle entre les entreprises (une firme de textile n'a pas le même profil qu'une firme de métallurgie) mais aussi que les situations de défaillance à ce niveau de détail sont rarement observables. Et de conclure que pour obtenir une réelle fiabilité statistique des prévisions, il est nécessaire que les modèles soient construits à un niveau suffisamment agrégé de ce secteur.

est mesurée selon des montants d'actifs variables, mais aussi selon d'autres critères comme les effectifs ou le chiffre d'affaires¹⁴, et là aussi en fonction d'intervalles instables.

Manifestement, ce sont les conditions expérimentales qui guident les processus de sélection, mais éclairées néanmoins par des pratiques donnant des indications très générales sur des critères d'importance (secteur, classe d'actifs) en dehors de toute référence à de quelconques seuils.

1.4.2 STATIONNARITÉ ET STABILITÉ DES DONNÉES

La plupart des modèles de prévision font l'hypothèse que la relation entre les variables est stable dans le temps et que leur distribution l'est aussi. D'un côté, le lien entre la variable dépendante (généralement celle qui indique un seuil de risque ou une probabilité de défaillance) et les variables indépendantes (les ratios...) est présumé stationnaire. Ceci est au principe de tout modèle de ce type, puisque construit sur la base de données historiques pour prévoir un état futur. De l'autre, les ratios ou les indicateurs servant de variables explicatives sont supposés ne pas être affectés par des facteurs exogènes à l'entreprise (inflation, taux d'intérêt, environnement concurrentiel, cycle économique...). Ce n'est pas systématique, mais à chaque fois qu'un modèle est calculé à partir de données relevées sur plusieurs années, cette hypothèse est bien en jeu.

Ces hypothèses sont-elles valides ? Existe-t-il un délai à l'intérieur duquel l'hypothèse de stationnarité tient ? Par extension, cette question renvoie à la durée de vie d'un modèle aussi bien du point de vue de ses coefficients que du point de vue des variables qu'il incorpore. On voit alors se dessiner une alternative : la structure essentielle peut-elle demeurer stable dans le temps, et reposer sur un socle identique de variables, et se satisfaire, comme condition de validité future, de réactualisations périodiques des poids affectant tel ou tel de ses éléments, ou cette structure possède-t-elle une durée de vie limitée ? Par ailleurs, dans quelle mesure peut-on accepter l'hypothèse de stabilité des données et quelle est l'incidence réelle de facteurs macro- ou méso-économiques sur la distribution des variables, et par ricochet, sur les modèles ?

La première hypothèse relative à la stabilité et à la durée de vie d'un modèle a été en partie testée à plusieurs reprises. Ainsi, les modèles d'Altman (1968), Ohlson (1980) et Zmijewski (1984)

¹⁴ Cette diversité, on la retrouve un peu partout. Ainsi, la Banque de France retient plutôt une définition de la taille fondée sur le chiffre d'affaires (très petites entreprises : CA ≤ 1,5 millions d'euros ; petites et moyennes entreprises : 1,5 < CA ≤ 50 millions d'euros ; grandes entreprises : > 50 millions d'euros) alors que l'Insee adopte une définition liée d'abord au nombre d'employés (très petites entreprises : 0-19 salariés ; petites et moyennes entreprises : 0-249 salariés ; grandes entreprises ≥ 250 salariés).

ont-ils donné lieu à de nombreux tests dits de « généralisabilité ». Ces modèles ont donc été appliqués en l'état à d'autres données que celles qui ont servi à leur élaboration, « autres » désignant des données comptables provenant de périodes différentes et d'entreprises appartenant à des secteurs différents de ceux et celles employés dans l'échantillon d'origine, puis ont été réestimés pour tenir d'éventuels changements (Moyer, 1977 ; Elmer et Borowski, 1988 ; Grice et Ingram, 2001 ; Grice et Dugan, 2003). Il ressort de ces expériences deux choses. Tout d'abord, les modèles appliqués en l'état donnent systématiquement de mauvais résultats et ce, quelles que soient les variations étudiées. Ensuite, une fois les coefficients des fonctions réestimés, ils donnent de bien meilleurs résultats, sans pour autant présenter le même degré de précision que les modèles originaux. L'écart provient essentiellement des entreprises défaillantes, que les modèles ne parviennent pas à classer correctement, contrairement aux entreprises saines, qui obtiennent des taux bien supérieurs.

Comment ces résultats peuvent-ils s'interpréter ? Ooghe et al. (1999) ont tenté d'apporter une réponse à cette question. Ils ont étudié le comportement de 7 modèles (Altman, 1968 ; Bilderbeek, 1979 ; Ooghe et Verbaere, 1982 ; Zavgren, 1985 ; Gloupos et Grammaticos, 1988 ; Keasey et McGuinness, 1990 ; Ooghe et al., 1991) face à des comptes d'entreprises belges, après transformation, afin de lisser les différences provenant du système comptable belge. Ces modèles ont obtenu des résultats conformes aux autres études, faisant apparaître des différences notables entre les modèles originaux et les répliques. Et quand il s'est agi d'expliquer ces écarts, aucun des critères étudiés n'a permis de fournir une réponse. Ni l'âge des modèles, ni la taille des entreprises utilisées pour leur mise au point initiale, ni la technique de discrimination, ni le nombre de variables – donc la complexité des modèles – ni la nationalité des firmes d'origine ne paraissent être en cause. Le seul point de convergence semble résider dans la proximité « géographique » des échantillons. Ainsi, les modèles d'Altman, Zavgren et Keasey et McGuinness, conçus à partir de comptes d'entreprises britanniques ou nord-américaines, présentent-ils dans l'ensemble de moins bonnes prévisions que ceux construits sur des échantillons de firmes européennes. Malgré ce point commun, aucune donnée objective ne permet d'expliquer les écarts. La seule chose qui demeure, c'est qu'aucun modèle ne peut être transposable tel quel dans un autre contexte que celui dont il est issu, vraisemblablement parce que trop marqué par les propriétés des échantillons d'apprentissage.

Les résultats des réajustements ne sont pas non plus probants et bien que ceci ne paraisse pas avoir été testé, il est assez probable que des structures de modèles radicalement différentes donneraient de bien meilleurs résultats. Les coefficients ne sont donc pas les seuls paramètres en jeu : les variables aussi ont un caractère conjoncturel, montrant que la défaillance est polymorphe et que ses formes varient en plus dans le temps. Bardos (1995) souligne qu'il existe des ratios au pouvoir discriminant général et permanent, toujours révélateurs de risques quand ils se dégradent. Mais en structurant son analyse plutôt sur des familles de risques et des ensembles de ratios au pouvoir prédictif, Bardos présente une argumentation qui reconnaît implicitement que seules les grandes forces qui concourent à structurer la défaillance sont stables, et que les variables dans lesquelles on les retrouvera sont contingentes¹⁵.

La seconde hypothèse ayant trait à la stabilité des données a aussi été étudiée en regard de pratiques comme celle d'Altman (1968) qui, en agrégeant des données sur une période de vingt ans, supposait une réelle homogénéité de cette période. C'est d'ailleurs l'hypothèse formulée souvent implicitement par de nombreux auteurs qui utilisent des données prélevées sur plusieurs années, comme indiqué dans le tableau 18 ci-après.

Tableau 18 : Périodes de prélèvement des entreprises défaillantes utilisées pour constituer les échantillons

Études	Périodes au cours desquelles les entreprises défaillantes ont été prélevées		Études	Périodes au cours desquelles les entreprises défaillantes ont été prélevées	
	Année de début	Année de fin		Année de début	Année de fin
AGARWAL (1999)	?	?	LEE et URRUTIA (1996)	1985	1991
AGARWAL et al. (2001)	?	?	LENNOX (1999)	1987	1994
ALAM et al. (2000)	1991	1991	LENSBERG et al. (2006)	1996	1997
ALTMAN (1968)	1946	1965	LESHNO et SPECTOR (1996)	1984	1988
ALTMAN (1973)	1939	1970	LI et GUPTA (2002)	1971	1981

¹⁵ Pour Bardos (1995), les ratios les plus discriminants illustrent le risque patrimonial, le risque d'illiquidité et les risques liés à l'activité productive, à la rentabilité financière et au financement. « Le risque patrimonial et le risque d'illiquidité peuvent être appréhendés au travers de l'équilibre de la structure du bilan par l'examen des ratios rapprochant des postes d'actif et de passif, habituellement d'une grande stabilité. Leur dégradation peut donc constituer des signes sérieux de la défaillance. Il en est ainsi de la couverture des immobilisations et du BFR d'exploitation par les ressources stables et du financement du fond de roulement par les concours bancaires courants. Viennent enrichir ces informations, le taux d'endettement qui rapporte l'endettement à une notion de financement propre et les délais de paiement (client et fournisseur) qui rapportent un poste de bilan à un poste du compte de résultats et dont l'allongement signale des difficultés ». Et de conclure, p. 60 : « Mais une analyse plus économique est nécessaire reliant activité productive, rentabilité et financement. Les ratios particulièrement significatifs dans ces domaines sont : les indicateurs de rentabilité (économique, globale et financière), les taux de marge, les coûts de production et les coûts de financement, les contraintes de solvabilité (frais fixe/EBE) et la capacité de remboursement, ainsi que l'effet de levier (taux d'endettement * (rentabilité nette du capital financier - taux d'intérêt apparent)). » Nul doute qu'il s'agit là d'indicateurs généraux. Mais l'absence d'un niveau supplémentaire de précision est bien la preuve que l'on ne peut pas aller bien au-delà dans le consensus.

ALTMAN (1994)	1982	1992
ALTMAN et al. (1977)	1969	1976
ALTMAN et al. (1995)	1990	1993
ANANDARAJAN et al. (2001)	1989	1996
ATIYA (2001)	?	?
AZIZ et al. (1988)	1971	1982
BACK et al. (1994)	1986	1989
BACK et al. (1996 A)	1986	1989
BACK et al. (1996 B)	1986	1989
BACK et al. (1997)	1985	1993
BAEK et CHO (2003)	1993	2000
BARANOFF et al. (2000)	1991	1992
BARNIV et al. (1997)	1980	1991
BARNIV et al. (2002)	1980	1992
BARNIV et HERSHBARGER (1990)	1975	1983
	1975	1985
BARNIV et McDONALD (1992)	1974	1988
BARNIV et RAVEH (1989)	1971	1981
BEAVER (1966)	1954	1964
BEAVER (1968)	1954	1964
BELL et al. (1990)	1985	1986
BERG (2005)	1995	2000
BEYNON et PEEL (2001)	1997	1998
BLUM (1974)	1954	1968
BOOTH (1983)	1964	1979
BORITZ et KENNEDY (1995)	1971	1984
BOSE et PAL (2006)	1993	2003
BRABAZON et KEENAN (2004)	1991	2000
BRABAZON et O'NEILL (2004)	1991	2000
BROCKETT et al. (1994)	1992	1989
CHARALAMBOUS et al. (2000)	1983	1994
CHARITOU et al. (2004)	1988	1997
CHAVA et JARROW (2004)	1962	1999
	1962	1999
CIELEN et al. (2004)	1994	1996
COATS et FANT (1993)	1973	1989
CYBINSKI (1995)	1980	1983
DAMBOLENA et KHOURY (1980)	1969	1975
DEAKIN (1972)	1964	1970
DIMITRAS et al. (1999)	1986	1990
DORSEY et al. (1995)	1989	1989

LIN et PIESSE (2004)	1985	1994
LINDSAY et CAMPBELL (1996)	1983	1992
LUSSIER (1995)	?	?
MCGURR et de VANEY (1998)	1989	1993
MCKEE et GREENSTEIN (2000)	1985	1989
	1984	1988
	1983	1987
	1982	1986
	1981	1985
MCKEE et LENSBERG (2002)	1991	1997
MENSAH (1984)	1972	1973
	1973	1975
	1976	1977
	1978	1980
MERWIN (1942)	1927	1936
MEYER et PIFER (1970)	1948	1965
MIN et al. (2006)	1999	2002
MIN et LEE (2004)	1999	2002
MIN et LEE (2005)	2000	2002
MOSSMAN et al. (1998)	1980	1991
	1980	1991
MOYER (1977)	1965	1975
NAM et JINN (2000)	1997	1998
NANDA et PENDHARKAR (2001)	1987	1995
NEOPHYTOU et MAR-MOLINERO (2004)	1988	1999
NORTON et SMITH (1979)	1971	1975
ODOM et SHARDA (1990)	1975	1982
OHLSON (1980)	1970	1976
PACEY et PHAM (1990)	1958	1985
PANTALONE et PLATT (1987)	1981	1984
PEEL et PEEL (1987)	1982	1985
PENDHARKAR (2005)	1987	1995
PHILOSOPHOV et PHILOSOPHOV (2002)	1980	1988
PINCHES et TRIESCHMANN (1974)	1966	1971
PIRAMUTHU et al. (1998)	1987	1989
PLATT et al. (1994)	1982	1988
PLATT et PLATT (1990)	1972	1986
PLATT et PLATT (2002)	1989	1998
POMPE et BILDERBEEK (2005 A)	1986	1994
POMPE et BILDERBEEK (2005 B)	1986	1993
POMPE et FEELDERS (1997)	1988	1994

	1990	1990
	1991	1991
DOUMPOS et ZOPOUNIDIS (1999)	1986	1993
DREZNER et al. (2001)	?	?
EDMISTER (1972)	1958	1969
	1969	1969
EL HENNAWY et MORRIS (1983)	1960	1971
ELMER et BOROWSKI (1988)	1986	1987
ETHERIDGE et SRIRAM (1997)	1989	1989
FAN et PALANISWAMI (2000)	?	?
FANNING et COGGER (1994)	1947	1965
FITZPATRICK (1932)	1920	1929
FLAGG et al. (1991)	1981	1985
FOREMAN (2003)	2000	2001
FRYDMAN et al. (1985)	1971	1981
GENTRY et al. (1985)	1970	1981
GENTRY et al. (1987)	1970	1981
GENTRY et al. (2002)	1971	1987
GILBERT et al. (1990)	1975	1983
GODLEWSKI (2003)	1997	1982
GOMBOLA et al. (1987)	1970	1982
GOSS et RAMCHANDANI (1995)	1975	1985
GRICE et DUGAN (2003)	1985	1991
GRICE et INGRAM (2001)	1985	1991
GUPTA et al. (1990)	?	?
HÄRDLE et al. (2005)	2001	2002
HER et CHOE (1999)	1992	1998
	1992	1998
HUA et al. (2007)	1999	2004
HUANG et al. (1994)	1989	1991
HUNTER et ISACHENKOVA (2001)	1996	1997
	1990	1991
HWANG et al. (1997)	1985	1988
JO et al. (1997)	1991	1993
JONES et HENSHER (2004)	1996	2000
KANE et al. (1998)	1968	1990
KARELS et PRAKASH (1987)	1972	1972
KEASEY et MCGUINNESS (1990)	1976	1984
KEASEY et WATSON (1987)	1970	1983
KIM et al. (2003)	1991	2000
KIM et HAN (2003)	2001	2002

RAHIMIAN et al. (1993)	1975	1982
RAHMAN et al. (2004)	1998	1998
RAMSER et FOSTER (1931)	1919	1925
RICHARDSON et al. (1998)	1968	1990
SALCEDO-SANZ et al. (2004)	1983	1994
SALCEDO-SANZ et al. (2005)	1983	1994
SALCHENBERGER et al. (1992)	1986	1987
SARKAR et SRIRAM (2001)	1986	1988
SEN et al. (2004)	1968	1990
SERRANO-CINCA (1997)	1977	1985
SEXTON et al. (2003)	1989	1989
SHARMA et ISELIN (2003)	1970	1976
SHARMA et MAHAJAN (1980)	1970	1976
SHIN et al. (2005)	1996	1999
SHIN et LEE (2003)	1995	1997
SHIRATA (1998)	1986	1996
SHUMWAY (2001)	1962	1992
SINKEY (1975)	1973	1986
SKOGSVIK (1990)	1966	1980
SUN et al. (2005)	1997	2000
SUNG et al. (1999)	1991	1995
	1997	1998
TAFFLER (1983)	1969	1976
TAM et KIANG (1990)	1985	1987
TAM et KIANG (1992)	1985	1987
TAN et DIHARDJO (2001)	1989	1991
TANG et CHI (2005)	1995	2002
THEODOSSIOU (1991)	1980	1984
THEODOSSIOU (1993)	1973	1986
TIRAPAT (1999)	1997	1997
TSAKONAS (2006)	1986	1993
TSENG (2005)	1985	1994
TUNG et al. (2004)	1989	2000
TYREE et LONG (1996)	?	?
VAN GESTEL et al. (2006)	1991	1997
VARETTO (1998)	1982	1995
VIEIRA et al. (2003)	1998	1999
WALLRAFEN et al. (1996)		
WANG (2004)	1999	2001
WEISS et CAPKUN (2005)	2003	2003
	2002	2002

KIM et MCLEOD (1999)	1981	1985	2001	2001
KIVILUOTO (1998)	?	?	2000	2000
KOLARI et al. (2002)	1989	1992	1999	1999
KOTSIANTIS et al. (2005)	2003	2004	1998	1998
KUMAR et al. (1997)	?	?	WEST et al. (2005)	? ?
KURUPPU et al. (2003)	1987	1993	WESTGAARD et VAN DER WIJST (2001)	1996 1998
LACHER et al. (1995)	1972	1989	WILCOX (1973)	1954 1971
LAITINEN (1991)	?	?	WILSON et al. (1995)	1972 1979
LAITINEN et KANKAANPAA (1999)	1986	1989	WILSON et SHARDA (1994)	1975 1982
LAITINEN et LAITINEN (1995)	?	?	WINAKOR et SMITH (1935)	1923 1931
LAITINEN et LAITINEN (1998)	1986	1991	WU et al. (2007)	1998 2000
LAITINEN et LAITINEN (2000)	1985	1993	YANG et al. (1999)	1984 1989
LANINE et VANDER VENNET (2005)	1997	2004	YANG et HARRISON (2002)	1989 1995
	1997	2004	YIM et MITCHELL (2002)	1995 1999
	1997	2004	YIM et MITCHELL (2005)	1999 2000
	1997	2004	ZAVGREN (1985)	1972 1978
LAU (1987)	1976	1977	ZHANG et al. (1999)	1980 1991
LEE (2004)	1998	2000	ZMIJEWSKI (1984)	1972 1978
LEE et al. (1996)	1979	1992	ZOPOUNIDIS et DOUMPOS (2002)	1986 1993
LEE et al. (2005)	1995	1998	ZURADA et al. (1998)	1987 1988

Le « ? » signifie qu'aucune indication n'a été donnée au sujet des dates de prélèvement des comptes.

On constate, à la vue de ce tableau, que les périodes agrégées varient en moyenne entre 2 et 10 ans, avec quelques extrêmes comme Altman (1973), dont les données s'étalent sur plus de 30 ans. C'est bien pour pallier la rareté des données que les périodes sont aussi longues, rareté que l'on peut constater dans les tableaux 19 et 20 ci-après, qui précisent la taille des échantillons de chaque étude. Seulement 28 expérimentations sur les 190 s'appuient sur des échantillons de plus de 1 000 entreprises, et lorsque ceux-ci ont une taille inférieure à 1 000, leur taille moyenne est de 215 firmes.

Tableau 19 : Taille des échantillons employés par les études opposant des entreprises non défaillantes à des entreprises défaillantes

Études	Nombre d'entreprises par groupe		Études	Nombre d'entreprises par groupe	
	Non déf.	Déf.		Non déf.	Déf.
AGARWAL (1999)	70	70	LENSBERG et al. (2006)	568	568
ALTMAN (1968)	33	33	LESHNO et SPECTOR (1996)	44	44
ALTMAN (1973)	21	21	LI et GUPTA (2002)	142	58
ALTMAN et al. (1977)	53	58	LIN et PIESSE (2004)	45	32
ALTMAN et al. (1995)	61	34	LINDSAY et CAMPBELL (1996)	250	250
ANANDARAJAN et al. (2001)	104	418	LUSSIER (1995)	108	108
ATIYA (2001)	716	444	MCGURR et de VANEY (1998)	56	56

AZIZ et al. (1988)	39	39
BACK et al. (1994)	38	38
BACK et al. (1996 A)	37	37
BACK et al. (1996 B)	37	37
BACK et al. (1997)	570	285
BAEK et CHO (2003)	4 102	129
BARNIV et HERSHBARGER (1990)	69	69
	77	28
BARNIV et McDONALD (1992)	153	141
BARNIV et RAVEH (1989)	142	58
BEAVER (1966)	79	79
BEAVER (1968)	79	79
BELL et al. (1990)	1 834	233
BERG (2005)	100 000	1 000
BEYNON et PEEL (2001)	45	45
BLUM (1974)	115	115
BOOTH (1983)	35	35
BORITZ et KENNEDY (1995)	6 153	171
BRABAZON et KEENAN (2004)	89	89
BRABAZON et O'NEILL (2004)	89	89
BROCKETT et al. (1994)	183	60
CHARALAMBOUS et al. (2000)	139	139
CHARITOU et al. (2004)	51	51
CHAVA et JARROW (2004)	4 818	464
	16 263	1 197
CIELEN et al. (2004)	276	90
COATS et FANT (1993)	188	94
CYBINSKI (1995)	78	62
DAMBOLENA et KHOURY (1980)	23	23
DEAKIN (1972)	32	32
DIMITRAS et al. (1999)	40	40
DORSEY et al. (1995)	242	130
	370	127
	215	155
DOUMPOS et ZOPOUNIDIS (1999)	59	59
DREZNER et al. (2001)	185	155
EDMISTER (1972)	21	21
	281	281
EL HENNAWY et MORRIS (1983)	53	53
ELMER et BOROWSKI (1988)	3 206	60
ETHERIDGE et SRIRAM (1997)	991	148

MCKEE et GREENSTEIN (2000)	3 705	94
	2 935	91
	2 708	78
	3 019	72
	1 442	66
MCKEE et LENSBERG (2002)	145	145
MENSAH (1984)	110	110
MERWIN (1942)	381	558
MEYER et PIFER (1970)	30	30
MIN et al. (2006)	307	307
MIN et LEE (2004)	258	258
MIN et LEE (2005)	944	844
MOSSMAN et al. (1998)	37	37
	45	45
MOYER (1977)	25	23
NAM et JINN (2000)	46	46
NANDA et PENDHARKAR (2001)	85	85
NEOPHYTOU et MAR-MOLINERO (2004)	50	50
NORTON et SMITH (1979)	30	30
ODOM et SHARDA (1990)	64	65
OHLSON (1980)	2 058	105
PACEY et PHAM (1990)	285	74
PANTALONE et PLATT (1987)	29	15
PENDHARKAR (2005)	100	100
PHILOSOPHOV et PHILOSOPHOV (2002)		24
PINCHES et TRIESCHMANN (1974)	26	26
PIRAMUTHU et al. (1998)	91	91
PLATT et al. (1994)	89	35
PLATT et PLATT (1990)	57	57
POMPE et BILDERBEEK (2005 A)	3 000	1 369
POMPE et BILDERBEEK (2005 B)	3 000	1 356
POMPE et FEELDERS (1997)	288	288
RAHIMIAN et al. (1993)	64	65
RAMSER et FOSTER (1931)		178
RICHARDSON et al. (1998)	2 000	128
SALCEDO-SANZ et al. (2004)	36	36
SALCEDO-SANZ et al. (2005)	36	36
SALCHENBERGER et al. (1992)	158	158
SARKAR et SRIRAM (2001)	991	148
SEN et al. (2004)	160	129
SERRANO-CINCA (1997)	37	29

FAN et PALANISWAMI (2000)	88	88
FANNING et COGGER (1994)	190	190
FITZPATRICK (1932)	19	19
FLAGG et al. (1991)	177	25
FOREMAN (2003)	63	14
FRYDMAN et al. (1985)	142	58
GENTRY et al. (1985)	33	33
GENTRY et al. (1987)	33	33
GENTRY et al. (2002)	99	99
GODLEWSKI (2003)	1 727	120
GOMBOLA et al. (1987)	244	77
GRICE et DUGAN (2003)	3 523	671
GRICE et INGRAM (2001)	1 678	296
GUPTA et al. (1990)	30	30
HÄRDLE et al. (2005)	42	42
HER et CHOE (1999)	35	35
	22	22
HUA et al. (2007)	60	60
HUANG et al. (1994)	2 000	44
HUNTER et ISACHENKOVA (2001)	27	27
	20	20
HWANG et al. (1997)	2 020	658
JO et al. (1997)	271	271
KANE et al. (1998)	2 000	128
KARELS et PRAKASH (1987)	350	25
KEASEY et MCGUINNESS (1990)	43	43
KEASEY et WATSON (1987)	73	73
KIM et al. (2003)	331	331
KIM et HAN (2003)	386	386
KIM et MCLEOD (1999)	30	30
KIVILUOTO (1998)	833	304
KOLARI et al. (2002)	50	50
KOTSIANTIS et al. (2005)	100	50
LACHER et al. (1995)	188	94
LAITINEN (1991)	40	40
LAITINEN et KANKAANPAA (1999)	38	38
LAITINEN et LAITINEN (1995)	40	40
LAITINEN et LAITINEN (1998)	41	41
LAITINEN et LAITINEN (2000)	285	285
LANINE et VANDER VENNET (2005)	390	78
	325	85

SEXTON et al. (2003)	991	148
SHARMA et ISELIN (2003)	25	25
SHARMA et MAHAJAN (1980)	23	23
SHIN et al. (2005)	1 160	1 160
SHIN et LEE (2003)	264	264
SHIRATA (1998)	300	686
SHUMWAY (2001)	3 182	300
SKOGSVIK (1990)	328	51
SUNG et al. (1999)	49	21
	54	30
TAFFLER (1983)	46	46
TAM et KIANG (1990)	81	81
TAM et KIANG (1992)	81	81
TAN et DIHARDJO (2001)	2 078	66
TANG et CHI (2005)	128	128
THEODOSSIOU (1991)	420	81
THEODOSSIOU (1993)	197	62
TIRAPAT (1999)	341	55
TSAKONAS (2006)	59	59
TSENG (2005)	45	32
TUNG et al. (2004)	2 555	548
TYREE et LONG (1996)	91	55
	22	34
VAN GESTEL et al. (2006)	348	74
VARETTO (1998)	2 369	2 369
VIEIRA et al. (2003)	225	225
WALLRAFEN et al. (1996)	?	?
WANG (2004)	46	1 946
WEISS et CAPKUN (2005)	2 595	41
	2 643	59
	2 568	86
	2 455	74
	2 082	31
	1 729	36
WEST et al. (2005)	236	93
WESTGAARD et VAN DER WIJST (2001)	69 620	1 908
WILCOX (1973)	52	52
WILSON et SHARDA (1994)	64	65
WINAKOR et SMITH (1935)		183
WU et al. (2007)	66	22
YANG et al. (1999)	89	33

	490	98
	445	89
LEE (2004)	70	35
LEE et al. (1996)	83	83
LEE et al. (2005)	84	84
LEE et URRUTIA (1996)	76	76
LENNOX (1999)	949	90

YANG et HARRISON (2002)	2 244	164
YIM et MITCHELL (2002)	80	20
YIM et MITCHELL (2005)	92	29
ZAVGREN (1985)	45	45
ZHANG et al. (1999)	110	110
ZMIJEWSKI (1984)	1 600	81
ZOPOUNIDIS et DOUMPOS (2002)	59	59

Tableau 20 : Taille des échantillons employés dans les études opposant plus de deux catégories d'entreprises

Études	Groupes	Nbre
AGARWAL et al. (2001)	• Non défaillance	279
	• Réduction de dividendes	29
	• Défaut de paiement	37
	• Faillite	30
ALAM et al. (2000)	• Sur-performance	17
	• Non défaillance	80
	• Faillite	3
ALTMAN (1994)	• Non défaillance	404
	• Vulnérabilité	404
	• Défaut de paiement	404
	• Non défaillance	151
	• Vulnérabilité	151
	• Défaut de paiement	151
ANANDARAJAN et al. (2001)	• Défaillance	418
	• Faillite	104
BARANOFF et al. (2000)	• Non défaillance	1 522
	• Détresse	60
BARNIV et al. (1997)	• Acquisition	49
	• Restructuration	119
	• Liquidation	69
BARNIV et al. (2002)	• Acquisition	49
	• Restructuration	119
	• Liquidation	69
BOSE et PAL (2006)	• Non défaillance	120
	• Mauvaise santé financière	120
GILBERT et al. (1990)	• Non défaillance	304
	• Défaillance	304
	• Faillite	76
GOSS et RAMCHANDANI (1995)	• Solvabilité	20
	• Insolvabilité	20
JONES et HENSHER (2004)	• Non défaillance	2 838
	• Insolvabilité	78
	• Faillite	116

Études	Groupes	Nbre
KUMAR et al. (1997)	• Liquidation	76
	• Réorganisation	76
KURUPPU et al. (2003)	• Mauvaise santé	50
	• Liquidation	85
LAU (1987)	• Non défaillance	700
	• Réduction de dividendes	40
	• Défaut de paiement	30
	• Mise sous protection des lois sur la faillite	20
	• Faillite	10
PEEL et PEEL (1987)	• Non défaillance et profits	56
	• Non défaillance et pertes	34
	• Défaillance	56
PLATT et PLATT (2002)	• Non défaillance	62
	• Détresse	24
RAHMAN et al. (2004)	• Non défaillance	76
	• Problématique	52
SINKEY (1975)	• Non problématique	110
	• Problématique	110
SUN et al. (2005)	• Mauvaise santé	297
	• Faillite	58
WILSON et al. (1995)	• Survie	40
	• Problématique	32
	• Défaillance	40
ZURADA et al. (1998)	• Non défaillance	253
	• Réduction de dividendes	28
	• Défaut de paiement	34
	• Faillite	30

On comprend bien l'intérêt pratique d'agréger des données issues de différentes périodes pour pallier la rareté de celles décrivant des défailtantes. Mais quelles incidences ceci peut-il avoir sur les modèles ? Mensah (1984) a comparé un modèle construit à partir de données qui s'étalent largement dans le temps, avec d'autres construits sur des intervalles de temps restreints, où régnaient à chaque fois des conditions économiques homogènes. Il a ainsi pris en compte

quatre périodes : 1972-1973, qui est une période de croissance, 1973-1975, qui est une période de récession, 1976-1977, qui est aussi une période de croissance, et 1978-1980, qui est une période de stagflation. Il a constaté que les modèles calculés au cours d'une phase de croissance effectuent des prévisions bien meilleures lorsqu'ils sont appliqués sur des données mesurées lors de phases identiques que lorsqu'ils sont appliqués sur des données issues de périodes de récession, et inversement. Platt et Platt (1990) se sont intéressés, comme dans le cas précédent, aux échantillons prélevés sur de longues périodes mais aussi composés d'entreprises provenant de divers secteurs. Ils ont étudié la façon de stabiliser les ratios pour se prémunir des effets conjugués des différences temporelles et intersectorielles, c'est-à-dire éviter que leur moyenne ne change entre la période d'estimation du modèle et la période de prévision. Ils ont suggéré de pondérer les ratios par la moyenne du secteur d'où provient l'entreprise et ont pu montrer qu'une telle méthode de normalisation donnait des résultats plus précis que si elle n'était pas employée. Platt et al. (1994) ont réexaminé la problématique de Platt et Platt (1990) mais cette fois en proposant une pondération spécifique à chaque ratio, et en utilisant comme facteur de pondération, les taux d'intérêt et le prix du baril de pétrole. Là aussi, les pondérations conduisent à de meilleurs résultats que si elles ne sont pas employées. Skogsvik (1990) s'est intéressé à l'influence de l'inflation sur les ratios, en comparant des modèles calculés sur des données brutes avec d'autres élaborés avec des données ajustées, sans que les résultats ne mettent en valeur de différences, la hausse des prix jouant donc un rôle négligeable. Pompe et Bilderbeek (2005 B) ont analysé la performance d'un modèle conçu lors d'une période économique plutôt favorable aux entreprises, mais avec des données prélevées au cours d'une récession, c'est-à-dire à un moment où le taux de défaillance est bien supérieur à la normale. Ils ont constaté, en étudiant les résultats obtenus sur plusieurs années, que la détérioration des prévisions coïncide précisément avec l'apparition d'une période de récession. Tout ceci montre bien le côté chaotique des modèles, leur extrême sensibilité aux conditions initiales de calcul et d'utilisation, et la difficulté de maîtriser en même temps les sources de biais et de variance. Il est en effet difficile à la fois de construire des modèles « proches » du véritable modèle, ce qui implique un biais non négligeable et des modèles qui ne sont pas « *sample specific* », comme disent souvent les anglo-saxons, donc qui sont relativement indépendants des échantillons dont ils dérivent, ce qui produit de la variance. Ceci suggère aussi que la période au cours de laquelle les données doivent être sélectionnées, pour le paramétrage des modèles, soit la plus courte possible et que cette même période soit, d'un point de vue macro-économique, sensiblement identique à la période de test et d'exploitation.

1.4.3 HISTORIQUE ET HORIZON PRÉVISIONNEL

On vient de le voir, le facteur temps joue un rôle considérable et ce, d'autant plus que la période de référence pour mesurer des variables financières est assez proche de la date officielle de la défaillance, puisqu'elle se situe en moyenne un an avant. On peut alors penser que si les règles de prévision sont instables dans le temps, leur horizon d'utilisation risque d'être fortement limité. Ceci a été évoqué dans la partie consacrée à la définition de la défaillance. Nous allons le préciser ici.

Beaucoup d'auteurs se sont préoccupés du délai au-delà duquel il n'est plus envisageable de prévoir ce qu'il adviendra d'une entreprise et en filigrane, se sont aussi posés la question de l'étendue de l'historique des données à utiliser ; jusqu'où peut-on alors aller dans la prévision et jusqu'où peut-on remonter dans le temps pour trouver des signes avant-coureurs de la défaillance ?

Les variables explicatives, le plus souvent de nature comptable, sont calculées à partir d'informations annuelles, car ce rythme correspond à celui imposé par la loi pour l'arrêté et la publication des comptes. Tout autre intervalle est difficilement envisageable pour réaliser des mesures car aucune source ne permet de recueillir aisément des données « intermédiaires ».

Plusieurs stratégies de calcul des variables explicatives ont été envisagées au cours du temps, mais il y en a une qui domine largement les autres. Elle consiste à repérer une date, qui fixe l'horizon de la prévision, et à prélever les données un certain laps de temps avant : moins d'un an, un an, deux ans avant... Ainsi, pour une date t , par exemple, on décide de ne retenir que les données de $t-1$, et de $t-1$ uniquement, sans tenir compte des années précédentes. De rares auteurs ont cependant observé qu'une telle pratique conduisait à se couper de l'historique des comptes d'une entreprise ($t-2, t-3 \dots$) qui pourrait renfermer plus d'informations que ne le pourrait une seule série. Faisant l'hypothèse que la défaillance s'inscrit dans un processus dont on peut trouver des traces plusieurs années à l'avance, quelques-uns ont suggéré d'utiliser toutes ces informations. Shumway (2001) propose ainsi un modèle de survie fondé sur un historique de dix ans et l'oppose à plusieurs modèles qui, eux, ne reposent que sur une seule série de données annuelles, et montre que les résultats obtenus avec le premier sont nettement plus précis qu'avec les autres. D'autres proposent de coder l'évolution de chaque variable financière dans autant de variables statistiques qu'il y a d'années à prendre en compte. Pedersen (1997) utilise ainsi un historique qui s'étale sur deux ans. D'autres encore, comme Gentry et al. (1985), emploient des moyennes calculées sur plusieurs années. Mais ces pratiques

sont marginales, surtout celle envisagée par Shumway (1991), car dans ce dernier cas, la structure des méthodes statistiques n'autorisent pas ce qui peut être fait avec un modèle de survie.

Maintenant, s'agissant des horizons de prévision, les délais couramment étudiés varient entre un et trois ans, parfois entre un et cinq ans, mais jamais au-delà de sept ans. Les modèles sont construits et testés de deux manières différentes. La première revient à calculer un modèle sur un échantillon d'apprentissage contenant des variables mesurées un an avant la défaillance, puis à le tester sur un autre échantillon, mais cette fois avec des données mesurées un, deux, trois... ans avant. La seconde consiste aussi à calculer et à tester les modèles sur des échantillons différents, mais cette fois en les spécialisant par année antérieure à la défaillance : on ne cherche plus à tester un même modèle sur des données appartenant à des périodes différentes (modèle calculé en t pour prévoir une défaillance en $t+1$ et ce, avec des données relevées en $t, t-1, t-2 \dots$), mais à construire un modèle par période précédent la défaillance (modèles calculés en $t-1, t-2$ pour prévoir une défaillance en t). Dans un cas comme dans l'autre, plus l'écart se creuse entre la date de l'événement à prévoir et la mesure des variables, plus les prévisions sont mauvaises. À titre d'illustration, nous donnons dans le tableau 21 ci-après les résultats d'un certain nombre de travaux relatifs à la première catégorie décrite ci-dessus.

Les résultats présentés correspondent tous à des taux calculés sur des échantillons de test (*out of sample*), et montrent assez clairement une dégradation nette des performances dès lors que le délai de prévision dépasse un an, qui est de l'ordre de 8 % en moyenne. Et ces taux, chutent de manière considérable à partir de trois ans (nous ne les avons pas indiqués par souci de clarté).

Bien que certains auteurs aient pu montrer la présence de signes avant coureur de la défaillance bien avant que celle-ci ne se produise¹⁶, le délai moyen qui favorise une bonne prévision est réduit, et est compris entre 12 et 18 mois. Ce délai est en fait le pendant du laps de temps couramment retenu entre la date légale de défaillance et la date de calcul des modèles ; un bon modèle est donc un modèle qui, à un instant t , prévoit bien dans un intervalle compris entre t et $t+1$, dès lors qu'il a été calculé en $t-1$.

¹⁶ Nam et Jinn (2000), par exemple, ont analysé la structure financière d'entreprises coréennes défailtantes lors de la crise économique de 1998-1999, au cours des années qui ont précédé leur faillite et ont montré que la plupart d'entre elles, bien avant la survenue de la crise, présentaient déjà des signes avancés de mauvaise santé financière, ce qui leur a permis de construire un modèle capable de prévoir correctement, six avant la défaillance des firmes, 77,8 % des cas. Il s'agit bien évidemment d'une situation exceptionnelle, aucun autre modèle n'ayant jamais pu présenter une telle précision aussi longtemps à l'avance.

Tableau 21 : Résultats des prévisions effectuées un et deux ans avant la défaillance

Études	% de prévisions correctes		Études	% de prévisions correctes	
	1 an avant	2 ans avant		1 an avant	2 ans avant
ALTMAN (1968)	95,00%	72,00%	DIMITRAS et al. (1999)	76,30%	60,50%
ATIYA (2001)	87,04%	57,14%	DOUMPOS et ZOPOUNIDIS (1999)	71,05%	60,53%
AZIZ et al. (1988)	91,80%	84,70%	KOTSIANTIS et al. (2005)	71,17%	70,14%
BACK et al. (1996)	97,00%	77,00%	LACHER et al. (1995)	94,70%	89,40%
	84,00%	72,00%	LACHER et al. (1995)	94,70%	89,40%
BARNIV et HERSHBARGER (1990)	91,10%	85,70%	LAITINEN et KANKAANPAA (1999)	94,70%	82,90%
BARNIV et McDONALD (1992)	86,60%	85,70%	LAU (1987)	80,00%	79,00%
	82,90%	80,00%	NAM et JINN (2000)	84,40%	76,10%
BRABAZON et KEENAN (2004)	80,67%	72,00%	PIRAMUTHU et al. (1998)	89,10%	87,00%
BRABAZON et O'NEILL (2004)	78,67%	74,67%	TAM et KIANG (1990)	85,20%	88,70%
CHARITOU et al. (2004)	83,33%	76,19%	YIM et MITCHELL (2002)	92,00%	90,00%
COATS et FANT (1993)	92,90%	86,20%	ZURADA et al. (1998)	81,58%	73,05%
DEAKIN (1972)	91,20%	84,80%			

Nous n'avons pas indiqué dans ce tableau les études qui présentaient de multiples résultats rendant impossible tout choix. Et lorsqu'il a fallu choisir, nous avons retenu systématiquement les meilleurs, donc ceux qui correspondent aux taux de classement les plus élevés.

Si le moment où les variables sont mesurées est trop proche de la date légale, des distorsions peuvent apparaître. Elles peuvent être dues aux stratégies de présentation des comptes selon l'issue judiciaire envisageable ou envisagée par les dirigeants ; en cas de rachat possible de l'entreprise, par exemple, ou dans un optique de continuation, une équipe dirigeante peut être amenée à présenter plus favorablement une structure financière qu'elle ne le ferait en cas de liquidation. Mais elles peuvent aussi être engendrées par des stratégies post-défaillance mises en place précisément pour éviter une issue fatale. C'est le cas notamment des entreprises qui, sentant le moment fatidique arriver, vont tenter de rester à flot en jouant sur leur niveau de trésorerie, et qui vont finalement défaillir tout en présentant une excellente structure de liquidité qui n'est absolument pas représentative de leur réel état financier. Un délai d'un an permet donc de se prémunir des effets de ce type de stratégies qui visent à remodeler un profil financier ou à tout le moins, d'atténuer leurs effets dans l'hypothèse où elles peuvent être enclenchées plus d'un an avant le moment ultime de la vie d'une firme.

Dans le cas inverse, c'est-à-dire si la mesure intervient trop en amont de la date de défaillance, donc à partir de deux ans avant celle-ci, les modèles perdent leur capacité de discrimination. Tout se passe comme si, dépassé ce seuil, la population des firmes à risque n'était plus homogène, une partie ayant tendance à se fondre dans la catégorie des entreprises saines, certes au profil

présentant des déficiences financières, mais pas suffisamment pour en être radicalement différente. La frontière est alors trop ténue pour autoriser toute discrimination.

En définitive, deux arguments plaident en faveur de ce délai moyen compris entre un an et un an et demi : d'un côté, c'est le moment où les firmes défailtantes présentent le moins de similarités avec les autres et de l'autre, c'est un laps de temps que l'on peut considérer comme étant suffisamment court pour que des fluctuations macro-économiques n'affectent pas sensiblement les modèles. On retrouve cette constante quelle que soit la méthode statistique employée, le type de variables utilisées, ou les caractéristiques des échantillons : taille, secteur...

1.4.4 CRITÈRE DE PERFORMANCE ET CRITÈRE DE CLASSIFICATION

Il reste un dernier sujet important qui fait débat et qui renvoie au critère de décision qu'un modèle devra employer pour effectuer ses prévisions ainsi qu'au critère d'évaluation de ses performances. Comme nous sommes en présence d'une problématique de classification, la précision d'une règle se calcule à partir d'une matrice de confusion indiquant la répartition des classements prévisionnels par groupe. Une fois la répartition effectuée, il faut choisir une mesure qui permettra de qualifier la qualité du modèle, qui s'appuie la plupart du temps sur un critère de *Rand*, rapportant le nombre d'observations mal classées de chaque groupe au total général des observations de l'échantillon. Cette mesure brute peut être avantageusement complétée par une mesure de complexité du modèle, pondérant l'estimation de l'erreur par un calcul faisant intervenir le nombre de variables, si d'aventure on souhaite comparer plusieurs modèles aux performances analogues entre eux. En effet, lorsque l'on réalise une modélisation statistique dans une optique prédictive, on cherche en définitive un modèle qui dispose d'une bonne capacité de généralisation donc qui, une fois confronté à des données qui n'ont pas servi à sa construction, est capable de donner de bons résultats. En ce sens, un « bon » modèle est avant tout un modèle « parcimonieux », c'est-à-dire un modèle disposant d'un nombre réduit de paramètres ajustables. Donc, à performance égale, il vaut mieux choisir entre deux modèles, celui qui est le moins complexe. C'est pourquoi, si on désire les comparer, il faut tenir compte à la fois de leur performance et de leur complexité, et pondérer la première par la seconde.

Maintenant, s'agissant du critère de classification, son choix n'est pas donné de prime abord. En effet, l'affectation d'une observation à un groupe suppose que l'on ait déterminé *a priori* un seuil (*cut*), délimitant d'une certaine manière la frontière entre les groupes, qui peut varier

selon la stratégie de calcul ; le modèle produit pour chaque observation un « score », qui une fois comparé au *cut*, déterminera sa classe d'appartenance.

Si l'on dispose d'un modèle continu, la règle de classification peut être formulée ainsi, avec y_i^* correspondant à la classe prévue de l'entreprise i et y^* à la valeur du *cut* :

$$y_i^* \begin{cases} 1 & \text{si le score } \hat{y}_i \text{ de l'entreprise } i \text{ produit par le modèle } > y^* \\ 0 & \text{si le score } \hat{y}_i \text{ de l'entreprise } i \text{ produit par le modèle } \leq y^* \end{cases}$$

Il est possible d'envisager plusieurs méthodes pour calculer le *cut*. La plus simple, et aussi la plus répandue car la plus économique, consiste à prendre la valeur qui minimise le taux d'erreur global. Mais, lorsque les groupes ont des effectifs très différents, cette stratégie est susceptible de conduire à des taux très élevés, tout en produisant des résultats médiocres pour la classe sous représentée. Le tableau 16, p. 56, en est une parfaite illustration. Il indique bien comment la sur représentation d'une classe conduit à une asymétrie très forte des taux de prévision correcte selon le groupe. Par ailleurs, cette mesure du *cut* ne tient absolument pas compte de l'asymétrie du coût d'une erreur, qui a été souvent relevée dans la littérature (Edmister, 1972 ; Altman et al., 1977 ; Eisenbeis, 1977 ; Pacey et Pham, 1990 ; Salchenberger et al. 1992 ; Tam et Kiang, 1992 ; Hsieh, 1993 ; Huang et al., 1994 ; Dorsey et al., 1995 ; Barniv et al., 1997 ; Etheridge et Sriram, 1997 ; Pompe et Feelders, 1997 ; Kiviluoto, 1998 ; Agarwal et al., 2001 ; Nanda et Pendharkar, 2001 ; Kim et al. 2003 ; Tang et Chi, 2005) et que nous avons abordée au cours du paragraphe 1.4.1, p. 55. C'est pourquoi de nombreux auteurs recommandent d'en tenir compte lors de la détermination de la valeur optimale du *cut*, en l'incorporant explicitement dans le calcul de la fonction de classement. Ceci revient, à paramètres constants, à tenter de diminuer le taux de mauvais classements des entreprises défailtantes tout en maximisant l'écart entre le taux de mauvais classements des deux groupes, avec le taux des défailtantes tendant vers zéro.

La fonction de coût suivante à minimiser, peut être utilisée pour déterminer le *cut* et calculer le coût attendu (CA) des erreurs d'un modèle :

$$CA = \frac{p_1 \frac{e_{12}}{n_1} c_1}{p_2 \frac{e_{21}}{n_2} c_2}$$

avec p_1 et p_2 correspondant respectivement aux probabilités *a priori* de défaillance et de non défaillance, e_{12} et e_{21} aux erreurs respectives de type 1 (entreprise défailtante considérée comme

non défailante) et de type 2 (entreprise non défailante considérée comme défailante), n_1 et n_2 aux nombres d'entreprises défailtantes et non défailtantes dans l'échantillon et c_1 et c_2 aux coûts respectifs d'une erreur de type 1 et de type 2.

Toute la difficulté ici, réside dans la fixation des valeurs que prendront c_1 et c_2 , p_1 et p_2 étant assez facilement ajustables en fonction du taux moyen de défaillance observé au moment où le modèle est élaboré. En effet, c_1 et c_2 , comme le note Pacey et Pham (1990), diffèrent selon la firme considérée, mais aussi selon la situation de celui qui emploie le modèle. Une banque ou un investisseur, par exemple, évaluera au cas par cas un risque de non remboursement d'une créance par rapport à un risque de perte d'une affaire, le risque de non remboursement devant aussi être pondéré par la capacité de chacun à recouvrir une dette.

L'aspect conjoncturel du rapport pouvant exister entre c_1 et c_2 a conduit beaucoup d'auteurs à ne pas employer cette méthode, donc à supposer une égalité des coûts.

Une autre forme de calcul du *cut* a été proposée et employée notamment par Kim et al. (2003). Elle repose sur un test statistique (*Kolmogorov-Smirnov*) fondé sur la distribution des scores obtenus par les observations des deux groupes, et permet de s'abstraire de la prise en compte de leur proportion réelle comme du coût d'une erreur. La valeur du *cut* est celle pour laquelle les deux distributions des scores sont le plus significativement différentes, au sens du test utilisé.

On voit bien qu'il n'y a pas de valeur absolue définissant le *cut* et qu'il est souhaitable d'étudier plusieurs solutions. C'est aussi pour cette raison que certains auteurs, comme Ooghe et al. (1999), ont envisagé un autre critère de performance que ceux que nous avons évoqués précédemment, et qui s'appuie sur cette idée. Ce critère repose sur une courbe de concentration (courbe de *Lorentz*) faisant apparaître la distribution des fréquences cumulées des erreurs de prévision de chaque groupe. On dispose alors d'une représentation graphique avec, par exemple, en abscisse, l'erreur de type 1, et en ordonnée, l'erreur de type 2, la courbe indiquant toutes les valeurs possibles du *cut* lorsque chaque erreur varie de 0 à 100 %. Plus la courbe est proche des axes, plus le modèle correspondant est précis, le meilleur modèle étant celui représenté par les axes eux-mêmes. L'espace qui existe entre la courbe et les axes est formalisable au travers d'un indice (indice de *Gini*) que l'on peut utiliser pour caractériser la précision générale d'un modèle et surtout la comparer à d'autres.

Dans la pratique, les classements ainsi que les évaluations de performance, y compris lorsque ces dernières le sont dans une optique comparative, se font quasi systématiquement sur la base

d'un taux global de bons classements, en dehors de toute considération relative au coût des erreurs ou à la complexité des modèles. L'absence de référence au coût peut se justifier, si l'on considère que le nombre de combinaisons des différents paramètres complique singulièrement la tâche puisqu'il conviendrait de tester, à l'image d'Altman (1977), au moins deux probabilités *a priori* de défaillance, en les combinant avec 8 ou 10 valeurs associées au coût relatif de chaque erreur (type 1 et type 2), et ce pour tous les modèles étudiés. En revanche, le manque de prise en compte de la complexité est plus problématique, surtout lorsque les études proposent des comparatifs de modèles comprenant des nombres très différents de variables et/ou de paramètres ajustables. Ce qui est souvent le cas avec les réseaux de neurones, où la prise en considération du nombre de neurones cachés n'intervient jamais dans la pondération des performances lorsque les modèles présentent des performances similaires.

1.5 CONCLUSION DU CHAPITRE

Ce chapitre a permis de montrer, malgré les différentes formes que peut revêtir la défaillance, que c'est sa définition légale qui est la plus souvent employée. Parce qu'elle permet de dater une situation qui est par nature critique, elle facilite la modélisation : on dispose bien d'un repère temporel pour mesurer un état et d'un niveau de gravité suffisant pour le discriminer franchement d'un état opposé. Ainsi, un modèle de défaillance peut être vu comme n'étant pas simplement une règle dichotomique, prédisant ou non la survenue d'une issue fatale à plus ou moins brève échéance, mais comme une règle calculant une mesure de « distance à la défaillance », renseignant ainsi sur la plus ou moins grande proximité entre l'état financier d'une entreprise et un état critique de référence.

Ce chapitre a aussi montré que la défaillance ne se laisse pas appréhender au travers d'une série limitée de causes, que celles-ci sont multiples et pas toutes prévisibles, et qu'elles ne se révèlent pas non plus systématiquement dans les états comptables. Malgré cette diversité, c'est bien l'information de nature financière qui est le plus souvent privilégiée, au travers de l'emploi de ratios calculés d'abord à partir de bilans et de comptes de résultat, pour des raisons doublement économiques : en premier lieu, économique au sens des moyens à mettre en œuvre pour obtenir une information, les ratios étant plus facilement accessibles que toute autre forme de données ; en second lieu, économique au sens du contenu informationnel des modèles obtenus, ceux fondés sur des ratios étant plus précis que les autres.

Ce chapitre, en complément, a mis en évidence les conditions de validité des modèles ou du moins, celles que l'on peut dégager des pratiques, car recueillant une sorte de consensus implicite sans être théoriquement démontrées. Deux séries de conditions émergent. La première plaide pour que les modèles ne soient pas généraux, mais plutôt spécialisés par catégories d'entreprises, ces dernières n'étant pas toutes égales devant de la défaillance selon quelles sont plus ou moins grandes, qu'elles interviennent dans tel ou tel secteur, ou qu'elles soient de création récente ou non. Il n'y a pas de véritable consensus sur la façon de mesurer chacun de ces critères et sur les subdivisions à leur appliquer. Il n'y a de consensus que sur le principe de la spécialisation et sur la primauté de la taille appréhendée par les actifs et du secteur comme paramètres délimitant leur domaine de validité. La seconde plaide pour des modèles temporels, suggérant de les calculer avec des données de la même période, plutôt qu'avec des données s'étalant sur plusieurs années, et ayant un horizon de prévision limité.

Enfin, ce chapitre a donné un aperçu des critères de classification habituellement employés et des modalités d'évaluation de la performance d'un modèle. Il a mis en évidence le caractère rudimentaire des pratiques, en raison de l'absence quasi systématique de prise en compte de l'asymétrie du coût d'une erreur et du manque de prise en considération de la complexité des modèles lors des études comparatives.

CHAPITRE 2 : CONSTRUCTION DE MODÈLES DE PRÉVISION À L'AIDE DE RÉSEAUX DE NEURONES

La première utilisation des réseaux de neurones dans le domaine de la prévision de défaillance, qui ait eu un certain retentissement, remonte aux travaux d'Odom et Sharda (1990). Bien qu'ayant obtenu des résultats très moyens (81 % de bons classements) et ce, sur un petit échantillon (129 entreprises), elle a ouvert la voie à tout un courant de recherche qui y a vu une technique prometteuse de par ses propriétés mathématiques, laissant supposer qu'elle résoudrait notamment la question liée à l'absence de prise en compte, par les méthodes habituelles, de la non linéarité des relations entre les ratios financiers.

Les réseaux de neurones ne constituent pas la seule famille de méthodes non linéaires, mais c'est néanmoins vers elle, et presque exclusivement, que tous les chercheurs se sont tournés.

Ce chapitre leur est consacré. Nous dresserons tout d'abord et rapidement, un panorama des études qui s'appuient sur les réseaux depuis maintenant plus de quinze ans, afin de dégager, parmi tous les types qui existent, celui ou ceux qui font office de techniques de référence. Nous expliquerons ensuite les arguments qui ont été avancés pour justifier le choix de leur emploi, et surtout les avantages qu'ils présentent par rapport aux autres méthodes. Nous étudierons aussi leur aspect algorithmique au travers des techniques d'optimisation mises en jeu pour la réalisation de modèles et les contraintes afférentes aux modèles issus d'un apprentissage. Enfin, nous indiquerons comment les différents auteurs auxquels nous nous sommes référés et qui ont construit des modèles de défaillance avec des réseaux, ont procédé pour y parvenir.

2.1 TECHNIQUES DOMINANTES

Les réseaux peuvent être classés de différentes manières : selon leur architecture (bouclé ou non), leur technique d'apprentissage (supervisée ou non) ou l'algorithme employé au cours de la procédure d'apprentissage. Afin d'éviter toute ambiguïté, nous utiliserons pour classer les réseaux, les termes habituellement retenus dans la littérature, chaque réseau étant alors caractérisé par un « nom » auquel on y associe souvent celui de l'algorithme intervenant dans le processus d'entraînement, algorithme assimilé la plupart du temps et de manière hâtive, à la procédure d'entraînement elle-même. Nous avons retenu cette façon de les nommer afin de pouvoir différencier, pour un même type de réseau, les différentes méthodes d'apprentissage employées dans la pratique.

Le tableau 22 indiqué ci-après dresse la liste des réseaux utilisés par toutes les études sur la prévision de la défaillance que nous avons recensées, et précise le nom du réseau ainsi que celui de l’algorithme caractérisant l’apprentissage.

Tableau 22 : Algorithmes neuronaux employés par les principales études sur la défaillance d’entreprise

Études	Algorithmes employés	Études	Algorithmes employés
AGARWAL (1999)	MLP-BP	TUNG et al. (2004)	MLP-BP GSOFNN
AGARWAL et al. (2001)	MLP-BP	FAN et PALANISWAMI (2000)	MLP-BP LVQ
ALTMAN et al. (1994)	MLP-BP	LESHNO et SPECTOR (1996)	MLP-BP MLP- ?
ATIYA (2001)	MLP-BP	ANANDARAJAN et al. (2001)	MLP-BP MLP-AG
BACK et al. (1996 B)	MLP-BP	PENDHARKAR (2005)	MLP-BP MLP-AG
BARNIV et al. (1997)	MLP-BP	BORITZ et KENNEDY (1995)	MLP-BP MLP-FLBP MLP-PBP MLP-PCBP
BOSE et PAL (2006)	MLP-BP	TYREE et LONG (1996)	MLP-BP PNN
BROCKETT et al. (1994)	MLP-BP	YANG et al. (1999)	MLP-BP PNN
GOSS et RAMCHANDANI (1995)	MLP-BP	YANG et HARRISON (2002)	MLP-BP PNN RHPNN
JO et al. (1997)	MLP-BP	BAEK et CHO (2003)	MLP-BP RAA-LM
KIM et HAN (2003)	MLP-BP	BACK et al. (1994)	MLP-BP SOM BM
KUMAR et al. (1997)	MLP-BP	LEE et al. (1996)	MLP-BP SOM-MLP
LAITINEN et KANKAANPAA (1999)	MLP-BP	BACK et al. (1997)	MLP-AG
LEE (2004)	MLP-BP	BRABAZON et KEENAN (2004)	MLP-AG
MIN et LEE (2004)	MLP-BP	DORSEY et al. (1995)	MLP-AG
MIN et LEE (2005)	MLP-BP	HUANG et al. (1994)	MLP-AG
ODOM et SHARDA (1990)	MLP-BP	KIM et MCLEOD (1999)	MLP-AG
PIRAMUTHU et al. (1998)	MLP-BP	KIM et al. (2003)	MLP-AG
POMPE et FEELDERS (1997)	MLP-BP	LI et GUPTA (2002)	MLP-AG
POMPE et BILDERBEEK (2005 A)	MLP-BP	SEXTON et al. (2003)	MLP-AG
POMPE et BILDERBEEK (2005 B)	MLP-BP	WALLRAFEN et al. (1996)	MLP-AG
SALCHENBERGER et al. (1992)	MLP-BP	CHARALAMBOUS et al. (2000)	MLP-GC
SEN et al. (2004)	MLP-BP	CHARITOU et al. (2004)	MLP-GC
SERRANO-CINCA (1997)	MLP-BP	ZHANG et al. (1999)	MLP-GRG2
SHIN et al. (2005)	MLP-BP	LEE et al. (2005)	MLP-LM SOM
TAM et KIANG (1990)	MLP-BP	BELL et al. (1990)	MLP-NCB
TAM et KIANG (1992)	MLP-BP	KOTSANTIS et al. (2005)	MLP-RBF

TAN et DIHARDJO (2001)	MLP-BP	BACK et al. (1996 A)	MLP- ?
TANG et CHI (2005)	MLP-BP	BERG (2005)	MLP- ?
WEST et al. (2005)	MLP-BP	COATS et FANT (1993)	CasCor
WILSON et SHARDA (1994)	MLP-BP	LACHER et al. (1995)	CasCor
WILSON et al. (1995)	MLP-BP	VIEIRA et al. (2003)	HLVQ G-Prop
WU et al. (2007)	MLP-BP	ALAM et al. (2000)	SOM CNN
YIM et MITCHELL (2002)	MLP-BP	ETHERIDGE et SRIRAM (1997)	SOM PNN
YIM et MITCHELL (2005)	MLP-BP	KIVILUOTO (1998)	SOM SOM-RBF
ZURADA et al. (1998)	MLP-BP	MCKEE et GREENSTEIN (2000)	?
RAHIMIAN et al. (1993)	MLP-BP ANN	MIN et al. (2006)	?
FANNING et COGGER (1994)	MLP-BP GANN		

Ce tableau indique, pour chaque étude, le ou les types de réseau employés, suivis, lorsque nécessaire, du nom de l'algorithme faisant partie de la procédure d'apprentissage, que la littérature désigne, un peu vite, comme étant l'algorithme d'apprentissage lui-même. Ainsi, MLP-BP désigne un réseau de type Multi-Layer Perceptron (Perceptron multi-couches) entraîné à l'aide de l'algorithme de rétro-propagation de l'erreur (Back-Propagation). Cet algorithme est employé dans 75 % des cas cités ici, alors que le réseau de type MLP l'est dans la quasi-totalité des expérimentations. La signification des acronymes est donnée ci-après. Ces acronymes sont ceux utilisés par les auteurs. Le « ? » signifie qu'aucune indication n'a été donnée au sujet du réseau et/ou de la méthode d'apprentissage employés.

Réseaux employés

ANN :	Athena Neural Network	GSOFFNN :	Generic Self-Organizing Fuzzy Neural Network
BM :	Boltzman Machine (machine de Boltzmann)	HLVQ :	Hidden Layer Vector Quantization (MLP suivi d'une carte de Kohonen)
CasCor :	Cascade Correlation	MLP :	Multi-Layer Perceptron (Perceptron multicouches).
CNN :	Competitive Neural Network (carte de Kohonen disposant d'une couche de sortie indiquant la classe d'appartenance d'une entrée)	PNN :	Réseau de neurones probabiliste (Probabilistic Neural Network ou fenêtre de Parzen)
G-Prop :	Réseau combinant un MLP avec un algorithme génétique pour effectuer l'apprentissage	RAA :	Réseau Auto-Associatif
GANN :	Generalized Adaptive Neural Network (réseau employant un algorithme génétique pour assurer l'apprentissage)	RHPNN :	Robust Heteroscedastic Probabilistic Neural Network (réseau de neurones probabiliste)
		SOM :	Self-Organizing Map (carte de Kohonen)

Algorithmes intervenant au cours de l'apprentissage

AG :	Algorithme Génétique	LM :	Levenberg-Marquardt
BP :	Back-Propagation	NCB :	Normalized Cumulative Back-Propagation
FLBP :	Functional Link Back-Propagation	OET :	Optimal Estimation Theory
GC :	Gradient Conjugué	PBP :	Pruned Back-Propagation
GRG2 :	Generalized Reduced Gradient	PCBP :	Predictive Cumulative Back-Propagation

Ce tableau fait ressortir la domination absolue d'un réseau, en l'occurrence le *Perceptron multi-couches* (MLP), qui fait partie de la famille des réseaux non-bouclés à apprentissage supervisé et d'un d'algorithme particulier employé au cours de son entraînement, appelé algorithme de *rétropropagation* (*Back-Propagation*) et ce, presque sans interruption depuis le début des années 90, laissant loin derrière, à la fois d'autres types de réseaux comme les *cartes de Kohonen*¹⁷,

¹⁷ Les *cartes de Kohonen* sont employées généralement pour réaliser des typologies, c'est-à-dire dans des situations où le problème consiste à découvrir si des observations peuvent être partitionnées en classes. Dans le cas de la prévision de défaillance, elles permettent aussi de construire des classifieurs. En effet, une carte constitue un résumé d'un ensemble d'observations opéré après un processus de quantification et

mais aussi d'autres algorithmes d'apprentissage, comme le *gradient conjugué* ou celui de *Levenberg-Marquart*. Cet engouement pour le *Perceptron multi-couches* s'explique en partie par la nature même des modèles de défaillance qui sont tous construits sur la base d'un processus qui, à partir d'échantillons de données, cherche à capter des régularités et à les engramer au cours d'un apprentissage. Parce que ce réseau participe d'une logique identique de construction de modèles, il s'est présenté comme pouvant s'inscrire facilement et avantageusement dans des expérimentations reposant sur des hypothèses similaires : aucun modèle n'a besoin d'être défini *a priori*, et seules les informations contenues dans les données sont de nature à en révéler un. À cet argument, s'en est ajouté un second, plus fondamental, qui a trait aux propriétés mathématiques d'approximation de la famille de réseau dont le *Perceptron multi-couches* fait partie, laissant espérer qu'il pouvait incarner une alternative crédible aux méthodes de classification ou de régression généralement employées.

Maintenant, l'engouement pour l'algorithme d'apprentissage s'explique moins par ses propriétés intrinsèques que par sa commodité d'utilisation, et surtout sa facilité d'implémentation. En effet, dans l'ensemble des techniques d'optimisation dont il fait partie, cet algorithme est loin de faire partie des plus performants, même si certains auteurs considèrent, pour des problèmes de classification, qu'il peut largement rivaliser avec des techniques bien plus sophistiquées (Le Cun et al., 1998).

Nous allons maintenant détailler¹⁸ tous ces aspects, en commençant par les propriétés, avant de nous arrêter sur les questions relatives aux modalités d'apprentissage et de paramétrage.

2.2 CARACTÉRISTIQUES DES RÉSEAUX DE NEURONES

Les caractéristiques que nous allons présenter sont importantes au regard de la nature des modèles de défaillance qui découlent d'un apprentissage. La problématique sous jacente à la conception de ces modèles est la suivante : aucune hypothèse n'est faite *a priori* sur le modèle qui régit le phénomène étudié, et l'on dispose uniquement d'une série de variables x_i (ratios,

chacun de ses neurones peut être vu comme un prototype d'un certain nombre d'entre elles. On peut donc affecter à chaque neurone le code de la classe qu'il représente le mieux, puis utiliser ce code comme critère de classement ; pour toute observation dont on souhaite prévoir la classe d'appartenance, on cherche le neurone dont elle est le plus proche au sens de la métrique utilisée par la carte et on affecte à l'observation en question la classe codée par le neurone.

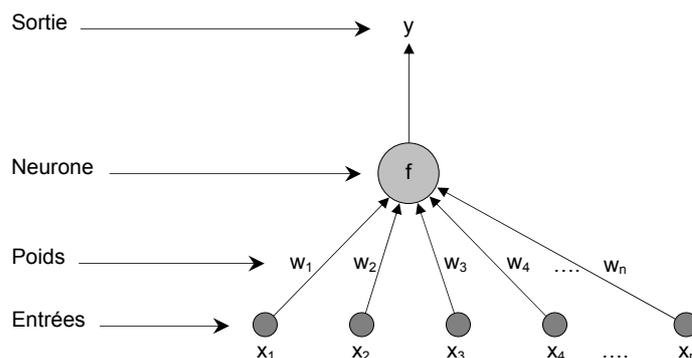
¹⁸ Nous ne détaillerons dans ce qui suit que ce qui est en lien direct avec le *Perceptron multi-couches*. Le lecteur intéressé par toutes les variantes de réseau, d'architecture, de topologie, de stratégie et d'algorithme d'apprentissage..., et plus généralement, par tout ce qui a trait à la problématique des réseaux, à leur formalisation mathématique, leur paramétrage et leur champ d'application, pourra se référer utilement à l'ouvrage de Dreyfus et al. (2004).

par exemple) contenant des mesures décrivant un échantillon d'observations (entreprises), et d'un ensemble de mesures z_k d'un élément caractéristique du phénomène en question (statut vis-à-vis de la défaillance par exemple). On suppose qu'il existe une relation entre les variables x_i et l'élément z_k et l'on cherche la fonction qui permet d'approcher au mieux cette relation, sachant que le nombre d'observations est restreint, que le nombre de variables d'intérêt est limité et que les mesures peuvent être entachées d'erreur. Mais ce qui nous intéresse dans la fonction à découvrir, c'est moins son exactitude au regard de l'échantillon employé, que sa capacité de généralisation, donc son aptitude à se comporter correctement face à des données qui ne sont pas intervenues lors de l'apprentissage. On cherche donc, *in fine*, un moyen de construire des modèles aptes à restituer la non linéarité entre un ensemble de variables financières et une probabilité de défaillance et présentant de bonnes capacités de généralisation. C'est précisément là, pour chacun de ces critères, que les propriétés de certains réseaux comme les *Perceptrons multicouches* sont attractives et c'est en ce sens que ces derniers offrent une alternative intéressante à des méthodes comme l'analyse discriminante ou la régression logistique.

2.2.1 ARCHITECTURE

La figure 1 illustre le formalisme que l'on emploie pour représenter un neurone. Il s'inscrit dans un autre formalisme élaboré par deux biologistes américains pour rendre compte du fonctionnement d'un neurone particulier du système nerveux central : le motoneurone. Leurs travaux ont mis en évidence un principe de fond sur lequel repose tous les modèles de neurones artificiels : sous l'influence des cellules nerveuses qui l'entourent, un neurone va réagir selon une fonction d'activation qui lui est propre et qui va lui permettre de transmettre ou non l'influx reçu.

Figure 1 : Architecture d'un neurone



Cette fonction représente en fait la relation qui lie la fréquence moyenne des potentiels pré-synaptiques à son potentiel de réponse ou potentiel somatique. Dans le cas d'un neurone artificiel, elle exprime une fonction non linéaire d'une combinaison des entrées x_i pondérées par des paramètres w_i , que l'on appelle des poids, ou efficacités synaptiques.

Le potentiel le plus utilisé correspond à une somme pondérée des entrées par les poids auquel s'ajoute un terme constant w_0 , appelé « biais ». Ce potentiel v s'exprime ainsi :

$$v = w_0 + \sum_{i=1}^{n-1} w_i x_i$$

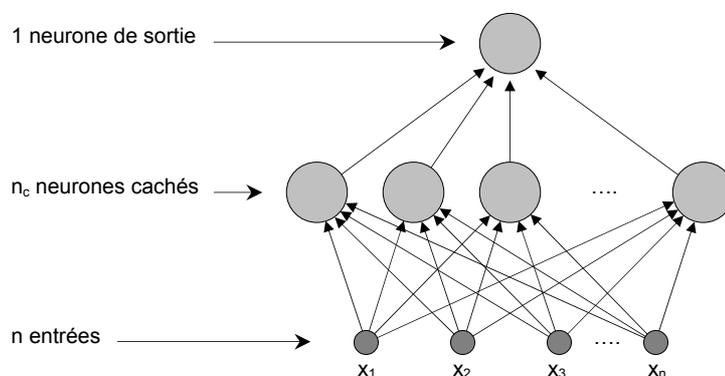
Dans de nombreux modèles, la fonction d'activation f est une sigmoïde, de type tangente hyperbolique ou arctangente, qui peut se représenter graphiquement par une fonction en forme de « s », symétrique par rapport à l'origine. La sortie d'un neurone s'exprime alors sous la forme :

$$y = f\left(w_0 + \sum_{i=1}^{n-1} w_i x_i\right)$$

Un réseau tire ses propriétés de la façon dont il va combiner n neurones, réalisant chacun une transformation non linéaire de ses entrées. C'est ce qui lui confère une forte aptitude à s'adapter et à modéliser des phénomènes non linéaires.

Un réseau de type *Perceptron* peut alors être représenté par un graphe orienté, à la topologie plus ou moins complexe, reliant une série d'entrées, correspondant aux variables décrivant le problème que l'on cherche à modéliser, à un ou plusieurs neurones de sortie, comme indiqué sur la figure 2 ci-après.

Figure 2 : Exemple d'architecture d'un réseau de neurones comprenant n entrées, n_c neurones cachés et 1 neurone de sortie



Ce genre de réseau est non bouclé, l'information étant véhiculée entre les entrées et la sortie sans possibilité de retour en arrière, en traversant une ou plusieurs couches dites « cachées ». Sur cette base, on peut imaginer un grand nombre de topologies, mais dans l'exemple du

Perceptron qui nous intéresse, et pour des tâches de classification, la topologie souvent utilisée est la suivante : chaque entrée est connectée à tous les neurones de la première couche cachée, qui peuvent eux-mêmes être connectés à tous les neurones d'une couche supplémentaire, et ainsi de suite selon qu'il y a plusieurs couches cachées, les derniers neurones étant enfin tous reliés à un seul neurone de sortie, lorsque le problème se résume à une classification en deux groupes.

Pour un problème de classification, la fonction d'activation de tous les neurones cachés est de forme identique à celle indiquée précédemment, alors que celle du neurone de sortie devra plutôt être de la forme suivante :

$$y = \frac{1 + f(v)}{2}$$

En effet, pour que les valeurs de sortie puissent être interprétées comme des probabilités, il est nécessaire qu'elles soient bornées entre 0 et 1.

Pour qu'un réseau puisse effectuer une classification, il faudra découvrir au préalable une fonction F qui sera susceptible de transformer un vecteur d'entrée x en une sortie y . En réalité, on ne disposera que d'une approximation $\hat{F}(x, w)$ de F qui, dans le meilleur des cas, sera telle que $\hat{F}(x, w) - F(x) \approx 0$. Dans cette formule, \hat{F} représente l'ensemble des fonctions de transformation possibles et w un ensemble de poids propre à cette fonction. Toute la difficulté sera de calculer un jeu de poids \hat{w} tel que, à l'intérieur d'une région à modéliser, l'écart entre la fonction F et son approximation soit inférieur à un seuil de divergence tolérable.

Les poids sont déterminés par apprentissage, selon une stratégie « supervisée ». Ce caractère « supervisé » réside dans le fait que l'on contraint le réseau à adopter un comportement prédéfini à partir d'une série de réponses connues *a priori*. Dans le cas qui nous intéresse, comme nous l'indiquions précédemment, on dispose d'un certain nombre de descripteurs (ratios) de chaque observation (entreprise) et l'on connaît, pour chacune d'elle, son statut vis-à-vis de la défaillance, que l'on peut coder par l'intermédiaire d'une valeur numérique. L'apprentissage consiste à faire en sorte que le réseau découvre progressivement la relation permettant d'associer à chaque observation, une valeur désignant son statut qui soit aussi proche que possible de la valeur théorique de référence, par un mécanisme itératif d'ajustement des poids. La contrainte s'exerce à partir des réponses du réseau et vise à orienter les modifications de poids de façon à ce que les réponses soient les plus conformes à ce qui est attendu.

Un réseau de neurones est donc un objet mathématique qui, à l'aide d'un algorithme d'apprentissage, permettra de rechercher dans un ensemble de fonctions paramétrées par les poids w , celle qui apparaîtra comme la plus adaptée à un échantillon de données. Pour autant, puisqu'il s'agit de découvrir une fonction inconnue, rien ne garantit qu'un réseau soit à même d'approximer n'importe quel type de fonction. Il faut donc pouvoir qualifier son pouvoir d'expression mais aussi l'analyser au regard de celui d'autres méthodes.

2.2.2 PROPRIÉTÉS

2.2.2.1 APPROXIMATION UNIVERSELLE

La première propriété des réseaux de neurones est celle d'approximation universelle que l'on doit à Hornik et al. (1989), Hornik et al. (1990) et Hornik (1991), citée de manière identique par Stoppiglia (1997), Monari (1999) et Dreyfus et al. (2004). Elle s'exprime ainsi :

« toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire. »

Et Hornik et al. (1989) rajoutent : « en ce sens, les réseaux de neurones à couches constituent une classe d'approximateurs universels »¹⁹.

Cette propriété signifie qu'un réseau peut approcher toute fonction connue. Or, ceci correspond rarement à une situation pratique, où l'on chercherait à reproduire avec un réseau une fonction que l'on connaît, et encore moins dans le domaine qui nous intéresse où l'éventuelle fonction en question nous est totalement inconnue. En revanche, les réseaux ayant toutes les qualités pour approcher une fonction de régression qui obéit aux critères énoncés, et notamment au critère de régularité, ils sont tout à fait à même d'estimer des probabilités *a posteriori* (Dreyfus et al., 2004), donc de résoudre des problèmes de classification²⁰.

¹⁹ Hornik et al. (1989), p. 359.

²⁰ Le pouvoir d'expression des réseaux a été étudié à de nombreuses reprises au travers d'études les comparant aux méthodes classiques de prévision (modèles Arma, modèles autorégressifs à seuils, modèles bilinéaires, modèles à noyau...) et Kouam et al. (1992), notamment, ont montré qu'il existe une inclusion de chacun de ces modèles dans les réseaux de neurones, donc que chacun d'entre eux peut être représenté par un réseau. En conséquence, dès lors que, dans la pratique, on ne peut pas déterminer le type de modèle le plus adapté à un problème donné, il est souhaitable de recourir à un réseau qui représente une généralisation d'un grand nombre de modèles possibles, plutôt que d'imposer a priori un modèle particulier.

C'est cette première propriété qui a fondé tout l'intérêt qu'on a pu porter aux réseaux en matière de prévision de la défaillance, puisque l'objectif premier était bien de résoudre des problèmes où la fonction de classement n'est pas connue tout en se départissant des insuffisances des méthodes linéaires. Or, la propriété la plus intéressante, n'est pas celle-ci, mais plutôt celle relative à leur capacité à fournir des modèles parcimonieux.

2.2.2.2 PARCIMONIE

L'objet d'une modélisation réside dans la mise au point de modèles de complexité minimale (Rivals et Personnaz, 2003) pour deux raisons. D'abord, la précision avec laquelle il est possible d'estimer les paramètres d'un modèle diminue en général avec le nombre de paramètres à estimer, au point de rendre parfois impossible leur identification si ces derniers sont trop nombreux. Ensuite, un modèle qui possède trop de paramètres libres présente généralement de mauvaises capacités de généralisation, car trop ajusté aux données qui ont permis de le créer. En ayant appris les particularités d'un échantillon, il n'est plus capable de s'en abstraire.

La recherche de la parcimonie permet de s'affranchir de ces deux écueils et là aussi, les réseaux de neurones apportent une réponse intéressante. La seconde propriété qui est la leur, que l'on doit à Hornik et al. (1994), citée aussi par Stoppiglia (1997) et Monari (1999), s'exprime ainsi :

« si le résultat de l'approximation (c'est-à-dire la sortie du réseau de neurones) est une fonction non linéaire des paramètres ajustables, elle est plus parcimonieuse que si elle est une fonction linéaire de ces paramètres. De plus, pour des réseaux de neurones à fonction d'activation sigmoïdale, l'erreur commise dans l'approximation varie comme l'inverse du nombre de neurones cachés, et elle est indépendante du nombre de variables de la fonction à approcher. Par conséquent, pour une précision donnée, donc pour un nombre de neurones cachés donné, le nombre de paramètres du réseau est proportionnel au nombre de variables de la fonction à approcher. »

Le véritable intérêt des réseaux à fonction d'activation sigmoïdale, pour qui la sortie des neurones n'est pas linéaire par rapport à leurs paramètres, comparativement à d'autres méthodes d'approximation, comme les polynômes, s'incarne dans une double propriété : pour une précision donnée, ils nécessitent moins de paramètres à ajuster et à nombre de paramètres identiques, ils sont plus précis (Dreyfus et al., 2004). Par construction, ils sont moins gourmands en nombre d'observations nécessaires à l'apprentissage, puisque ce nombre est une fonction du nombre de coefficients.

L'ensemble de ces résultats théoriques a donné une réelle légitimité aux réseaux surtout dans des domaines où les phénomènes étudiés ne sont pas réductibles à de simples combinaisons linéaires de variables. Il est vrai qu'il est souvent préférable d'employer des modèles linéaires ou des modèles polynomiaux, linéaires par rapport à leurs paramètres, car ils sont plus faciles à mettre en œuvre et à contrôler. Par ailleurs, le recours à d'autres méthodes doit se faire uniquement lorsque l'on est en présence de fortes non linéarités.

Manifestement, la problématique de prévision de la défaillance présente des caractéristiques qui justifient l'emploi des réseaux. Mais ceci ne doit pas faire oublier les difficultés liées à leur utilisation dues essentiellement aux techniques mathématiques en jeu, mais aussi aux contraintes induites par le processus d'apprentissage dont ils dépendent.

2.3 APPRENTISSAGE DES RÉSEAUX

Comme nous l'avons vu, un modèle élaboré à l'aide d'un réseau se construit au travers d'un processus d'apprentissage qui repose sur des données empiriques. On pose donc une hypothèse forte sur ces dernières, car on suppose qu'elles renferment une structure, inconnue, mais formalisable. On suppose aussi que cette structure présente suffisamment de régularités pour pouvoir être appréhendée sur la base d'un seul échantillon puisque, dans le cas qui nous intéresse, on ne disposera jamais d'une infinité de mesures.

Face à ces hypothèses, on se propose de découvrir la dite structure, à partir d'un échantillon fini, en recherchant parmi toutes les fonctions paramétrées par les coefficients du réseau (poids), celle qui se rapproche le plus de la fonction à l'origine des données. Typiquement, pour chaque observation d'un échantillon donné, on connaît les occurrences des descripteurs retenus (variables indépendantes), mais aussi l'occurrence de leur classe d'appartenance, définie a priori (variable dépendante), et on cherche la fonction qui permet d'expliquer la variable dépendante au travers d'une combinaison des variables indépendantes.

Pour découvrir cette fonction, on utilise une procédure d'adaptation progressive des paramètres du réseau en vue de le contraindre à donner une réponse désirée. C'est en ce sens que l'apprentissage est supervisé puisqu'il faut constamment contrôler les réponses et vérifier si elles sont ou non en adéquation avec ce qui est attendu, et si ce n'est pas le cas, effectuer les corrections nécessaires. On y parvient par l'intermédiaire d'une fonction coût qui permet d'évaluer l'écart entre les réponses du réseau et celles que l'on souhaite obtenir, et qui pénalise les erreurs.

2.3.1 FONCTION DE COÛT

Dans le cas de la régression, l'ensemble d'apprentissage est constitué d'observations pour lesquelles la sortie désirée est une variable continue. La fonction de coût la plus utilisée dans ce contexte est alors l'erreur quadratique : pour une observation i , issue d'un échantillon E , la fonction de coût des moindres carrés est égale à la somme, mesurée sur les n neurones de la couche de sortie, des carrés des écarts calculés entre les sorties du réseau y^i et les sorties désirées d^i , et s'exprime ainsi :

$$J^i(w) = \sum_{k=1}^n (y_k^i - d_k^i)^2$$

Pour l'ensemble de l'échantillon E , la fonction de coût correspond à la moyenne des carrés des écarts obtenus pour les N observations qui en font partie, soit :

$$J^E(w) = \frac{1}{N} \sum_{i=1}^N J^i(w)$$

Cette fonction de coût est issue du principe de maximum de vraisemblance avec une hypothèse gaussienne sur la distribution des sorties. Pour les problèmes de classification qui ne mettent en jeu que deux classes, la sortie désirée est une variable binaire codée 0 ou 1 selon que l'observation appartient à l'un ou l'autre groupe. L'hypothèse gaussienne sur la distribution des sorties n'est alors clairement plus vérifiée. C'est pourquoi on lui préfère, dans ce dernier cas, la fonction d'entropie croisée suivante :

$$J^i(w) = \sum_{k=1}^n d_k^i \log \left[\frac{y_k^i}{d_k^i} \right] + (1 - d_k^i) \log \left[\frac{1 - y_k^i}{1 - d_k^i} \right]$$

La fonction de coût dépend donc étroitement des observations employées. À cet égard, une attention particulière doit être portée à la nature de l'échantillon qui servira à l'estimation de l'erreur, car l'objectif principal n'est pas de trouver le modèle qui ajustera au mieux les données qui lui sont soumises, mais bien de construire celui qui présentera les meilleures capacités de généralisation. Nous y reviendrons ultérieurement.

2.3.2 TECHNIQUES D'OPTIMISATION

L'apprentissage revient donc à estimer un vecteur de paramètres qui minimise la fonction de coût choisie. Pour ce faire, on dispose de différentes techniques d'optimisation, plus ou moins complexes et efficaces, pour rechercher un minimum acceptable à cette fonction.

Les réseaux étant des modèles non linéaires par rapport à leurs paramètres, l'estimation des poids ne peut s'opérer qu'à l'aide de méthodes itératives, dès lors que la fonction de coût est dérivable par rapport aux paramètres du réseau. La stratégie consiste à effectuer une descente de coût dans l'espace des poids, étape par étape : on fixe aléatoirement un point de départ dans cet espace et on calcule la direction de descente de plus grande pente, donc celle qui correspond à la plus grande réduction d'erreur. Une fois la direction repérée, on franchit un pas dans ce sens, puis on recommence jusqu'à ce qu'une condition d'arrêt soit satisfaite.

La modification des poids w du réseau s'effectue alors ainsi (notation de Stoppiglia, 1997)²¹ :

$$w_i = w_{i-1} + \alpha_{i-1} d_{i-1}$$

avec i représentant l'itération où se produit le calcul, α le pas réglant l'amplitude de la descente et d indiquant le sens de la descente.

Les algorithmes existant se différencient par le choix du pas, qui peut être constant ou non, et de la technique de descente. Cette dernière s'appuie soit sur un algorithme de premier ordre, qui calcule la direction à partir des dérivées premières de la fonction de coût par rapport aux poids (vecteur appelé gradient de la fonction de coût et noté ∇J), soit sur un algorithme de second ordre qui, cette fois, calcule la direction à partir des dérivées secondes de cette même fonction par rapport au poids (matrice appelée Hessien de la fonction et notée H).

2.3.2.1 DESCENTE DE GRADIENT

Cette méthode est la plus simple, car elle ne fait appel qu'aux dérivées premières. Le calcul du gradient s'appuie sur l'algorithme de Rumelhart et al. (1986) dit de « rétropropagation de l'erreur », et la modification des poids obéit aux règles suivantes :

$$w_i = w_{i-1} - \alpha \nabla J_{i-1}$$

avec α représentant le pas, considéré ici comme une constante, et $d_{i-1} = -\nabla J_{i-1}$.

L'ajustement des paramètres de la fonction d'erreur s'opère dans le sens de sa descente, c'est-à-dire de la réduction de l'erreur globale, la mesure utilisée étant la dérivée première de la

²¹ La modification des poids n'intervient pas au même moment selon la méthode employée ou la stratégie d'apprentissage. Elle peut se produire soit après la présentation au réseau d'une observation, soit après la présentation de toutes les observations de l'échantillon. Une descente de gradient opérée observation après observation est dite « stochastique », par opposition à une descente dite en mode « batch », car réalisée une seule fois pour l'ensemble des observations. Certaines techniques, et notamment de second ordre, ne fonctionnent que dans ce dernier mode.

fonction d'erreur, donc sa pente. La pente est une mesure locale de la variation de la fonction d'erreur suite à une variation d'une unité du paramètre analysé. Le coefficient d'apprentissage représente l'ampleur du déplacement que l'algorithme va réaliser dans le sens de la pente. Si ce coefficient est petit, les déplacements successifs le seront aussi, permettant une convergence précise vers un minimum. En effet, à chaque itération, la pente sera estimée à nouveau, permettant ainsi d'épouser précisément la fonction d'erreur.

Cet algorithme a été souvent critiqué, surtout à cause de sa lenteur, et a fait l'objet d'un nombre considérable de travaux visant à améliorer le processus et la vitesse de convergence, en intégrant des calculs permettant à l'algorithme de s'adapter aux variations de la surface d'erreur.

De nombreuses études se sont focalisées sur la façon de calculer le pas. Certains auteurs proposent d'employer un pas adaptatif selon l'erreur obtenue, comme Battiti (1989) ou Kruschke et Movellan (1991), d'autres selon le cycle d'apprentissage de l'échantillon employé (Eaton et Olivier, 1992)... D'autres encore suggèrent d'utiliser un pas propre à chaque poids afin de s'assurer qu'ils convergent bien tous à la même vitesse. Ainsi, Le Cun et al. (1998), constatant que les dérivées secondes de la fonction de coût par rapport aux poids proches de la sortie sont souvent plus faibles que celles des poids proches des entrées, recommandent d'employer des pas plus importants sur les dernières couches que sur les premières, pour compenser ces écarts.

Quelques travaux se sont attachés à apporter des compléments à l'algorithme lui-même. Ainsi, Rummelhart et al. (1986) proposent-ils d'y adjoindre un terme de « moment » afin d'accélérer la descente dans la direction où la pente est la plus forte, tout en stabilisant le processus de recherche lorsqu'il est susceptible d'osciller. La technique revient à combiner des gradients successifs en ajoutant une proportion fixe de la valeur employée pour mettre à jour un poids à l'itération $i-1$, à la valeur employée pour cette même mise à jour, mais à l'itération i . La règle d'adaptation des poids est alors la suivante :

$$w_i = w_{i-1} - \alpha \nabla J_{i-1} + \mu \Delta w_{i-1}$$

avec μ représentant le terme de moment et Δw_{i-1} la variation des poids à l'itération $i-1$.

D'autres encore ont cherché à apporter des modifications plus substantielles à l'algorithme de base, comme Fahlman (1988) ou Riedmiller et Braun (1993). Le premier tente de limiter l'incidence de la saturation de la fonction d'activation des neurones. Il constate en effet que, lorsque

l'on se rapproche du minimum ou du maximum de la fonction sigmoïde, la pente se rapproche de zéro et donc que l'ajustement des poids tend aussi vers zéro. Il propose alors un algorithme (QuickProp) qui utilise une fonction d'activation modifiée pour résoudre ce problème, ainsi qu'un coefficient d'apprentissage et un moment dynamiques. Quant aux seconds, ils tentent, avec leur méthode (Rprop), d'amoindrir l'influence néfaste de l'amplitude de la dérivée partielle de l'erreur sur la variation des poids. Ici, seul le signe est employé pour indiquer la direction du changement des poids, l'amplitude de la variation étant contrôlée par un paramètre particulier δ , dont la valeur est déterminée selon le signe des gradients calculés lors de deux itérations successives. La logique est la suivante : lorsque la variation d'un poids est trop importante, la dérivée partielle du poids considéré change de signe. Ceci indique que l'algorithme a franchi un minimum local ; dans ce cas, δ est réduit d'un facteur η^- afin de revenir en arrière, facteur fixé aux alentours de 0,5. En revanche, lorsque la variation ne conduit pas à un changement de signe, δ est légèrement augmenté afin d'accélérer l'apprentissage dans les régions peu profondes, d'un facteur η^- de l'ordre de 1,2.

Malgré ces améliorations, il demeure des situations qui n'ont pas reçu de réponses algorithmiques satisfaisantes. C'est le cas lorsque le gradient, à un moment donné, tend vers zéro. Ceci peut être révélateur, soit de la proximité d'un minimum de la fonction de coût, soit de la présence d'une surface plane de cette même fonction et rien ne dit dans ce cas de figure, quelle est l'origine exacte d'un gradient presque nul : est-on en présence d'une portion de surface plane, ou d'un minimum ? C'est aussi le cas lorsque la surface d'erreur est fortement perturbée, la direction prise initialement pouvant être située à l'opposé de celle qui conduit réellement à un minimum. C'est pourquoi d'autres types de méthodes sont envisageables, qui ne s'appuient plus sur les dérivées premières pour modifier les paramètres du réseau, mais sur des données de second ordre.

2.3.2.2 MÉTHODES NEWTONIENNES

Les méthodes newtoniennes s'appuient sur les dérivées secondes de la fonction de coût pour parvenir à son minimum. La modification des paramètres du réseau s'effectue ainsi :

$$w_i = w_{i-1} - H_{i-1}^{-1} \nabla J_{i-1}$$

avec H^{-1} représentant l'inverse de la Hessienne, α étant constant est égal à 1, et $d_{i-1} = -H_{i-1}^{-1} \nabla J_{i-1}$.

Ces méthodes permettent, par exemple, pour des fonctions quadratiques, d'obtenir immédiatement, en une seule itération, le minimum global de la fonction. Leur avantage est double : elles convergent rapidement à l'approche du minimum et la vitesse de convergence ne dépend pas de la direction de recherche. En revanche, il n'y a de convergence possible qu'à la condition que la Hessienne soit définie positive, ce que rien ne garantit. De plus, le calcul de cette matrice et son inversion doivent être effectués à chaque itération, ce qui devient rédhibitoire au-delà de quelques variables. En effet, pour une matrice de dimension (n,n) , ce calcul requiert n^3 opérations, n représentant le nombre de poids.

2.3.2.3 MÉTHODES QUASI-NEWTONIENNES

Contrairement aux précédentes, les méthodes quasi-newtoniennes cherchent à utiliser l'information de la Hessienne mais sans l'inverser. Pour ce faire, elles utilisent une approximation \hat{H} de l'inverse de la Hessienne qui est réalisée à chaque itération, \hat{H} étant définie positive.

La modification des poids obéit aux règles suivantes :

$$w_i = w_{i-1} - \alpha_{i-1} \hat{H}_{i-1} \nabla J_{i-1}$$

avec $d_{i-1} = -\hat{H}_{i-1} \nabla J_{i-1}$ et α étant variable et évalué à chaque itération.

Une méthode quasi-newtonienne, comme BFGS, possède de bonnes propriétés de convergence tout en étant assez rapide, l'approximation de la Hessienne ne demandant que n^2 opérations, et numériquement stable grâce à une Hessienne symétrique et définie positive. Bien qu'étant plus rapide qu'une méthode newtonienne, elle n'en demeure pas moins beaucoup plus lente qu'une méthode qui repose sur une descente de gradient simple qui, elle, ne nécessite que n opérations par itération. BFGS est vraisemblablement plus efficace que l'algorithme de rétropropagation pour des problèmes de taille modérée traités avec de petits réseaux (Battiti, 1992 ; Le Cun et al., 1998).

2.3.2.4 METHODE DE LEVENBERG-MARQUARDT

La méthode de Levenberg-Marquardt s'appuie sur une approximation de la Hessienne à l'aide du carré de la matrice Jacobienne (matrice contenant les dérivées premières de la fonction de coût).

La modification des poids s'opère alors ainsi :

$$w_i = w_{i-1} - (H_{i-1} + \lambda_{i-1} I)^{-1} \nabla J_{i-1}$$

avec λ représentant un paramètre de régularisation, et I la matrice identité.

Le paramètre de régularisation permet à cette méthode de s'adapter à des situations où la Hessienne est indéfinie, donc dans des situations où certaines valeurs propres sont faibles, comparativement aux plus grandes, ou négatives. Bien que s'appuyant sur une approximation de la Hessienne, cette méthode présente une complexité identique à celle des méthodes newtoniennes et requiert n^3 opérations par itération. La méthode de Levenberg-Marquardt est une variante de celle de Gauss-Newton et elle n'en diffère qu'à un détail près : elle possède juste un paramètre de régularisation que l'autre n'a pas.

D'autres méthodes existent, que l'on peut difficilement classer dans ces catégories, comme celle du *gradient conjugué*, qui s'appuie notamment sur une comparaison des gradients obtenus au cours de deux itérations successives.

2.3.2.5 CHOIX D'UNE TECHNIQUE

D'un point de vue général, aucune méthode ne garantit une convergence vers un minimum global, sauf dans des cas de figures théoriques. Les surfaces d'erreur sont souvent accidentées et le choix d'une méthode sur la seule base des propriétés que nous venons d'évoquer n'est pas suffisant pour garantir que le réseau convergera vers un minimum acceptable, ni surtout que la solution trouvée sera de nature à remplir les objectifs du modèle recherché, car, par-delà l'algorithme et son paramétrage, c'est toute la stratégie d'apprentissage qui conditionnera effectivement les performances d'un modèle.

En pratique, le choix d'une méthode est fonction de la taille des échantillons, de la complexité de la fonction à approcher, et de l'économie de l'apprentissage. Dans bien des cas, ce sont des règles empiriques qui dictent l'emploi de tel ou tel algorithme comme ses conditions d'implémentation. À cet égard, Le Cun et al. (1998) prodiguent les recommandations suivantes :

- les techniques employant les données de la Hessienne (BFGS, Gauss-Newton et Levenberg-Marquardt) ne peuvent s'appliquer qu'à de petits réseaux entraînés en mode *batch*, alors que de tels réseaux ne sont pas ceux nécessitant le plus une accélération de l'apprentissage ;
- un *gradient stochastique* correctement paramétré peut difficilement donner de moins bons résultats qu'une autre méthode, lorsqu'il est confronté à un problème de classification portant sur un grand ensemble de données ;
- pour des jeux de données de petite taille, où l'on cherche à effectuer des calculs de précision, un *gradient conjugué* offre le meilleur compromis rapidité-efficacité-simplicité. Si l'échantillon

est important (plusieurs centaines d'observations) et comporte des données redondantes, pour une tâche de classification, il est préférable d'employer un *gradient stochastique* correctement réglé ou une version stochastique de la méthode de Levenberg-Marquardt. Si l'échantillon n'est pas trop important et que la tâche consiste à effectuer des calculs de précision, alors il est souhaitable d'utiliser un *gradient conjugué*.

2.3.3 STRATÉGIES D'APPRENTISSAGE

L'objectif de toute modélisation statistique ne consiste pas à ajuster fidèlement un modèle à un ensemble de données, mais à faire en sorte que le modèle réagisse correctement face à des données qui n'ont pas servi à son apprentissage, donc qu'il soit capable de généraliser. Comme on dispose d'un nombre limité d'observations, l'information qu'elles contiennent n'est pas forcément suffisante pour reconstituer la réalité que tente d'approcher le modèle aux endroits précis où les observations manquent. Il faut donc s'assurer que cette absence d'information ne sera pas préjudiciable à son comportement, ou tenter de l'atténuer.

La capacité de généralisation d'un modèle est fonction de deux paramètres : elle dépend des données de départ et de leur représentativité du phénomène étudié, mais aussi de la technique employée pour éviter un surajustement.

2.3.3.1 CHOIX DES DONNÉES

On sait depuis longtemps déjà que toute généralisation requiert une connaissance a priori du problème traité. Ceci suppose que, face à un problème prédéfini, on dispose des entrées pertinentes et que l'on puisse évaluer le degré de ressemblance des données dont on dispose avec celles pour lesquelles le réseau devra être capable de généraliser. Deux conditions doivent donc être remplies :

- les entrées du réseau doivent contenir suffisamment d'informations susceptibles de décrire la sortie attendue du réseau, de telle façon qu'il existe une fonction mettant en relation les entrées et les sorties avec une degré de précision suffisant ;
- l'échantillon d'apprentissage doit être suffisamment large et représentatif de la population étudiée pour éviter que le réseau ne soit un jour confronté à une situation par trop éloignée de celle qu'il a appris à reconnaître (Le Cun, 1989). La généralisation intervient soit grâce à un mécanisme d'interpolation, soit grâce à un mécanisme d'extrapolation. Le premier s'applique

aux observations qui sont plus ou moins proches des observations employées pour l'apprentissage. Le second, lui, intervient dans tous les autres cas de figure et notamment lorsque les observations sont en dehors de l'espace délimité par l'échantillon d'apprentissage ou dans des régions vides de cet espace. De fait, l'extrapolation n'est pas un processus fiable. Il faut donc posséder suffisamment de données pour éviter d'y avoir recours.

2.3.3.2 SURAJUSTEMENT

Le « surajustement » incarne une situation extrême où le modèle a appris les caractéristiques propres des données au dépend de leur structure. Le modèle possède donc suffisamment de degrés de liberté pour « passer » au plus juste d'un certain nombre d'observations. Or, seule la structure générale à un réel intérêt car elle représente ce qui peut être commun et permanent à tout échantillon provenant d'une même population, alors que les particularismes d'un échantillon n'ont qu'un aspect conjoncturel.

La difficulté de l'apprentissage réside dans le contrôle de ce surajustement car il n'y a pas de méthode ou critère définitifs garantissant que le seuil n'a pas été franchi. À tout le moins dispose-t-on d'un ensemble de techniques permettant de le limiter.

Le surajustement, ou « surapprentissage », s'explique traditionnellement au travers de la décomposition de l'erreur obtenue avec les modèles de régression. Elle comporte trois termes : un terme de « bruit » qui est une sorte d'erreur irréductible, un terme de « biais » qui mesure la distance entre la « véritable » fonction et la meilleure fonction admissible dans la classe que l'on a choisie, et un terme de variance qui représente la distance moyenne entre la fonction estimée sur l'ensemble d'apprentissage et la meilleure fonction dans la classe choisie (Chapados, 2003).

Si la fonction est trop simple compte tenu du phénomène à modéliser, le biais sera important, car la probabilité qu'elle soit proche de la véritable fonction dans la classe choisie sera faible, mais pas la variance, car l'erreur dépendra peu de l'échantillon. En revanche, si elle est trop complexe, son biais sera faible, car la fonction estimée sera proche de la fonction la plus représentative de l'ensemble d'apprentissage, mais sa variance élevée, car cette fois, l'erreur dépendra d'abord de la structure de l'échantillon employé : c'est cette situation que désigne le surajustement.

Mais cette décomposition de l'erreur, comme la vision qui va avec, où l'augmentation d'un terme se traduit pas la diminution de l'autre, puisque la relation est additive, s'adapte mal aux

réseaux de neurones à couches et vaut essentiellement pour les problèmes de régression et non de classification. Comme le rapporte Stricker (2000), Gallinari et Cibas (1999) ont montré, au travers de leurs expérimentations sur des problèmes de régression, qu'un réseau avec quinze neurones cachés pouvait présenter une plus grande variance qu'avec soixante, dans une situation où l'apprentissage est effectué sur un échantillon de petite taille, indiquant dans ce cas que le biais et la variance ne changent pas nécessairement en sens contraire. De plus, Stricker (2000) montre que le surajustement selon la tâche à effectuer n'intervient pas de la même manière : pour la régression, il se fait sur l'ensemble des données, de manière égale, alors que pour la classification, il a plutôt tendance à se concentrer autour de la frontière de séparation de classes.

Ceci aura donc des conséquences sur la façon de contrôler le surapprentissage avec les réseaux que nous allons maintenant présenter.

2.3.3.3 CONTRÔLE DU SURAJUSTEMENT

Le contrôle de l'ajustement se pose avec n'importe quel modèle en des termes identiques dès lors que les qualités de généralisation sont recherchées. Mais avec les réseaux non bouclés à fonction d'activation sigmoïde, le contrôle de la complexité des modèles se fait plus facilement qu'avec d'autres méthodes en raison de leur parcimonie.

Pour juger de la capacité de généralisation, on peut procéder de deux manières :

- une fois l'apprentissage terminé, sur la base d'une estimation de l'erreur dite de « généralisation » calculée sur un échantillon de test ou par une technique de rééchantillonnage, comme la validation croisée, ou sur la base de mesures d'information. Ces méthodes ont pour objet de sélectionner les meilleurs modèles parmi un ensemble considéré ;
- au cours de l'apprentissage, à l'aide de techniques de régularisation, comme l'arrêt prématuré ou la modération des poids, qui servent à contraindre le modèle pour qu'il ne soit pas sujet à un surapprentissage. Cependant, l'emploi de ces techniques n'est pas exclusif de l'emploi des précédentes si l'on souhaite estimer les performances des modèles.

La question de l'évaluation des capacités de généralisation n'est traitée dans ce chapitre qu'indépendamment de celle qui concerne le choix des variables ou la définition d'une architecture optimale. Les questions relatives aux techniques de sélection des variables seront

abordées dans le chapitre 3, alors que celles relatives à l'optimisation de l'architecture ne le seront pas en tant que telles²².

2.3.3.3.1 CONTRÔLE UNE FOIS L'APPRENTISSAGE TERMINÉ

Le contrôle à la fin de l'entraînement s'appuie sur l'idée que l'erreur empirique mesurée sur un échantillon d'apprentissage n'a pas le même comportement que celle mesurée sur un autre échantillon, tiré de la même distribution. En effet, à mesure que la complexité d'un modèle croît, sa précision sur l'ensemble d'apprentissage augmente, comme son ajustement aux données, et lorsque l'ajustement devient trop important, le modèle se trouve dans une situation qui pénalisera ses capacités de généralisation. Ce qui laisse supposer qu'au cours de l'apprentissage le modèle est susceptible de franchir un seuil au-delà duquel il ne sera plus apte à généraliser correctement. C'est précisément ce seuil, ou cet instant, que les méthodes de contrôle « *ex post* » se proposent de détecter.

2.3.3.3.1.1 ENSEMBLE DE TEST

La méthode la plus classique consiste à partitionner les données dont on dispose en deux ensembles disjoints : un échantillon d'apprentissage et un échantillon de test. La détection du surajustement s'opère donc en observant, à intervalles réguliers, le comportement de l'erreur sur les deux échantillons. Généralement, tant que le modèle n'est pas surajusté, l'erreur sur les deux ensembles diminue à mesure que l'entraînement progresse. Lorsqu'un surajustement se produit, alors que l'erreur continue de décroître sur le premier ensemble, elle commence à augmenter sur le second, laissant présager qu'un surapprentissage vient de se produire.

Des techniques dérivées de cette méthode, comme la validation croisée ou le *leave-one out*, consistent, non pas à utiliser un seul échantillon de test, mais plusieurs. Dans le cas de la validation croisée, on découpe l'échantillon initial en k parties (souvent avec $k = 10$)²³, de taille égale, et on réalise k apprentissages en laissant à chaque fois une partie de côté pour

²² Les techniques d'optimisation de l'architecture d'un réseau concourent à trouver la structure la plus adaptée à un problème donné, tout en faisant en sorte qu'elle soit la plus simple possible. Elles s'appuient soit sur des algorithmes destructifs, qui élaguent les poids ou les neurones cachés, soit sur des algorithmes constructifs, qui sont des méthodes de construction incrémentale. Elles peuvent donc, par ricochet, être avantageusement employées pour trouver les meilleures variables. La limite entre ces techniques et celles de sélection des variables est ténue, c'est pourquoi, nous n'aborderons ici que celles qui, dans la littérature, ont un lien direct avec la problématique de sélection. Par exemple, tout ce qui a trait à l'élagage des neurones, comme les méthodes de « *skeletonization* », ne sera pas abordé.

²³ La littérature recommande en effet $k = 10$. En pratique, k dépend du nombre de données disponibles, et il est recommandé de tester différentes valeurs pour ce paramètre.

valider les résultats obtenus à partir d'un échantillon composés des $k-1$ parties restantes. L'estimation de l'erreur est donc calculée par la moyenne pondérée des erreurs obtenues sur chaque sous-échantillon. Le *leave one out* représente la limite du processus de validation croisée, c'est-à-dire lorsque le nombre d'observations dans chacune des k parties est égal à 1.

La validation croisée apporte de meilleurs résultats que l'emploi d'un seul échantillon de test lorsque le nombre d'observations est faible, car elle permet de mieux tirer partie de toute l'information contenue dans les données. En effet, isoler un échantillon de test revient à se priver d'une partie des données pour l'apprentissage et soustrait ainsi au futur modèle un pan peut-être important de la réalité qu'il doit apprendre.

Le *bootstrap* offre une alternative à ces méthodes. En effet, le *leave-one-out* fournit une estimation de l'erreur en généralisation presque non biaisée, mais avec une variance importante. Le *bootstrap* peut alors sensiblement réduire cette variance (Efron et Tibshirani, 1995), mais au prix de calculs bien plus considérables. Dans sa forme la plus simple, au lieu de répéter l'analyse de sous-ensembles des données, on répète celle de sous-échantillons. Chaque sous-échantillon représente un échantillon aléatoire avec remise de l'échantillon initial.

De nombreux résultats aussi bien théoriques qu'expérimentaux ont montré qu'une augmentation significative des coûts de calcul pour estimer l'erreur d'un classifieur ne conduit pas nécessairement à des résultats bénéfiques, surtout si la précision relative d'une procédure de classification est plus importante que l'obtention de valeurs exactes (Kohavi, 1995). Or les temps de calcul qui séparent une validation à partir d'un échantillon de test d'une procédure de *bootstrap* sont tellement importants, qu'une solution couramment retenue revient à effectuer une validation croisée.

Pour autant, ces méthodes de contrôle sont surtout efficaces pour les problèmes de régression, car le surajustement affecte uniformément l'espace, ce qui n'est pas le cas des problèmes de classification, puisque, comme nous l'avons indiqué précédemment, le surajustement se produit dans des zones spécifiques, à proximité de la frontière entre les classes. Si le surajustement ne concerne que quelques points, il sera donc difficilement détectable. De telles méthodes ne semblent donc pas les plus appropriées pour détecter un surapprentissage d'un modèle de classification (Stricker, 2000).

2.3.3.3.1.2 MESURES D'INFORMATION

Les mesures d'information complètent les précédentes en ce sens qu'elles visent à fournir un critère qui ne tient pas seulement compte de l'estimation de l'erreur en généralisation, mais aussi de la complexité du modèle. C'est une façon de pénaliser l'erreur en la pondérant par une valeur qui dépend du nombre de paramètres et qui vise surtout à privilégier les modèles parcimonieux. Elles sont particulièrement utiles pour arbitrer entre plusieurs règles qui présentent le même degré de performance ou des degrés assez similaires. La fonction correspondante comprend d'un côté, une variable dont les valeurs décroissent avec la précision, et de l'autre, une variable dont les valeurs augmentent avec la complexité. Le modèle retenu *in fine* sera donc celui pour lequel la fonction renverra la plus petite valeur. Il existe un certain nombre de mesures de ce type parmi lesquelles on compte, le critère d'Akaike (Akaike's Information Criterion : AIC), le critère de Bayes (Bayesian Information Criterion : BIC) ou le critère de Schwarz (Schwarz's Bayesian Criterion : SBC). Elles s'expriment ainsi :

$$\text{AIC} = n[\log(\text{EQM})] + 2p$$

$$\text{BIC} = n[\log(\text{EQM})] + 2(p + 2)q - 2\left(\frac{\sigma^2}{\text{EQM}}\right)^2$$

$$\text{SBC} = n[\log(\text{EQM})] + p[\log(n)]$$

avec n représentant la taille de l'échantillon, p le nombre de paramètres (ou nombre de variables pour un modèle linéaire par rapport aux paramètres), σ l'estimation de la variance de l'erreur du modèle complet et EQM (écart quadratique moyen) la moyenne du carré des écarts entre la sortie du réseau et la valeur attendue.

Toutes ces techniques jouent en réalité sur le nombre de paramètres en jeu et tentent de favoriser, à précision égale, les modèles disposant du plus petit nombre de neurones cachés. Elles ne permettent pas de détecter un surajustement, mais tentent juste de le prévenir et sont délicates à employer pour la classification.

2.3.3.3.2 CONTRÔLE EN COURS D'APPRENTISSAGE

Le contrôle en cours d'apprentissage se fait par l'intermédiaire de techniques de régularisation qui, comme leur nom l'indique, ont pour objet de transformer les modèles qui disposent de singularités en certains endroits de leur espace de représentation, en modèles les plus réguliers possibles, notamment aux endroits où le surajustement s'est produit.

Deux ensembles de méthodes permettent d’y parvenir : le premier, dit « d’arrêt prématuré » (*early stopping*), consiste à mettre un terme à l’apprentissage avant que le réseau n’ait pu converger vers un minimum ; le second, dit de « pénalisation de la fonction d’erreur », revient à ajouter un terme à cette fonction (un *weight decay* par exemple) afin de la rendre plus régulière.

2.3.3.3.2.1 ARRÊT PRÉMATURÉ

L’arrêt prématuré met un terme à l’apprentissage avant que le réseau n’ait eu le temps de s’ajuster trop précisément aux données qui lui ont été soumises, même si aucun minimum n’a été atteint. Pour y parvenir, comme précédemment, il faut suivre l’évolution de l’entraînement sur un autre échantillon au fur et à mesure de sa progression et observer le comportement de l’erreur sur ce deuxième ensemble. Une heuristique courante consiste à arrêter l’apprentissage lorsque l’erreur sur l’échantillon de test commence à augmenter, révélant ainsi quelque signe de surapprentissage. Mais se pose alors le même problème que précédemment pour les tâches de classification, car l’erreur en question peut ne pas évoluer dans le sens que l’on imagine alors que le surajustement du réseau progresse.

Par ailleurs, l’arrêt prématuré ne permet pas de définir une architecture optimale en termes de complexité et de précision et va, par construction, à l’encontre du principe de parcimonie, puisqu’il autorise des modèles surparamétrés, c’est-à-dire contenant trop de paramètres à estimer compte tenu du nombre d’observations sur lesquelles prendront appui les estimations.

Le tableau 23, p. 108, illustre bien ce point. Il décrit l’architecture de réseaux employés pour expérimenter des modèles de défaillance et les tailles des échantillons qui ont servi à leur mise au point : on constate que certains réseaux comportent bien plus de paramètres que d’observations qui ont concouru à leur apprentissage. Dans ces cas de figures, les auteurs ont généralement procédé par arrêt prématuré pour mettre fin à l’apprentissage, sans toujours contrôler la complexité du réseau.

2.3.3.3.2.2 MODÉRATION DES POIDS

Au cours d’un apprentissage, les poids sont constamment modifiés dans le sens d’une réduction d’erreur et ceci peut conduire certains d’entre eux à enfler considérablement, entraînant alors une saturation des fonctions d’activation des neurones. En effet, la valeur du potentiel somatique est tellement élevée que ces fonctions renvoient des valeurs proches de la valeur limite de réponse. Or, pour obtenir une fonction régulière (de classement ou de régression), il est

préférable que le réseau travaille sur la partie plus ou moins linéaire des fonctions d'activation, donc que les poids ne soient pas trop grands en valeur absolue.

Sticker (2000) a montré, au travers de plusieurs expérimentations, que des poids trop grands conduisent à des surfaces de séparation comportant des changements brusques, ce qui signifie que la moindre variation pouvant affecter les entrées risque de conduire à de grandes variations dans les prévisions. La régularisation par modération des poids vise alors à lisser la surface de réponse pour la rendre plus régulière donc moins sujette à de telles amplitudes de variation. Pour y parvenir, la fonction de coût à minimiser est complétée par un terme de régularisation qui dépend de la somme des carrés des paramètres du réseau. Ainsi, la fonction de coût J décrite précédemment est-elle remplacée par une nouvelle fonction J' qui peut s'exprimer ainsi :

$$J' = J + \frac{\beta}{2} \sum_{k=1}^n w_k^2$$

avec k désignant le nombre de poids et β un hyperparamètre réglant l'amplitude de la régularisation. Cet hyperparamètre se détermine empiriquement, car s'il est trop grand, les poids tendent vers zéro, et s'il est trop petit, son influence disparaît. Il doit donc être adapté au problème concerné. Comme le souligne Stricker (2000), rappelant McKay (1992), le paramètre β ne peut pas être unique, car il faut tenir compte de sa position selon les couches ; c'est pourquoi, plusieurs hyperparamètres doivent être employés à la place d'un seul : par exemple, un premier qui est propre aux poids reliant les entrées aux neurones cachés, un second propre à ceux reliant les biais aux neurones cachés et un troisième propre aux poids qui sont reliés au neurone de sortie. Là aussi, ces paramètres doivent être déterminés empiriquement selon la nature des données en testant différentes combinaisons, ou en employant une approche bayésienne, ainsi que l'a montré Stricker (2000).

2.4 PARAMÉTRAGE DES MODÈLES DE DÉFAILLANCE

En introduction de ce chapitre, nous donnions un aperçu rapide des différents types de réseaux et algorithmes d'optimisation employés dans la littérature financière pour réaliser des modèles de défaillance. Comme la logique que nous poursuivons vise à offrir un cadre de comparaison avec ce qui a été fait dans le domaine jusqu'à présent, nous avons étudié la façon dont tous les auteurs que nous avons référencés ont procédé pour paramétrer leurs réseaux et gérer les problèmes liés à l'apprentissage. Les tableaux 23, 24 et 25 présentés ci-après offrent une synthèse de leurs principaux choix méthodologiques que nous allons commenter.

En règle générale, ces auteurs donnent peu d'indications sur les méthodes mises en œuvre pour paramétrer leurs réseaux. On sait assez peu de choses notamment sur la façon dont ils ont contrôlé la descente de gradient et sur les éventuelles techniques de régularisation auxquelles ils ont pu avoir recours. Et lorsqu'ils ne précisent rien en ces domaines, la problématique sous-jacente à ces méthodes n'est pas plus détaillée. Nous n'avons donc retenu, dans les tableaux qui suivent, que des paramètres ayant été cités à quelques reprises.

Le tableau 23 indique que presque toutes les architectures sont organisées autour d'une couche cachée, avec un seul neurone de sortie, la couche cachée comportant dans l'ensemble un nombre assez limité de neurones, compris entre 6 et 10 en moyenne et le nombre d'entrées variant en moyenne entre 10 et 15.

Tableau 23 : Taille des échantillons et architectures des réseaux de type MLP employés pour la construction de modèles de prévision de la défaillance

Études	Taille des échantillons		Architecture des réseaux	Études	Taille des échantillons		Architecture des réseaux
	App.	Test			App.	Test	
AGARWAL (1999)	80	60	4:8:1	LESHNO et SPECTOR (1996)	44	132	41:?:?
AGARWAL et al. (2001)	220	115	10:15:1		44	44	41:?:??
ALTMAN (1994)	404	404	[?:15]:[?:?:]1		132	44	
	808	808	[?:15]:[?:15]:[?:12]:2	LI et GUPTA (2002)	180	20	4:[2-6]:1
ANANDARAJAN et al. (2001)	418	104	5?:1		170	30	6:[2-6]:1
ATIYA (2001)	491	669	[5-6]:2:1	MIN et LEE (2004)	310	103	10:[8-32]:1
BACK et al. (1994)	38	38	[3-11]:[2-10]:1	MIN et LEE (2005)	1 132	378	11:[8-32]:1
BACK et al. (1996 A)	74			ODOM et SHARDA (1990)	Var	Var	5:5:1
BACK et al. (1996 B)	36	38	[2-21]:[7-12]:1	PENDHARKAR (2005)	100	100	5:10:1
BACK et al. (1997)	400	170	[12-19]:4:1	PIRAMUTHU et al. (1998)	166	16	[10-11]:[6-8]:1
	200	170		POMPE et BILDERBEEK (2005 A)	Var.	Var.	
	100	170		POMPE et BILDERBEEK (2005 B)	Var.	Var.	9:[0-4]:1
BAEK et CHO (2003)	3 634	4 570	5:10:1	POMPE et FEELDERS (1997)	288	288	10:[0-6]:1
BARNIV et al. (1997)	114	123		RAHIMIAN et al. (1993)	74	55	5:5:1
BELL et al. (1990)	204	1 059	11:[6]:1	SALCHENBERGER et al. (1992)	200	116	5:3:1
BERG (2005)	60 600	40 400	23?:1	SEN et al. (2004)	160	129	4:25:2
BORITZ et KENNEDY (1995)	230	112	[5-13]:[5-13]:1				12:25:02
			[5-13]:?:1	SERRANO-CINCA (1997)	66		[10-11]:[6-8]:1
BOSE et PAL (2006)	192	48					9:4:1
BRABAZON et KEENAN (2004)	128	50	[2-6]:[1-4]:1	SEXTON et al. (2003)	1 139	215	[21-37]:[4-14]:1
			[4-12]:[1-4]:1	SHIN et al. (2005)	Var.	Var.	9:9:1

BROCKETT et al. (1994)	145	49	8:3:1	TAM et KIANG (1990)	118	44	19:1
CHARALAMBOUS et al. (2000)	164	114	[5-28]:[0-3]:1				19:1:1
CHARITOU et al. (2004)	50	52	3:4:1	TAM et KIANG (1992)	118	44	19:1
DORSEY et al. (1995)	200				118	40	19:10:1
	372			TAN et DIHARDJO (2001)	1 495	715	17:5:1
	497			TANG et CHI (2005)			[2-32]:[1-65]:1
	370			TUNG et al. (2004)	620	2 483	9:?:?:?:?
FAN et PALANISWAMI (2000)	174						9:10:1
FANNING et COGGER (1994)	150	230	3:6:7:1	TYREE et LONG (1996)	146		4:3:1
			3:2:1		56		?:2:1
GOSS et RAMCHANDANI (1995)	40		4:Var.:1	WALLRAFEN et al. (1996)	667	3029	[10-73]:[1-3]-1
HUANG et al. (1994)			12:20:01				73:[20-26]:[3-20]: [20-31]:[20-30]:1
JO et al. (1997)			[9-20]:[1-41]:1	WEST et al. (2005)	230	50	5:[2-5]:1
KIM et al. (2003)	100	60	[7-14]:10:1	WILSON et al. (1995)	108	8	18:13:3
KIM et HAN (2003)	309	232		WILSON et SHARDA (1994)	Var	Var	5:10:2
KIM et MCLEOD (1999)	55		7:14:1	Wu et al. (2007)	88	538	17:7:1
KOTSIANTIS et al. (2005)	150					534	
KUMAR et al. (1997)	76	76	7:10:1	YANG et al. (1999)	44	40	8:?:1
LAITINEN et KANKAANPAA (1999)	38	38	3:[5-15]:1				8:?:?:1
LEE (2004)	69	36	5:9:1	YANG et HARRISON (2002)			
			8:17:1	YIM et MITCHELL (2002)	100	46	[3-7]:[1-3]:1
LEE et al. (1996)	66	100	[7:8]:[7:18]:2				[1-6]:[1-4]:1
	96	70		YIM et MITCHELL (2005)	121	29	[3-5]:[2-4]:1
	126	40		ZHANG et al. (1999)	176	64	6:[1-15]:1
LEE et al. (2005)	Var	Var	5:[1-10]:1	ZURADA et al. (1998)	204	141	9:[12-30]:2
							9:[12-30]:4

« Var. » signifiant variable, différentes tailles ayant été testées

Le formalisme employé ici pour décrire les architectures des réseaux est le suivant :

- le « : » symbolise la présence d'une couche ;
- le « ? » indique que le paramètre n'a pas été indiqué par le ou les auteurs ;
- la première valeur désigne le nombre d'entrées ;
- la dernière valeur indique le nombre de neurones de sortie ;
- la ou les valeurs intermédiaires correspondent au nombre de neurones par couche ;
- les valeurs indiquées entre « [] » désignent l'intervalle à l'intérieur duquel les auteurs ont fait varier un paramètre : soit le nombre d'entrée, soit le nombre de neurones cachés.

Exemples :

Agarwal (1999) emploie une architecture 4:8:1. Elle comporte ainsi 2 couches avec 4 entrées, 8 neurones cachés et 1 neurone de sortie.

Atiya (2001) emploie une architecture [5-6]:2:1. Ici, le nombre d'entrées varie entre 5 et 6, selon le modèle.

Sexton et al. (2003) emploient une architecture [21-37]:[4-14]:1. Le nombre d'entrées varie entre 21 et 37 et le nombre de neurones cachés varie entre 4 et 14, selon le modèle.

Berg (2005) emploie une architecture 23:?:1. Elle possède 23 entrées, un nombre de neurones cachés inconnu et 1 sortie.

Tung et al. (2004) emploient une architecture 9:?:?:?:?. Elle possède 9 entrées et 4 couches comportant chacune un nombre de neurones inconnu.

Le tableau 23 montre que l'on est donc en présence d'architectures assez simples, mais quand même quelque peu surdimensionnées compte tenu de la taille des échantillons d'apprentissage, qui est en moyenne inférieure à 200, si l'on ne tient pas compte des quelques études où ceux-ci comptent plus de 1 000 observations.

La tableau 24 indique qu'un nombre non négligeable d'auteurs n'a effectué aucun contrôle de la topologie, la taille de la couche cachée ayant été déterminée dans ces cas-là de manière arbitraire, parfois au moyen de « *rules of thumb* » comme certains le disent, donc par l'intermédiaire de règles totalement empiriques et dénuées de tout fondement théorique, qui postulent l'existence d'une relation entre le nombre d'entrées et le nombre de neurones cachés, indépendamment de toute analyse portant sur un éventuel surajustement. D'ailleurs, seuls deux auteurs indiquent avoir procédé à une régularisation des poids grâce à un *weight decay*, mais un *decay* unique, applicable à l'ensemble de ces paramètres. Pour le reste, on observe, lorsque c'est précisé, que deux fonctions d'activation sont plus particulièrement utilisées, la fonction logistique et la tangente hyperbolique, et que la fonction de coût est plutôt du type « écart quadratique moyen » que du type « entropie croisée ». Pour le reste, trop peu d'indications sont données pour pouvoir en tirer quelques commentaires.

Tableau 24 : Modalités de paramétrage des réseaux de type MLP employés pour la construction de modèles de prévision de la défaillance

Études	Algorithmes	Contrôle de la topologie	Pas variable	Weight Decay	Validation en cours d'app.	Utilisation d'un moment	Fonction d'erreur	Fonction d'activation	Biais
AGARWAL (1999)	MLP-BP	?	Oui	?	?	Oui	?	?	?
AGARWAL et al. (2001)	MLP-BP	Non	Oui	?	?	?	?	?	?
ALTMAN (1994)	MLP-BP	Oui	?	?	Oui	Oui	?	?	?
ANANDARAJAN et al. (2001)	MLP-BP	?	?	?	?	?	EQM	Log.	?
	MLP-AG	?	-	-	?	-	EQM	Log.	?
ATIYA (2001)	MLP-BP	?	?	?	?	?	?	?	?
BACK et al. (1994)	MLP-BP	Oui	?	?	Oui	Oui	EQM	Log.	Oui
BACK et al. (1996 A)	MLP	?	?	?	?	?	?	?	?
BACK et al. (1996 B)	MLP-BP	?	?	?	?	?	?	?	?
BACK et al. (1997)	MLP-AG	?	-	-	?	-	?	?	?
BAEK et CHO (2003)	MLP-BP	Oui	?	?	?	?	?	?	?
BARNIV et al. (1997)	MLP-BP	?	Oui	?	?	?	EQM	?	?
BELL et al. (1990)	MLP-NCB	?	?	?	?	?	?	T. Hy.	Oui
BERG (2005)	MLP	?	?	Oui	?	?	Entropie	Log.	?

BORITZ et KENNEDY (1995)	MLP-BP	Oui	?	?	?	?	?	?	?
	MLP-FLBP	Oui	?	?	?	?	?	?	?
	MLP-PBP	Oui	?	?	?	?	?	?	?
	MLP-PCBP	Oui	?	?	?	?	?	?	?
	MLP-OET	Oui	-	-	-	-	-	-	-
BOSE et PAL (2006)	MLP-BP	?	?	?	?	?	?	?	?
BRABAZON et KEENAN (2004)	MLP-AG	Oui	-	-	?	-	?	Log. + T. Hy.	?
BROCKETT et al. (1994)	MLP-BP	Oui	?	?	?	?	?	Log.	?
CHARALAMBOUS et al. (2000)	MLP-GC	Oui	-	Oui	Oui	-	?	T. Hy. + Log	Non
CHARITOU et al. (2004)	MLP-GC	?	-	?	?	-	?	T. Hy. + Log +	Oui
DORSEY et al. (1995)	MLP-AG	?	-	-	?	-	EQM	Log.	?
FAN et PALANISWAMI (2000)	MLP-BP	?	?	?	?	?	?	?	?
FANNING et COGGER (1994)	MLP-BP	Oui	?	?	?	?	?	Log.	?
GOSS et RAM-CHANDANI (1995)	MLP-BP	Oui	Oui	?	Non	Non	?	?	?
HUANG et al. (1994)	MLP-AG	?	-	-	?	-	?	?	?
JO et al. (1997)	MLP-BP	Oui	?	?	?	Oui	?	Log.	Oui
KIM et al. (2003)	MLP-AG	Oui	-	-	?	-	?	?	?
KIM et HAN (2003)	MLP-BP	?	?	?	?	?	?	?	?
KIM et MCLEOD (1999)	MLP-AG	?	?	?	?	?	?	?	?
KOTSIANTIS et al. (2005)	MLP-RBF	?	-	-	-	-	-	Gaus.	-
KUMAR et al. (1997)	MLP-BP	Oui	?	?	?	?	v	Log.	?
LAITINEN et KANKAANPAA (1999)	MLP-BP	?	?	?	?	?	?	?	?
LEE (2004)	MLP-BP	Oui	?	?	?	?	EQM	?	?
LEE et al. (1996)	MLP-BP	Oui	?	?	?	?	?	?	?
LEE et al. (2005)	MLP-LM	Oui	-	?	?	-	EQM	Log.	Oui
LESHNO et SPECTOR (1996)	MLP-BP	?	?	?	?	?	?	?	Oui
	MLP- ?	?	?	?	?	?	?	?	?
LI et GUPTA (2002)	MLP-AG	Oui	-	-	?	-	?	?	?
MIN et LEE (2004)	MLP-BP	Oui	?	?	?	Oui	?	T. Hy.	?
MIN et LEE (2005)	MLP-BP	Oui	?	?	?	?	?	T. Hy.	?
ODOM et SHARDA (1990)	MLP-BP	Oui							
PENDHARKAR (2005)	MLP-BP	Non	?	?	?	?	?	?	?
	MLP-AG	Non	-	-	?	-	?	?	?
PIRAMUTHU et al. (1998)	MLP-BP	Oui	?	?	?	?	?	?	?
POMPE et BILDER-BEEK (2005 A)	MLP-BP	?	?	?	?	?	?	?	?

POMPE et BILDER-BEEK (2005 B)	MLP-BP	?	?	?	?	?	?	?	?
POMPE et FEELDERS (1997)	MLP-BP	Oui	?	Oui	?	?	Entropie	Log.	Oui
RAHIMIAN et al. (1993)	MLP-BP	Non	?	?	?	Oui	?	?	?
SALCHENBERGER et al. (1992)	MLP-BP	Oui	?	?	?	Oui	?	Log.	?
SEN et al. (2004)	MLP-BP	Oui	?	?	Oui	Non	?	Log.	?
SERRANO-CINCA (1997)	MLP-BP	Non	?	?	Non	?	EQM	Lin. + Log.	?
SEXTON et al. (2003)	MLP-AG	Oui	-	-	?	-	?	?	?
SHIN et al. (2005)	MLP-BP	?	?	?	?	?	?	?	?
TAM et KIANG (1990)	MLP-BP	Oui	?	?	?	?	?	?	?
TAM et KIANG (1992)	MLP-BP	Oui	?	Oui	Non	Oui	EQM	Log.	Non
TAN et DIHARDJO (2001)	MLP-BP	Oui	Non	Non	Oui	Oui	EQM	Log.	Non
TANG et CHI (2005)	MLP-BP	Oui	?	?	?	?	EQM	?	?
TUNG et al. (2004)	MLP-BP	?	?	?	?	?	?	?	?
TYREE et LONG (1996)	MLP-BP	Oui	?	?	?	Oui	?	T. Hy.	?
WALLRAFEN et al. (1996)	MLP-AG	Oui	-	-	Oui	-	?	?	?
WEST et al. (2005)	MLP-BP	Oui	?	?	?	?	?	T. Hy.	?
WILSON et al. (1995)	MLP-BP	Non	?	?	Non	Oui	EQM	Log.	Non
WILSON et SHARDA (1994)	MLP-BP	Non	?	?	?	?	?	?	?
WU et al. (2007)	MLP-BP	Non	?	?	?	?	?	T. Hy. + Li.	?
YANG et al. (1999)	MLP-BP	?	?	?	?	?	?	?	?
YANG et HARRISON (2002)	MLP-BP	?	?	?	?	?	?	?	?
YIM et MITCHELL (2002)	MLP-BP	Oui	?	?	?	Oui	?	?	?
YIM et MITCHELL (2005)	MLP-BP	Oui	?	?	?	Oui	?	?	?
ZHANG et al. (1999)	MLP-GRG2	Oui	?	?	?	?	?	Log.	Oui
ZURADA et al. (1998)	MLP-BP	Oui	?	?	?	?	?	Log.	?

Les différentes colonnes du tableau indiquent les paramètres et les techniques employées au cours de l'apprentissage. Elles s'interprètent ainsi :

- contrôle de la topologie : limitation du nombre de neurones cachés ;
- pas variable : emploi d'un pas adaptatif ;
- Weight Decay : utilisation de cette technique de régularisation ;
- validation en cours d'apprentissage : moment où s'est produit la validation des résultats, lorsque celle-ci a été effectuée ;
- utilisation d'un moment : le gradient de la fonction de coût ayant été calculé la plupart du temps par une simple méthode de descente de gradient, la colonne précise si la formule de modification des poids a été complétée par un terme évitant les oscillations ;
- fonction d'erreur : fonction de coût minimisée (EQM : écart quadratique moyen ; Entropie : entropie croisée) ;
- fonction d'activation : type de fonction employée pour propager l'information de couche en couche (Lin. : linéaire ; Log. : logistique ; T. Hy. : tangente hyperbolique ; Gaus. : fonction gaussienne) ;
- biais : ajout d'un biais au potentiel somatique des neurones ;
- le « ? » signifie que le ou les auteurs n'ont pas donné d'indications et le « - » indique que pour le réseau concerné, le paramètre n'a pas lieu d'être.

Nous n'avons pas mentionné d'autres paramètres ou techniques qui auraient pu être utilisés lors de la phase d'apprentissage, comme les techniques de « pruning » par exemple, car aucune indication n'a été donnée à leur égard dans les différents articles.

Enfin, le tableau 25, précise que les techniques de validation reposent soit sur l'emploi d'un échantillon de test, soit sur une validation croisée, soit sur une combinaison des deux, l'erreur étant alors approchée à partir de répliques de l'échantillon de test, sorte de *bootstrap* particulièrement simplifié. Au demeurant, seules deux études se sont appuyées réellement sur une technique de *bootstrap* : celle de Pendharkar (2005) et celle de Wu et al. (2007)²⁴. Par ailleurs, l'estimation de l'erreur est effectuée assez souvent, conjointement avec l'estimation des paramètres du réseau, et seul le réseau conduisant à l'erreur la plus faible est alors retenu. En ce sens, l'erreur est biaisée favorablement en direction du ou des échantillons de test et représente donc une version optimiste de l'erreur en généralisation. Seules de rares études ont utilisé un troisième échantillon pour une telle estimation.

Tableau 25 : Modalités de validation des modèles de prévision de la défaillance élaborés à partir de réseaux de type MLP

Études	Taille des échantillons		Techniques de validation	Études	Taille des échantillons		Techniques de validation
	App.	Test			App.	Test	
AGARWAL (1999)	80	60	ET * 10 VC	LESHNO et SPECTOR (1996)	44	132	EA : t+0 ET : t+0, t+1, t+2
AGARWAL et al. (2001)	220	115	ET		44	44	EA : t+0 ET : t+0
ALTMAN (1994)	404	404	EA : t-3 ET : t-1		44	44	EA : t+1 ET : t+1
	808	808	EA : t-1 EV : t-1		44	44	EA : t+2 ET : t+2
		302	EV : t-1		132	44	EA : t+0, t+1, t+2 ET : t+1

²⁴ Personne n'emploie, par exemple, une technique comme le *bagging* (Breiman et al., 1996), technique qui permet d'améliorer sensiblement la précision d'une règle d'induction dans des situations où de petits changements peuvent l'affecter de manière conséquente. Cette technique consiste à *booster* l'échantillon d'origine, puis à construire autant de prédicteurs que d'échantillons *boosters*, et enfin à utiliser tous ces prédicteurs pour réaliser une prévision. Dans le domaine de la classification par exemple, la prévision finale est effectuée par un vote majoritaire ; la classe d'appartenance prévue est donc celle qui a recueilli la fréquence de réponses la plus élevée. Beaucoup d'auteurs y ont vu une technique permettant de stabiliser la variance d'un classifieur ou d'un régresseur, puisque Efron et Tibshirani (1995), notamment, ont montré que le *bootstrap* disposait déjà de cette propriété. Grandvallet (2001, 2004) a montré que le *bagging* ne réduit pas toujours la variance, et peut même l'augmenter (même s'il n'y a pas d'accord général sur ce que devrait être la décomposition entre biais et variance d'un problème de classification, avec toutes les définitions qui retiennent une décomposition additive où la variance représente la part variable de l'erreur de prévision, les résultats obtenus par Grandvallet montrent que la variance peut augmenter avec le *bagging*), mais néanmoins produit asymptotiquement une sorte de lissage des estimations conduisant à de meilleurs résultats que ceux obtenus avec d'autres méthodes. À quoi alors attribuer ces performances ? Pour Grandvallet, au cours d'une sélection de variables, le *bagging* est particulièrement bénéfique lorsque les observations les plus influentes sont perturbées. Pour lui, la stabilité n'est en fait pas liée à la variance mais à la présence d'observations qui ont un rôle plus prononcé que d'autres dans l'échantillon employé. Et le *bagging* permet précisément de stabiliser une sélection en égalisant les rôles respectifs. *A contrario*, le *bagging* peut être inutile lorsque toutes les observations ont le même poids sur le phénomène étudié, voire même négatif, lorsque ce dernier peut tirer partie de différences possibles entre observations. Dans un bon nombre de cas, les observations particulièrement influentes correspondent aux *outliers* et leur pondération par le *bagging* conduit à diminuer la variance. Donc, si une procédure est instable en raison de la présence d'observations atypiques, alors effectivement le *bagging* stabilise sa variance. Il s'agit donc d'une technique qui améliore la robustesse vis-à-vis des *outliers* en aplatissant les perturbations qu'ils peuvent engendrer, car ils n'apparaissent que dans 67 % des échantillons (le *bootstrap* étant une technique de rééchantillonnage avec remise, chaque observation a une probabilité d'être présente dans un échantillon à peu près égale à 0,67).

ANANDARAJAN et al. (2001)	418	104	ET	132	44	EA : t+0, t+1, t+2 ET : t+2	
ATIYA (2001)	491	669	ET	132	44	EA : t+0, t+1, t+2 ET : t+3	
BACK et al. (1994)	38	38	ET	Li et GUPTA (2002)	180	20	ET
BACK et al. (1996 A)	74		VC		170	30	ET
BACK et al. (1996 B)	36	38	ET	MIN et LEE (2004)	310	103	ET – EV
BACK et al. (1997)	400	170	VC	MIN et LEE (2005)	1 132	378	ET – EV
	200	170	VC	ODOM et SHARDA (1990)	Var	Var	ET
	100	170	VC	PENDHARKAR (2005)	100	100	EA : 1987-1992 * 5 Bootstraps EA : 1992-1995 * 5 Bootstraps
BAEK et CHO (2003)	3 634	4 570	EA : 1994-1996 ET : 1997-2000	PIRAMUTHU et al. (1998)	166	16	ET
BARNIV et al. (1997)	114	123	ET	POMPE et BILDERBEEK (2005 A)	Var.	Var.	16 EA * 10 VC
BELL et al. (1990)	204	1 059	ET	POMPE et BILDERBEEK (2005 B)	Var.	Var.	28 EA * 10 VC
BERG (2005)	60 600	40 400	EA : 1996 ET : 1996	POMPE et FEELDERS (1997)	288	288	ET – VC
BORITZ et KENNEDY (1995)	230	112	ET * 5 VC	RAHIMIAN et al. (1993)	74	55	ET
BOSE et PAL (2006)	192	48	ET * 10 VC	SALCHENBERGER et al. (1992)	200	116	ET
BRABAZON et KEENAN (2004)	128	50	ET	SEN et al. (2004)	160	129	ET
BROCKETT et al. (1994)	145	49	ET – EV	SERRANO-CINCA (1997)	66		LOO
CHARALAMBOUS et al. (2000)	164	114	EA : 1983-1990 ET : 1991-1994	SEXTON et al. (2003)	1 139	215	ET
CHARITOU et al. (2004)	50	52	EA : 1988-1994 ET : 1995-1997	SHIN et al. (2005)	Var.	Var.	ET – EV
DORSEY et al. (1995)	200		VC	TAM et KIANG (1990)	118	44	ET
	372		VC	TAM et KIANG (1992)	118	44	ET – LOO
	497		VC		118	40	ET – LOO
	370		VC	TAN et DIHARDJO (2001)	1 495	715	EA : 1989-1990 ET : 1991
FAN et PALANISWAMI (2000)	174		VC	TANG et CHI (2005)			EA : 1995-2001 ET : 2002
FANNING et COGGER (1994)	150	230	EA : 1947-1951 ET : 1961-1965	TUNG et al. (2004)	620	2 483	ET * 5 VC
GOSS et RAMCHANDANI (1995)	40		LOO	TYREE et LONG (1996)	146		VC
HUANG et al. (1994)			Non		56		VC
JO et al. (1997)				WALLRAFEN et al. (1996)	667	3029	ET – EV
KIM et al. (2003)	100	60	ET	WEST et al. (2005)	230	50	ET – EV
KIM et HAN (2003)	309	232	ET – EV	WILSON et al. (1995)	108	8	VC – LOO
KIM et MCLEOD (1999)	55		VC	WILSON et SHARDA (1994)	Var	Var	ET * 20 VC
KOTSIANTIS et al. (2005)	150		VC	WU et al. (2007)	88	538	EA : 1998-2000 ET 2001
KUMAR et al. (1997)	76	76	ET			534	ET 2002
LAITINEN et KANKAANPAA (1999)	38	38	ET		60	28	Bootstrap : 1998- 2000
LEE (2004)	69	36	ET	YANG et al. (1999)	44	40	ET – EV
LEE et al. (1996)	66	100	EA : 1979-1984 ET : 1985-1992	YANG et HARRISON (2002)			ET

	96	70	EA : 1979-1990 ET : 1991-1992	YIM et MITCHELL (2002)	100	46	EA : 1995-1999 ET 2000-2001
	126	40	EA : 1979-1991 ET : 1992	YIM et MITCHELL (2005)	121	29	ET
LEE et al. (2005)	Var	Var	ET * 4 VC EV * 4 VC	ZHANG et al. (1999)	176	64	ET * 5 VC
				ZURADA et al. (1998)	204	141	EA : 1988 ET : 1989

La colonne « techniques de validation » indique, pour chaque étude, la façon dont les résultats ont été validés et précise, lorsque nécessaire, l'année ou les années où les données ont été prélevées. On distingue 3 techniques essentielles : la validation à l'aide d'un échantillon de test (ET), la validation à l'aide d'une validation croisée (VC) ou d'un Leave-one-out (LOO), ou encore, la validation à l'aide d'un échantillon de test doublée d'un échantillon de validation (EV).

Certains auteurs utilisent parfois un échantillon de test doublé d'une validation croisée. Par exemple, Boritz et Kennedy (1995) emploient un échantillon de test avec lequel ils effectuent 5 validations croisées. Pompe et BilderbeekK (2005 A) utilisent 16 échantillons d'apprentissage, et pour chacun, effectuent 10 validations croisées. D'autres réalisent leur validation sur des données prélevées au cours d'une période différente de celle employée avec l'échantillon d'apprentissage. Ainsi, Baek et Cho (2003) réalisent l'apprentissage avec des données datant de 1994, 1995 et 1996, et la validation avec des données datant de 1997, 1998, 1999 et 2000. Parfois, les auteurs ne précisent pas de dates, mais un intervalle de temps : ainsi Altman (1994) utilise-t-il des données de la période t-3 avant la date de défaillance pour l'apprentissage, et des données de t-1 pour la validation.

2.5 CONCLUSION DU CHAPITRE

Ce chapitre a été l'occasion de détailler les propriétés essentielles des réseaux non bouclés à apprentissage supervisé, que sont l'approximation universelle et la parcimonie qui ont fondé leur légitimité pour traiter des problèmes d'approximation de fonction dans des contextes non linéaires, et notamment dans celui de la prévision de défaillance.

Il a permis de rappeler le mode de fonctionnement de cette famille de réseaux ainsi que les techniques d'optimisation mises en jeu à ce stade, de la descente de gradient simple aux méthodes plus complexes de second ordre.

Il a aussi permis de présenter la problématique de l'apprentissage à base d'observations et ses pièges numériques. Une large part a d'ailleurs été consacrée à l'étude du surajustement ainsi qu'aux stratégies d'évitement de ce phénomène qui permettent de s'assurer que les modèles construits sont aussi parcimonieux que possible.

Ce chapitre a été l'occasion de mettre en évidence la prédominance d'un réseau particulier (le *Perceptron multi-couches*) en tant que technique de modélisation de la défaillance, et de rendre compte des conditions expérimentales de son paramétrage. Il a fait ressortir, à cet égard, les points communs des stratégies d'implémentation, mais aussi les inconvénients de certaines pratiques déployées lors de l'apprentissage des modèles.

Enfin et surtout, il a permis de dégager un ensemble de règles, de techniques et de points essentiels nécessaires au bon apprentissage d'un modèle de prévision sur lesquels nous avons pris appui pour opérer nos propres choix expérimentaux et qui concernent notamment :

- le choix des données des échantillons qui doivent être suffisamment représentatives du phénomène étudié pour ne pas pénaliser la capacité de généralisation du réseau ;
- le choix d'une technique de contrôle du surajustement qui doit prendre en compte le type de tâche à effectuer, classification *versus* régression, et qui doit plutôt s'apparenter à une technique de régularisation en cours d'apprentissage lorsque l'on est confronté à une opération de classification ;
- le choix d'une stratégie de validation des résultats qui gagne à s'appuyer sur des techniques de rééchantillonnage dès lors que l'on est confronté à des données susceptibles de présenter certaines perturbations.

CHAPITRE 3 : MÉTHODES DE SÉLECTION DE VARIABLES

Après avoir étudié la problématique de la prévision de défaillance puis celle de la construction de modèles à l'aide de réseaux de neurones, il reste une question générale à aborder qui est celle du choix des variables. Nous avons souligné en introduction l'inadéquation entre les propriétés des réseaux et les hypothèses de nombreuses méthodes de sélection de variables, notamment paramétriques, comme le caractère arbitraire de certaines pratiques qui s'appuient sur la seule « popularité » d'un indicateur dans la littérature financière pour justifier un choix. Et nous avons rappelé que dans le domaine de la prévision de défaillance, l'absence d'un corpus théorique légitimait que l'on procède par induction pour isoler les prédicteurs d'un modèle.

Dès lors, puisque les méthodes ou pratiques traditionnelles apparaissent comme étant inadaptées à la création de modèles neuronaux, quelles alternatives peut-on envisager ? C'est la question à laquelle ce chapitre est consacré.

Pour ce faire, nous présenterons d'abord la problématique de la sélection au travers des difficultés liées à l'aspect combinatoire de la question et à l'étendue de l'espace de recherche qui en découle.

Nous détaillerons ensuite les particularités des méthodes employées pour ce type de tâche et nous dresserons un panorama de chacune de leurs principales composantes. Nous rendrons ainsi compte des procédures de recherche envisageables permettant de naviguer dans l'espace des descripteurs, des critères d'arrêt servant à interrompre les recherches et des critères d'évaluation susceptibles d'apprécier la ou les solutions trouvées.

Enfin, nous analyserons les spécificités de la sélection avec les réseaux de neurones, en soulignant les inconvénients posés par les méthodes traditionnelles, et notamment les méthodes paramétriques, qui expliquent pourquoi elles sont inutilisables dans leur contexte. Nous présenterons alors les méthodes les plus employées avec les réseaux et nous terminerons avec une question propre à bien des techniques de sélection, qui est leur extrême sensibilité aux conditions expérimentales initiales. En effet, le résultat d'une sélection est souvent très dépendant de la structure des données, donc très instable. C'est pourquoi, il est souhaitable d'utiliser, en parallèle de la sélection, des procédures particulières pour tenter de régulariser les résultats.

Nous ne reviendrons pas ici, à l'instar de ce que nous nous avons fait au cours des chapitres 1 et 2, sur la façon dont les auteurs de modèles de défaillance ont procédé pour sélectionner leurs variables, puisque ceci a déjà été effectué précédemment.

3.1 PROBLÉMATIQUE DE LA SÉLECTION

Choisir des prédicteurs parmi un ensemble de variables candidates est l'un des préalables à la construction de tout modèle. D'abord pour leur économie, car nous avons vu que la parcimonie est une condition essentielle de la généralisation. Ensuite, pour leur performance, car toutes les variables n'y concourent pas de manière similaire et égale. Certaines n'apportent que peu d'information, voire pas d'information du tout, d'autres sont entachées de bruit, d'autres encore sont redondantes donc susceptibles de donner plus de poids à une dimension du phénomène étudié qu'elle ne le mérite. La sélection revient donc à trouver un sous ensemble pertinent eu égard au problème posé, composé d'éléments aussi indépendants que possibles et en nombre suffisant pour expliquer convenablement le processus à modéliser.

La véritable difficulté tient au caractère souvent non monotone et conjoncturel du problème de sélection. En effet, le meilleur sous ensemble de p variables ne contient pas forcément le meilleur sous ensemble de q variables, avec $q < p$. Si tel était le cas, il suffirait d'effectuer une recherche séquentielle par ajout ou élimination successif de variables pour trouver aisément une solution. Mais ça ne l'est pas dans bien des situations. De plus, ce meilleur sous-ensemble n'existe pas en soi, mais dépend de la méthode employée pour construire un modèle, même si en règle générale les deux étapes sont traitées séparément.

Face à ce caractère non monotone, seule une exploration exhaustive de toutes les combinaisons possibles permettrait effectivement de découvrir le ou les meilleurs sous-ensembles. Mais l'explosion combinatoire qui en résulte, même pour un nombre assez faible de variables, rend cette procédure totalement inenvisageable. C'est pourquoi, la plupart des méthodes de sélection reposent sur des heuristiques qui parcourent de manière limitée l'espace de toutes les combinaisons possibles.

Ces heuristiques sont le plus souvent décrites au travers de leurs principaux composants (Langley, 1994 ; Dash et Liu, 1997 ; Leray, 1998) qui sont au nombre de trois :

- une procédure de recherche (génération) qui explore l'espace des combinaisons et génère les ensembles candidats à la sélection ;
- un critère d'évaluation des solutions afin de pouvoir comparer différents sous-ensembles et retenir celui ou ceux présentant les meilleures propriétés ;
- un critère d'arrêt afin d'interrompre la recherche et déterminer l'ensemble à retenir.

L'ensemble du processus se présente comme indiqué sur le schéma suivant²⁵ :

Figure 3 : Processus de sélection de variables

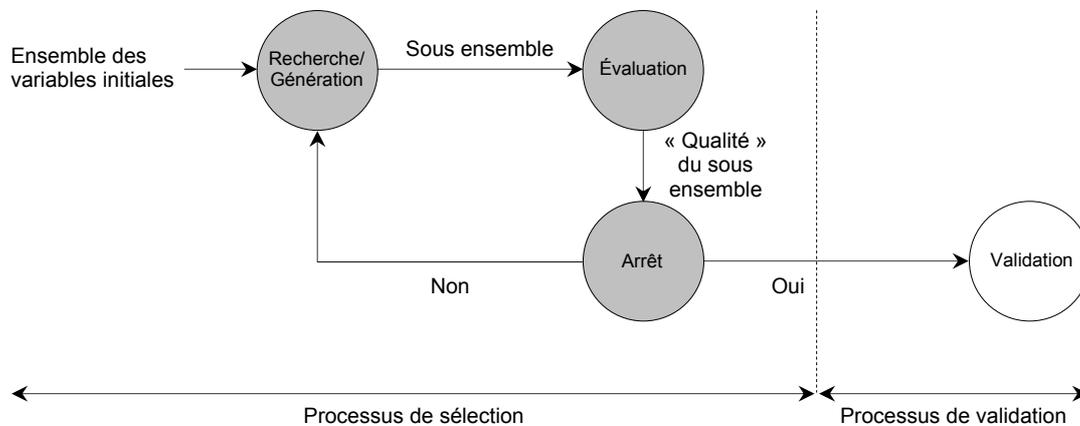


Schéma tiré de Dash et Liu (1997), p. 134

3.2 CARACTÉRISTIQUES DES MÉTHODES

3.2.1 PROCÉDURE DE RECHERCHE

Comme nous l'avons souligné, toutes les situations ne sont pas propices à une recherche dans tout l'espace des combinaisons de variables. Lorsque c'est le cas, donc quand le critère d'évaluation n'est pas régi par une fonction monotone du nombre de variables sélectionnées et quand le nombre de combinaisons est trop important (il est en effet nécessaire d'explorer $2^n - 1$ ensembles pour trouver le meilleur parmi n variables), on préfère limiter la recherche à un sous ensemble. En revanche, quand l'hypothèse de monotonie est respectée, on peut envisager une exploration plus étendue. Différentes typologies de méthodes ont été proposées dans la littérature. Dash et Liu (1997) comme Liu et Yu (2005) proposent une partition en trois : méthodes « complètes », méthodes « séquentielles » et méthodes « aléatoires ».

3.2.1.1 MÉTHODES COMPLÈTES

Ce type de méthode garantit de trouver la solution optimale sous réserve que le critère d'évaluation soit monotone. La recherche est dite « complète » et non « exhaustive », car elle ne parcourt pas l'ensemble des combinaisons, mais emploie différentes heuristiques pour réduire l'espace de

²⁵ La validation ne fait pas à proprement parler des techniques de sélection de variables. C'est pourquoi le schéma indique une frontière entre les deux processus. Mais elle fait partie de la problématique générale de sélection, de la même manière qu'elle fait partie de la problématique de l'apprentissage. Là aussi, l'évaluation d'une solution ne peut être faite uniquement à partir des données qui ont permis de la découvrir pour les mêmes raisons que celles évoquées précédemment liées au surajustement.

recherche sans pour autant diminuer la probabilité de trouver l'optimum. *Branch and Bound* (Narendra et Fukunaga, 1977), notamment, fait partie de cette famille. Des variantes de cette méthode ont été proposées pour tenir compte des problèmes où le critère n'est plus monotone, comme *Approximate Monotonic Branch and Bound*, mais cette fois, le caractère optimal n'est plus garanti. Certaines expériences ont néanmoins montré que *Branch and Bound* fonctionnait bien dans des situations où le critère n'était pas monotone (Jain et Zongker, 1997).

3.2.1.2 MÉTHODES SÉQUENTIELLES

Ces méthodes sont employées pour relâcher l'hypothèse très forte que *Branch and Bound*, notamment, fait peser sur le critère d'évaluation. On les distingue des autres par leur façon de parcourir l'espace : certaines débutent sans aucune variable, puis les ajoutent une à une (processus de type *forward*), d'autres démarrent avec toutes les variables, puis les éliminent une à une (*backward*). Elles sont simples et rapides d'emploi, mais présentent un caractère vraiment sous-optimal, car non seulement elles ne parcourent qu'une zone limitée de l'espace, mais en plus elles ne permettent à aucun moment de revenir en arrière, toute variable retenue ne peut alors plus être éliminée (*nesting*). Aussi, pour atténuer les effets de ce côté sous optimal, certaines méthodes alternent-elles les deux processus précédents. L'algorithme *Plus l – Take Away r* en est une illustration parfaite : à chaque étape, il commence par élargir le sous ensemble en répétant l fois une procédure *forward*, puis tente de le réduire en répétant r fois une procédure *backward*. D'autres méthodes dites flottantes qui dérivent de *Plus l – Take Away r*, utilisent aussi une telle mécanique *forward-backward*, mais en y ajoutant une contrainte sur le critère d'évaluation (Pudil et al., 1994) : ainsi SFFS (*Sequential Floating Forward Selection*) par exemple, applique-t-elle après chaque étape *forward*, autant de *backward* que le sous-ensemble de variables considérées à chaque itération permet d'améliorer le critère d'évaluation. Avec SBFS (*Sequential Floating Backward Selection*) la logique est la même, mais les étapes *forward* et *backward* sont inversées. Leur intérêt réside dans la détermination dynamique du nombre de variables ajoutées ou éliminées à chaque étape, selon la valeur que prend le critère d'évaluation, donc dans l'absence de paramètre à fixer *a priori*. De toutes les méthodes sous-optimales, ces dernières sont considérées comme les plus efficaces (Jain et Zongker, 1997). Une version adaptative de SFFS et SBFS existe aussi (Somol et al., 1999), qui permet de faire varier, là aussi selon les valeurs prises par le critère d'évaluation, le nombre d'étapes *forward* et *backward*. Bien évidemment, ce raffinement permet de mieux approcher une solution optimale, mais au prix d'un accroissement notable des temps de calcul.

3.2.1.3 MÉTHODES ALÉATOIRES

Contrairement aux précédentes qui sont déterministes, c'est-à-dire qui produisent toujours le même résultat quelle que soit l'initialisation opérée, ces méthodes conduisent à des solutions qui dépendent précisément des conditions initiales de calcul. Leur point de départ consiste à choisir au hasard un ensemble de variables, puis à parcourir l'espace selon deux stratégies possibles : soit séquentielle, donc identique à celles présentées précédemment (c'est le cas du recuit simulé – *Simulated Annealing*), soit aléatoire (c'est le cas des algorithmes génétiques, introduits dans ce domaine par Siedlecki et Sklansky, 1988). Dans ce dernier cas, l'aléa introduit dans le cheminement permet de limiter le risque que la méthode ne se retrouve piégée dans un minimum local, mais au prix d'un temps de calcul considérablement plus long qu'avec un parcours séquentiel. Malgré leurs propriétés, les algorithmes génétiques offrent des performances comparables aux méthodes traditionnelles, mais leur lourdeur rend leur utilisation souvent rédhibitoire.

3.2.2 CRITÈRE D'ÉVALUATION

Une fois qu'une procédure de recherche a choisi un sous-ensemble de variables, ce dernier doit être évalué eu égard au problème posé. Ainsi, pour une tâche de classification, l'évaluation portera sur sa capacité à bien discriminer les groupes. Il faut donc trouver un critère qui permettra, *in fine*, de comparer différents sous-ensembles et de retenir celui ou ceux présentant les meilleures propriétés. John et al. (1994) proposent de classer les méthodes d'évaluation en deux familles : d'un côté, on trouve celles pour lesquelles l'évaluation repose uniquement sur les caractéristiques intrinsèques des données sans faire intervenir l'algorithme d'induction²⁶ (méthodes de type *filter*), de l'autre, on trouve celles qui reposent sur les performances que l'algorithme d'induction présente en utilisant précisément les variables que l'on tente d'évaluer (méthodes de type *wrapper*). Les critères associés à ces méthodes seront donc, soit indépendants de l'algorithme d'induction, soit dépendants de lui, pour reprendre la terminologie de Liu et Yu (2005).

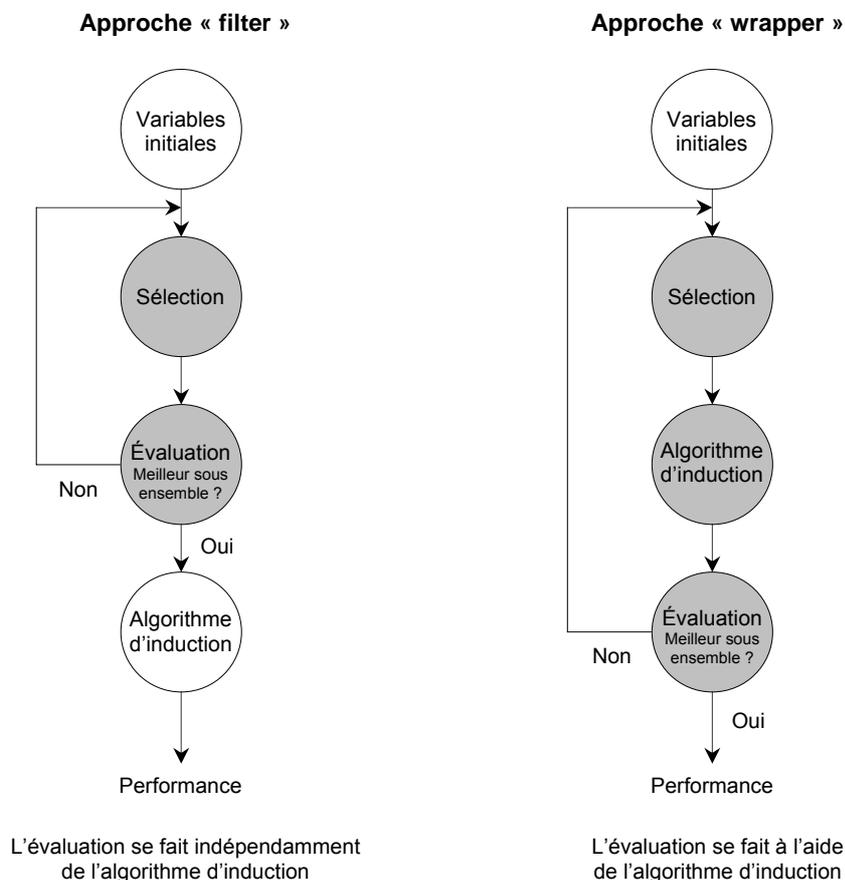
3.2.2.1 CRITÈRE INDÉPENDANT

Les quatre critères indiqués ci-après, sont couramment distingués et associés aux méthodes *filter* et peuvent être employés pour élaborer toute règle de classification :

²⁶ L'algorithme d'induction désigne ici celui qui, *in fine*, emploiera les variables sélectionnées : par exemple, un algorithme de classification (analyse discriminante, réseau de neurones...).

- mesures de distance : elles sont aussi appelées mesures de discrimination ou de divergence et se subdivisent en deux familles. La première tente d'appréhender les distances probabilistes (*Mahalanobis, Battacharyya, divergence*) entre les fonctions de densité ou entre les probabilités *a posteriori* des classes. La variable qui maximise l'une ou l'autre de ces distances sera retenue. Les secondes s'appuient sur des transformations des matrices de covariance inter et intra-classes – *Lambda de Wilks* – et sont bien adaptées aux situations où les classes ont une même structure de covariance et ne requièrent pas d'hypothèses probabilistes ;
- mesures d'information : elles évaluent le gain en information procuré par une variable par rapport à une autre à l'aide des probabilités *a posteriori* – *entropie, dérivée* ;
- mesures de dépendance : elles sont aussi appelées mesures de corrélation ou de similarité et quantifient le pouvoir de prédiction d'une variable à partir d'une autre variable, donc le degré de redondance – *corrélacion, information mutuelle, ambiguïté* (Semani et al., 2004) ;
- mesures de consistance : elles permettent de découvrir le plus petit sous-ensemble de variables disposant du même pouvoir de séparation que l'ensemble de départ.

Figure 4 : Deux approches de la sélection de variables



3.2.2.2 CRITÈRE DÉPENDANT

Mais avec certains autres algorithmes, les critères indépendants ne conduisent pas toujours à un résultat optimal, ou alors sont difficilement utilisables. C'est pourquoi, on se sert de l'algorithme d'induction (*i.e.* le classifieur) : chaque ensemble de variables sélectionné à un moment donné est alors évalué en fonction de l'erreur en généralisation du modèle correspondant, la meilleure solution étant alors celle qui procure l'erreur la plus faible.

Ce type de méthode (*wrapper*), trouve sa légitimité notamment dans les travaux de John et al. (1994) et Kohavi et John (1997, 1998). Selon John et al. (1994), une variable sélectionnée ne doit pas seulement dépendre de ce que l'on souhaite décrire (concept cible : *target concept*), mais aussi de l'algorithme d'induction employé. Cet argument dérive directement de la définition retenue de la pertinence d'une variable : est fortement (*strongly*) pertinente une variable qui est indispensable, c'est-à-dire qui ne peut pas être supprimée sans qu'une perte de précision de la prévision ne se produise, et faiblement (*weakly*) pertinente si elle est en mesure de contribuer, parfois, à la précision de la prévision. Les *wrappers* donnent souvent de bons résultats parce que les sous-ensembles obtenus sont adaptés à la méthode d'induction. Cependant, ils requièrent des temps de calculs élevés car l'estimation de l'erreur s'opère grâce à des procédures de validation, comme le *bootstrap*, qui sont particulièrement gourmandes en traitements (Kohavi, 1995).

Le tableau 26 ci-après fournit un bon aperçu des caractéristiques respectives de tous les critères passés en revue. Il indique la « généralité de la solution » trouvée, c'est-à-dire si le sous-ensemble optimal peut être utilisé quel que soit l'algorithme d'induction envisagé ou non, le temps de calcul pour parvenir à une solution, et la précision que l'on peut espérer du sous-ensemble optimal lorsqu'employé avec un algorithme d'induction.

Tableau 26 : Comparaison des critères dévaluation

Critère	Généralité de la solution	Temps de calcul	Précision
Distance	Oui	Faible	Dépend de l'algorithme d'induction
Information	Oui	Faible	Dépend de l'algorithme d'induction
Dépendance	Oui	Faible	Dépend de l'algorithme d'induction
Consistance	Oui	Modéré	Dépend de l'algorithme d'induction
Erreur	Non	Élevé	Élevée

Tiré de Dash et Liu (1997), p. 137.

3.2.3 CRITÈRE D'ARRÊT

L'évaluation de sous-ensembles pouvant se poursuivre indéfiniment ou du moins, jusqu'à ce que toutes les combinaisons aient été épuisées, l'algorithme doit être interrompu à un moment donné. Pour cela, on utilise un critère qui, lorsque satisfait, met un terme aux calculs. Les critères d'arrêt correspondent généralement soit à une limite prédéfinie fixant le terme du processus, comme un nombre maximum de variables ou d'itérations, soit à l'absence de variation significative d'un résultat attendu après ajout ou retrait d'un élément, soit à l'obtention d'une capacité prédictive satisfaisante, c'est-à-dire d'une erreur en généralisation ayant atteint un seuil considéré comme acceptable. Ils s'appuient souvent sur des heuristiques de calculs et parfois sur des tests statistiques. Ce dernier cas correspond notamment aux méthodes qui évaluent différents sous-ensembles selon leur pertinence eu égard au critère d'évaluation, lorsque l'on connaît la distribution empirique de ce dernier ou lorsqu'on peut lui associer une statistique. Dans une telle situation, il est alors possible de tester statistiquement la pertinence d'une variable et de ne retenir que celle qui répond à la condition posée. Une fois que plus aucune variable n'est considérée comme pertinente, la recherche s'interrompt²⁷.

3.3 SÉLECTION AVEC LES RÉSEAUX DE NEURONES

Avec un réseau de neurones, certaines méthodes ou critères que l'on vient de voir ne font pas partie des pratiques de sélection, parce qu'ils sont difficiles à mettre en œuvre et posent des problèmes d'implémentation, ou bien parce qu'ils ne sont pas utilisables en l'état. Il est en ainsi des méthodes paramétriques de sélection de variables. Comme le souligne Leray (1998), les méthodes qui sont indépendantes des modèles et qui reposent sur « des hypothèses soit de dépendance linéaire entre les entrées et les sorties, soit de corrélation linéaire entre les variables » ne sont pas adaptées aux réseaux, précisément parce que ces derniers sont des modèles non linéaires.

²⁷ Par exemple, lorsque l'on emploie un *Lambda de Wilks* pour sélectionner un ensemble de variables, dans le cas d'un problème de classification, on cherche le sous-ensemble qui minimise ce critère. Pour y parvenir, avec un algorithme de recherche *forward*, la séquence est la suivante :

1°) on choisit un ensemble de variables ;

2°) pour chaque variable i de cet ensemble, on calcule le *Lambda de Wilks* associé à chaque sous ensemble contenant toutes les variables sauf la variable i ;

3°) une fois tous les sous ensembles passés en revue, on supprime la variable qui minimise le *Lambda de Wilks* ;

4°) on recommence à l'étape 2 tant que le critère d'arrêt n'est pas satisfait.

On utilise généralement un F de *Fisher* comme critère d'arrêt : la suppression des variables se poursuit tant que le F calculé est inférieur à un seuil fixé *a priori*, ce qui statistiquement signifie que la variable supprimée n'apporte pas de contribution significative à la discrimination au sens de la loi de *Fisher* et du niveau de confiance associé à cette statistique.

Il en est ainsi, de la même manière, des tests statistiques pouvant présider aux choix des variables. Même si leur intérêt est indéniable pour asseoir la validité d'un modèle, ils sont étudiés depuis trop peu de temps pour que l'on dispose de résultats suffisants pour en tirer quelque généralité et ils renvoient toujours à des calculs assez lourds, car ils nécessitent souvent l'estimation d'une variance d'échantillonnage d'une mesure à partir de procédures toujours très longues de validation croisée ou de *bootstrap*²⁸. Ces tests portent soit sur les paramètres du réseau (poids), comme White (1989), cité par Leray (1998) et Refenes et Zapranis (1999), qui propose un test fondé sur leur distribution, soit sur une mesure de pertinence des variables, comme Dodier et Henze (2004) qui proposent un test dérivé d'un test de *Wald*, ou Refenes et Zapranis (1999) et La Rocca et Perna (2003, 2005) qui étudient des procédures de test couplées à une telle mesure, soit sur la topologie du réseau, comme Rivals et Personnaz (2003), qui suggèrent une procédure de sélection d'architecture à partir d'un test de *Fisher* afin de réduire la structure interne du réseau (nombre de neurones cachés) et par construction, le nombre d'entrées, donc de variables à utiliser. Il est en de même des techniques de rééchantillonnage qui permettent d'affranchir la sélection de certains pièges numériques. L'espace de recherche des paramètres d'un réseau étant parsemé de minima locaux, toute mesure d'évaluation est donc contrainte par le minimum vers lequel le réseau a convergé. Ceci nécessite alors de répliquer la sélection à de multiples reprises et de moyenniser les résultats et de fait, interdit l'emploi de toute procédure de recherche sophistiquée.

Maintenant, les réseaux ont des spécificités qui ne sont partagées par aucune méthode de régression ou de classification et qui sont à l'origine de techniques de sélection qui leur sont propres. Nous présenterons d'abord les spécificités de ces méthodes au regard des éléments constitutifs de toute procédure de sélection, puis nous détaillerons chaque famille, selon la typologie proposée par Leray (1998).

3.3.1 CARACTÉRISTIQUES GÉNÉRALES DES MÉTHODES

3.3.1.1 PROCÉDURE DE RECHERCHE

Les procédures de recherche sont soit des procédures qui portent uniquement sur les variables, soit des procédures où cette recherche est conditionnée par une recherche plus générale d'une

²⁸ Pour tester la significativité d'une variable, il faut estimer la variation d'échantillonnage de la mesure de pertinence retenue et utiliser un test d'hypothèse pour s'assurer que la variable est vraiment pertinente, ce qui est simple avec un modèle linéaire, mais beaucoup moins avec un modèle non linéaire. En effet, la significativité repose sur une estimation de l'erreur-type de la mesure et compte tenu du contexte non linéaire, cette estimation ne peut se faire qu'à l'aide de techniques non paramétriques qui reposent notamment sur des méthodes de rééchantillonnage.

architecture optimale. Si l'on s'intéresse uniquement aux variables, la recherche s'effectue plutôt sur un mode *backward*, l'argument essentiel justifiant cette pratique consistant à dire que c'est le seul moyen de véritablement tenir compte des interactions entre l'ensemble des variables avant toute élimination (Guyon et Elisseeff, 2003), ce qu'interdit une recherche *forward*. En revanche, lorsque la sélection découle d'un processus d'optimisation de la taille d'un réseau, les méthodes emploient plutôt une recherche *forward*, car elles sont calées sur des algorithmes constructifs où la taille du réseau augmente régulièrement par ajout de neurones jusqu'à ce qu'un seuil soit atteint. Dans un cas comme dans l'autre, se pose la question du réapprentissage du réseau. En effet, après un ajout ou une élimination, la nature des relations entre les variables restantes risque de n'être plus la même qu'à l'étape précédente et seul un réapprentissage laisse espérer que le réseau pourra réellement en tenir compte.

3.3.1.2 CRITÈRE D'ÉVALUATION

Les critères d'évaluation sont des mesures individuelles de pertinence de chaque variable fournies par des algorithmes qui utilisent l'information contenue dans les paramètres du réseau. C'est ici que les méthodes de sélection pour les réseaux divergent véritablement des autres. Il en existe un nombre assez considérable que Leray (1998) regroupe en trois familles :

- mesures d'*ordre zéro* : elles s'appuient sur les valeurs absolues des poids et sont issues soit d'un calcul tenant compte directement de leur amplitude, soit de tests statistiques ;
- mesures d'*ordre un* : elles emploient les dérivées premières des paramètres du réseau qui peuvent s'interpréter comme une mesure de la contribution des variables aux sorties ;
- mesures d'*ordre deux* : elles utilisent les dérivées secondes des paramètres et dérivent directement des techniques d'élagage (*pruning*) de poids.

3.3.1.3 CRITÈRE D'ARRÊT

Le calcul du critère d'arrêt relève aussi d'heuristiques de nature très diverse, faisant intervenir un seuil posé sur le critère d'évaluation ou un classement des variables selon un critère comme l'erreur en généralisation ou encore, un test statistique. Dans ce dernier cas, par exemple, Van de Laar et al. (1997) effectuent un test sur les erreurs. La logique qui préside à ce test vise à évaluer plusieurs modèles au regard de leur erreur et à ne retenir que le moins complexe parmi tous ceux dont l'erreur est statistiquement proche de la plus petite erreur

envisageable. Dans un même ordre d'idées, Rivals et Personnaz (2003) proposent un procédure qui vise d'abord à évaluer, pour un problème donné, l'architecture de plus grande complexité admissible, eu égard au conditionnement numérique du réseau, puis à élaguer ce dernier tant que l'estimation de son erreur demeure statistiquement proche du modèle complet. Leur hypothèse (dont une critique se trouve dans Dreyfus et al., 2004), consiste à considérer que la valeur prise par le nombre de conditionnement de la matrice Jacobienne des paramètres caractérise le degré de complexité du réseau pour le problème considéré, et qu'au-delà d'un certain seuil, qu'ils fixent à 10^8 , le réseau a dépassé sa limite de complexité.

3.3.2 MESURES D'ORDRE « ZÉRO »

Les mesures de cette catégorie portent sur la valeur absolue des paramètres du réseau. Intuitivement, on peut supposer que l'importance d'une variable est proportionnelle à l'amplitude des poids qui la relie aux neurones de la première couche cachée. Il suffit alors, à l'image de la régression linéaire, d'effectuer un test sur les poids pour savoir si une variable est significative ou non, comme on effectue un test de nullité des coefficients de régression. Les poids considérés alors comme nul peuvent être éliminés et toute variable qui n'est alors plus reliée au réseau (ou reliée seulement par un nombre minimal de poids) peut l'être aussi. Cette stratégie n'est pas directement transposable aux réseaux, car la non linéarité des fonctions d'activation interdit que l'on interprète un poids comme un coefficient de régression. Donc l'amplitude d'un poids n'est pas un bon indicateur de pertinence d'une variable (Refenes et Zapranis, 1999), même si celle-ci est parfois employée (Lee et al., 1993 ; Messer et Kittler, 1998). Dans ce cas, ce sont les plus faibles en valeur absolue qui sont éliminés.

Devant l'absence de relations théoriques et empiriques entre la magnitude des poids et l'intérêt qu'une variable peut présenter pour un problème donné, de nombreux auteurs ont cherché d'autres méthodes faisant toujours intervenir les poids, mais selon des modes assez différents. Par exemple, Cibas et al. (1994 A), Devena (1994) ou El Deredy et Branston (1995) proposent d'estimer la pertinence d'une variable par un calcul sur la variance des poids : l'hypothèse sous-jacente à leurs méthodes considère que l'absence de variation est révélatrice d'une variable aux effets constants, donc non porteurs d'information et qu'à ce titre, elle peut être éliminée. White (1989), cité par Leray (1998) et Refenes et Zapranis (1999), lui, suggère un test statistique pour éliminer des poids, qui est fondé sur une propriété asymptotique de leur distribution et qui peut servir à supprimer ensuite des variables (celles dont les poids sont du point de vue du test les moins significatifs).

À côté de ces mesures statistiques, on trouve une série de critères qui découlent plutôt d'heuristiques de calculs et qui s'appuient aussi sur les poids, mais où la pertinence d'une variable, appelée aussi « sensibilité » est définie selon des combinaisons pondérant tel ensemble de poids par tel autre. Contrairement aux calculs précédents, ceux-ci font intervenir explicitement la structure du réseau en même temps que la valeur des poids. La logique d'élimination est la même qu'avec les méthodes où seuls les plus petits poids sont supprimés, mais le calcul est plus raffiné. Nous en donnons deux exemples, ci-après, tirés de Yacoub et Bennani (1997) et de Howes (1999). Pour un réseau comportant n entrées, une couche cachée de h neurones et un neurone de sortie, et avec w_{ji} représentant le poids reliant l'entrée i au neurone j de la couche cachée, et w_j le poids reliant le neurone j au neurone de sortie, la pertinence P d'une variable i est définie par :

$$P_i = \sum_{j=1}^h \left(\frac{|w_{ji}|}{\sum_{k=1}^n |w_{jk}|} \frac{|w_j|}{\sum_{j=1}^h |w_j|} \right) \quad (\text{Yacoub et Bennani, 1997})$$

$$P_i = \frac{\sum_{j=1}^h \left(\frac{|w_{ji}|}{\sum_{k=1}^n |w_{jk}|} |w_j| \right)}{\sum_{j=1}^h |w_j|} \quad (\text{Howes, 1999})$$

La mesure revient ici à évaluer chaque chemin conduisant une entrée à la sortie relativement à l'ensemble des chemins, un chemin étant représenté par un produit de poids. Les variables éliminées sont donc celles pour lesquelles la magnitude des chemins qui les concernent est la plus faible. Ces critères nécessitent une recherche *backward*, ainsi qu'un réapprentissage après toute élimination et peuvent être couplés avec une évaluation de l'erreur en généralisation pour déterminer l'arrêt de la sélection.

3.3.3 MESURES D'ORDRE « UN »

Avec ces mesures, on cherche à appréhender la contribution des entrées aux variations des sorties, étant entendu qu'une variable non pertinente est une variable qui contribue faiblement à la formation de l'erreur, donc dont le retrait est supposé être négligeable de ce point de vue.

Un premier ensemble de méthodes utilise la variation de l'erreur calculée sur un échantillon de données lorsque l'on remplace une variable par une valeur fixe. C'est ce que proposent notamment Utans et Moody (1991) et Moody et Utans (1992) en remplaçant chaque variable par sa moyenne empirique, de manière à annuler son effet. La contribution des variables est ensuite calculée en effectuant la différence entre l'erreur produite en présence de la variable considérée et celle

produite avec sa moyenne. Les variables sont ensuite triées par ordre décroissant selon cette différence et éliminées une à une. La méthode initiale proposée par Moody et Utans ne prévoit qu'une seule évaluation de la pertinence des variables : une fois le classement effectué, un réseau est entraîné sur chaque ensemble de variables et celui présentant la plus faible erreur en généralisation est retenu. Moody (1994), faisant remarquer que cette méthode ne tient pas compte des corrélations possibles, propose d'orthogonaliser les entrées pour s'en affranchir. Leray (1998) suggère plutôt de recalculer la pertinence de chaque variable après toute élimination, au prix de temps de calcul accrus bien évidemment. Une variante de cette méthode revient à calculer l'erreur après retrait de chaque variable et à éliminer celle qui concoure le moins à sa formation, puis à recommencer jusqu'à ce qu'un critère d'arrêt soit satisfait. Setiono et Liu (1997) utilisent un principe similaire, mais cette fois la méthode repose sur un calcul de l'erreur après mise à zéro de tous les poids reliant chaque variable à la couche cachée : la variable éliminée est celle pour laquelle la mise à zéro des poids dont elle dépend a engendré la plus faible variation d'erreur.

Un second ensemble estime la pertinence d'une variable à partir d'un calcul faisant intervenir la dérivée de la sortie du réseau par rapport à cette variable, moyennée sur l'ensemble de données de l'échantillon. Ces calculs pouvant produire aussi bien des valeurs positives que négatives, si l'on n'annule pas l'effet du signe, la dérivée moyenne des sorties risque de ne pas refléter l'effet des variables. C'est pourquoi, on utilise soit son carré, soit sa valeur absolue. La pertinence s'exprime alors de cette manière si l'on retient sa valeur absolue :

$$P_i = \frac{1}{N} \sum_{j=1}^N \left| \frac{\partial y_i}{\partial x_{ji}} \right|$$

avec x_j représentant une variable, y la sortie du réseau calculée par un seul neurone et N la taille de l'échantillon.

Cette expression est à la base d'un grand nombre de mesures de sensibilité, dont nous donnons quelques exemples ci-après tirés de Refenes et Zapanis (1999).

$$P_i = \frac{1}{N} \sum_{j=1}^N \left| \frac{\partial y_i}{\partial x_{ji}} \right| \left| \frac{x_{ji}}{y_i} \right|, \quad y_i \neq 0$$

$$P_i = \left[\frac{1}{N} \sum_{j=1}^N \left(\frac{\partial y_i}{\partial x_{ji}} - \frac{1}{N} \sum_{j=1}^N \left(\frac{\partial y_i}{\partial x_{ji}} \right) \right)^2 \right]^{\frac{1}{2}}$$

$$P_i = \frac{1}{N} \sum_{j=1}^N \frac{\left(\frac{\partial y_i}{\partial x_{ji}} \right)^2}{\sum_{j=1}^N \left(\frac{\partial y_i}{\partial x_{ji}} \right)^2}$$

Leray (1998) indique d'autres mesures et notamment certaines qui permettent de ne pas tenir compte d'observations dites « pathologiques », donc aberrantes, à partir d'un *triming* de la distribution des dérivées des sorties par rapport aux entrées ou qui ne s'intéressent qu'aux observations proches des frontières entre les classes.

Comme précédemment, ces critères sont employés avec des méthodes *backward* et impliquent un réapprentissage dès qu'une variable est supprimée.

3.3.4 MESURES D'ORDRE « DEUX »

Cette dernière famille de mesures dérive directement des méthode d'élagage permettant de simplifier la structure d'un réseau en supprimant des connexions entre neurones : une connexion est coupée si sa pertinence ou « salience » est faible. Dans le cas de la sélection de variables, une variable est supprimée si la pertinence globale des connexions qui relie l'entrée correspondante à la première couche cachée du réseau est la plus faible.

La pertinence d'une variable se définit ainsi :

$$P_i = \sum_{j \in \text{fan out}(i)} P(w_j)$$

avec $\text{fan out}(i)$ représentant tous les poids partant de la variable i .

Cibas et al. (1994 B) proposent une méthode de calcul, *Optimal Cell Damage*, qui dérive de la méthode de Le Cun et al. (1990), *Optimal Brain Damage* pour y parvenir. Avec cette dernière, la pertinence d'un poids est définie comme la variation de l'erreur résultant de la suppression de ce poids, le poids ayant la plus faible pertinence étant celui dont l'élimination provoque la plus faible augmentation de l'erreur.

Le Cun et al. (1990) ont montré que si la Hessienne peut être approximée par une matrice diagonale et si la fonction de coût est localement quadratique, alors la salience d'un poids w_j peut être approximée localement au voisinage d'un minimum w_0 par :

$$P(w_j) = \frac{1}{2} H_{jj} w_j^2 = \frac{1}{2} \frac{\partial^2 J}{\partial w_j^2} w_j^2$$

avec H représentant la Hessienne, J la fonction de coût et w la matrice des poids.

La pertinence d'une variable s'exprime alors ainsi :

$$P_i = \sum_{j \in \text{fan out}(i)} \frac{1}{2} \frac{\partial^2 J}{\partial w_j^2} w_j^2$$

Optimal Brain Damage requiert que le réseau ait atteint un minimum local afin que l'estimation de l'erreur soit correcte. Cibas et al. (1994 B) proposent, pour y parvenir, de recourir à une méthode de type *backward* et, une fois ce minimum atteint, suggèrent de supprimer toutes les variables dont la salience est inférieure à un seuil défini *a priori*. La recherche se poursuit à partir du dernier ensemble de variables, en procédant à l'identique tant que l'erreur estimée sur un échantillon de test ne chute pas. Leray (1998) propose une variante de cette méthode à partir d'une critique du critère d'évaluation et du critère d'arrêt. En effet, Leray (1998) considère que de tels critères peuvent avoir des effets trop brutaux ; si la valeur du seuil est trop élevée, certaines variables d'intérêt risquent d'être éliminées à tort et si l'arrêt est prématuré, la recherche peut s'interrompre avant que le réseau n'ait atteint un minimum local. Il propose alors d'éliminer les variables les unes après les autres, et de continuer l'élagage jusqu'à ce qu'il n'y en ait plus en jeu. Ensuite, à partir de la configuration qui permet d'obtenir la plus faible erreur, il recherche tous les ensembles de variables qui conduisent à une erreur statistiquement proche de la plus faible, au sens d'un test de *Fisher*, et choisit parmi ceux-ci, l'ensemble contenant le moins de variables.

Stahlberger and Riedmiller (1997) offrent une méthode de suppression de variables à peu près similaire à celle qui vient d'être décrite, *Unit-Optimal Brain Surgeon*, mais elle s'appuie sur la technique d'élagage de Hassibi et Storck (1993), *Optimal Brain Surgeon*. Contrairement à *Optimal Brain Damage*, *Optimal Brain Surgeon* ne travaille plus à partir d'une approximation de la Hessienne, mais à partir d'un calcul direct des dérivées secondes. Ceci évite les simplifications fortes engendrées par la méthode de Le Cun et al. (1990), mais au prix d'un accroissement important des temps de calculs. *Unit-Optimal Brain Surgeon* tire donc partie de cette propriété mais en offrant une amélioration nette par rapport à *Optimal Brain Surgeon*. Cette fois, au lieu d'éliminer les poids un à un, la méthode permet de les détruire par groupes, ce qui représente un avantage indéniable pour sélectionner des variables.

Nous avons vu qu'*Optimal Brain Damage* nécessite que l'algorithme ait convergé vers un minimum local. Cependant, dans bien des cas, il est utile d'arrêter l'apprentissage avant toute convergence pour des raisons liées au surajustement. Or, lorsque l'apprentissage est arrêté prématurément, *Optimal Brain Damage* donne des résultats erronés. C'est pourquoi, Tresp et al. (1997) proposent une variante de cette méthode, *Early Cell Damage*, qui supporte que l'entraînement puisse être interrompu à tout moment. *Early Cell Damage* étant une méthode

d'élagage, Leray (1998) propose d'employer cette mesure de pertinence des poids pour en déduire celle d'une variable et de la coupler à un test statistique comme critère d'arrêt.

La pertinence d'un poids s'évalue, avec *Early Cell Damage*, ainsi (Tresp et al., 1997) :

$$P(w_j) = \frac{1}{2} \frac{\partial^2 J}{\partial w_j^2} w_j^2 - \frac{\partial J}{\partial w_i} w_i + \frac{1}{2} \left(\frac{\partial^2 J}{\partial w_i^2} \right)^{-1} \left(\frac{\partial J}{\partial w_i} \right)^2$$

Et Leray (1998) en déduit directement la mesure de pertinence suivante :

$$P_i = \sum_{j \in \text{fan out}(i)} \frac{1}{2} \frac{\partial^2 J}{\partial w_j^2} w_j^2 - \frac{\partial J}{\partial w_i} w_i + \frac{1}{2} \left(\frac{\partial^2 J}{\partial w_i^2} \right)^{-1} \left(\frac{\partial J}{\partial w_i} \right)^2$$

3.3.5 ÉVALUATION DES MÉTHODES ET STABILITÉ DES RÉSULTATS

Les conditions expérimentales dictent dans bien des situations les performances des méthodes de sélection. Donc, rien ne permet d'établir *a priori* une hiérarchie entre elles. Leray (1998) a d'ailleurs réalisé une étude comparative faisant intervenir des critères appartenant aux trois familles présentées précédemment, ainsi que deux critères qui ne sont pas propres aux réseaux (une mesure de distance entre les classes – *Lambda de Wilks* – et une mesure d'information – *entropie croisée*) et a étudié leur comportement dans différentes situations. Le test a d'abord porté sur des problématiques de classification avec, d'un côté, un problème où la frontière entre les classes est non linéaire, de l'autre, avec un problème décrit par de nombreuses variables inutiles. Ensuite, il s'est porté sur l'analyse des critères d'arrêt face à un jeu de données bien connu et constitué par les vagues de Breiman (Breiman et al., 1984). Enfin, le test a étudié le comportement des méthodes face à des ensembles de variables très corrélées.

Les conclusions de Leray (1998) vont bien dans le sens d'une absence de hiérarchie, aucun critère ne permettant, en général, d'obtenir de meilleurs résultats que d'autres. Mais, elles permettent de dégager quelques lignes directrices. En premier lieu, les critères fondés sur « des hypothèses de linéarité et de distribution unimodale sont mal adaptés aux autres problèmes »²⁹. En deuxième lieu, l'absence d'indépendance entre les variables plaide pour un réapprentissage du réseau après chaque suppression, afin qu'il tienne compte des changements affectant les relations entre les variables retenues. En troisième lieu, un critère d'arrêt qui ne fait que tenir compte des variations de performance de l'algorithme d'induction peut conduire à un arrêt par trop prématuré. C'est pourquoi, Leray (1998) propose d'employer conjointement une estimation de

²⁹ Leray (1998), p. 87.

l'erreur et un test statistique pour choisir le plus petit ensemble de variables parmi ceux présentant les meilleures performances.

Maintenant, toutes ces méthodes ont quand même une propriété commune : elles sont particulièrement sensibles à de petites perturbations des conditions expérimentales. Si les données sont redondantes, différents sous-ensembles de variables aux propriétés de prévision identiques peuvent être obtenus selon les conditions initiales de l'algorithme, l'ajout ou la suppression d'une variable ou encore, selon la présence de bruit. Donc la sélection peut être particulièrement dépendante de l'échantillon sur lequel elle est opérée et aboutir à des résultats très différents après de très petits changements dans sa composition. Pour stabiliser ce processus de choix, de nombreux auteurs recommandent d'employer des techniques de rééchantillonnage, avec lesquelles la sélection sera répétée sur différentes répliques de l'échantillon initial, de type *bootstrap*, de façon à lisser les résultats (Sauerbrei et Schumacher, 1992 ; Guyon et Elisseeff, 2003 ; Ferguson et al., 2003 ; Zellner, 2004).

3.4 CONCLUSION DU CHAPITRE

Ce chapitre a été l'occasion de cerner les difficultés et les spécificités des procédures de sélection de variables et de dresser un état de l'art des réponses algorithmiques qui ont été apportées pour résoudre trois questions : de quelle manière parcourir l'espace des variables, dès lors qu'il est illusoire d'explorer l'ensemble des combinaisons, et quelle méthode de recherche adopter ? Comment évaluer une solution trouvée et à l'aide de quel critère ? Comment arrêter une recherche et là aussi, quel critère prendre en compte ?

Ce chapitre a notamment mis en avant l'aspect très heuristique des méthodes associées aux réseaux de neurones en raison des propriétés intrinsèques des techniques de sélection mais aussi du caractère non linéaire des méthodes neuronales. Il a apporté ainsi des réponses aux questions précédentes.

Ce chapitre a aussi permis d'établir un certain nombre de règles pour la plupart assez empiriques, mais qui peuvent être utilisées avec profit, parmi lesquelles on retrouve celles présentées ci-après :

- les techniques paramétriques traditionnelles ne sont pas compatibles avec les propriétés des réseaux et il est pertinent d'employer d'autres alternatives ;
- les méthodes de recherche *backward* sont préférables aux *forward* et les méthodes de type *stepwise* sont particulièrement pénalisantes en raison des temps de calcul qu'elles imposent ;

- rien ne permet de déterminer *a priori* les performances des critères d'évaluation présentés et il est judicieux, pour un problème donné, d'en employer plusieurs et de comparer leurs performances sur la base de la capacité de généralisation des modèles produits ;
- le réapprentissage du réseau est nécessaire après toute suppression ou ajout d'une variable, malgré le coût additionnel en temps de calcul ;
- toute sélection risque de conduire à des solutions par trop tributaires de l'échantillon sur lequel elle est opérée et si l'on a affaire à des données présentant des perturbations, alors la sélection doit s'appuyer sur une technique de rééchantillonnage pour tenter de les atténuer.

PARTIE 2 : LES MÉTHODES DE SÉLECTION DE VARIABLES ET LES MODÈLES NEURONAUX DE DÉFAILLANCE

Après avoir passé en revue les questions relatives à la prévision de défaillance, notamment à l'aide de méthodes statistiques, puis détaillé les modalités d'élaboration de modèles neuronaux et de sélection des variables pour ces modèles, nous allons maintenant présenter les expérimentations que nous avons conduites pour répondre à la problématique posée en introduction.

Nous commencerons par exposer nos choix méthodologiques, puis nous présenterons les résultats auxquels nous avons aboutis.

Les choix méthodologiques portent sur trois séries de techniques : bien évidemment celles directement impliquées dans la construction des modèles (analyse discriminante, réseau de neurones...), mais aussi celles qui encadrent leur processus de création et de validation (*bagging*) et celles qui ont présidé aux choix des échantillons et des variables expérimentales.

Les résultats, eux, ont trait à la structure des modèles trouvés, à leur précision et à la représentation de la défaillance qu'ils renvoient. Nous analyserons donc les points communs entre tous les modèles pour tenter de dégager des éléments stables³⁰, des sortes de « formes fortes » qui pourraient révéler l'existence de variables explicatives de la défaillance qui ne sont pas que conjoncturelles, alors que l'examen de bon nombre de modèles tend plutôt à suggérer qu'en ce domaine, il y a fort peu de constante. Nous analyserons aussi la performance des modèles afin d'établir une hiérarchie de tous les couples « méthodes de prévision – technique de sélection » testés et savoir notamment, ce que l'emploi de méthodes de sélection adaptées aux réseaux peut apporter dans ce domaine. Nous étudierons, pour finir, le sens des modèles au regard de la défaillance et ce qu'ils permettent de mettre en lumière.

³⁰ Liu et al. (2002) faisaient d'ailleurs remarquer que les ensembles de variables sélectionnés par différentes méthodes devraient aussi être considérés, donc étudiés de près.

CHAPITRE 4 : MÉTHODOLOGIE

Notre problématique impliquant une analyse comparative, nous avons donc dû choisir un ensemble de méthodes de prévision et aussi de sélection de variables et pour ces dernières, retenir à chaque fois une procédure de recherche, un critère d'évaluation et un critère d'arrêt. Mais les choix ne sont pas arrêtés là. Une fois les méthodes retenues, il a fallu effectuer certains paramétrages, comme avec le réseau de neurones, puis définir et adopter un protocole encadrant la sélection puis la validation des modèles, pour s'affranchir autant que possible des pièges liés à tout emploi d'un échantillon. Enfin, nous avons dû sélectionner des observations, puis des variables. Ce sont tous ces choix que nous allons expliquer dans ce qui suit.

4.1 MÉTHODES DE CONSTRUCTION DES MODÈLES

Le tableau 27 présenté ci-après dresse un panorama assez complet des méthodes de classification ou de régression employées depuis les années 60 pour bâtir des modèles de défaillance. Il indique aussi, au travers de leur fréquence d'utilisation, quelles sont les techniques les plus souvent sollicitées. Si l'on fait abstraction des réseaux, on remarque que l'analyse discriminante et la régression logistique se détachent nettement des autres. Le tableau 22, p. 85, lui, donne des détails sur les réseaux employés dans un contexte similaire et comme nous l'avons déjà souligné, montre la prééminence des *Perceptrons multi-couches*. Le tableau 24, p. 110, montre les choix de paramétrages réalisés avec ces réseaux. Nous nous sommes donc appuyés sur toutes ces données pour effectuer nos propres choix, en retenant les méthodes ou critères les plus souvent employés, voire recommandés si nous ne disposons d'aucune information quant à leur utilisation par les auteurs des modèles de défaillance, afin que nos expérimentations se rapprochent le plus possible de ce qui a été fait jusqu'à présent.

Nous avons donc retenu comme méthodes de construction de modèles, l'analyse discriminante et la régression logistique et pour le réseau, nous avons choisi une descente de gradient simple comme technique d'optimisation, le calcul du gradient s'appuyant sur l'algorithme de rétropropagation de l'erreur de Rumelhart et al. (1986). Nous avons également inclus dans la formule de modification des poids du réseau, un terme de moment. Nous avons retenu une tangente hyperbolique comme fonction d'activation, adopté l'erreur des moindres carrés comme fonction de coût, ajouté un biais sur chaque couche et employé un terme de régularisation, là aussi propre à chaque couche.

Tableau 27 : Synthèse des principaux travaux relatifs à la prévision de défaillance à l'aide de techniques de classification et de régression

Méthodes	Études
AD	ALTMAN (1968) – EDMISTER (1972) – DEAKIN (1972) – ALTMAN (1973) – BLUM (1974) – MOYER (1977) – NORTON et SMITH (1979) – DAMBOLENA et KHOURY (1980) – SHARMA et MAHAJAN (1980) – BOOTH (1983) – TAFFLER (1983) – EL HENNAWY et MORRIS (1983) – KARELS et PRAKASH (1987) – LAITINEN (1991) – ALTMAN et al. (1995) – SHIRATA (1998) – GRICE et INGRAM (2001) – KURUPPU et al. (2003)
AD – AC	FRYDMAN et al. (1985) – SUNG et al. (1999)
AD – AC – GSS	WANG (2004)
AD – ADI	PINCHES et TRIESCHMANN (1974)
AD – ADNP – MDP	ZOPOUNIDIS et DOUMPOS (2002)
AD – ADQ	ALTMAN et al. (1977) – PANTALONE et PLATT (1987)
AD – ADQ – PR	GOMBOLA et al. (1987) – PACEY et PHAM (1990)
AD – ADQ – PR – RL	HER et CHOE (1999)
AD – AG	VARETTO (1998)
AD – AG – PO	NANDA et PENDHARKAR (2001)
AD – PO	GUPTA et al. (1990)
AD – RL	MENSAH (1984) – AZIZ et al. (1988) – MCGURR et DE VANEY (1998)
AD – RL – ADH	DOUMPOS et ZOPOUNIDIS (1999)
AD – RL – ADNP	BARNIV et HERSHBARGER (1990)
AD – RL – EG	DIMITRAS et al. (1999)
AD – RL – EG – AC	BEYNON et PEEL (2001)
AD – RL – EG – RNL	TSAKONAS (2006)
AD – RL – MS	SHUMWAY (2001)
AD – RL – PR	GENTRY et al. (1985) – LENNOX (1999)
AD – RL – PR – ADNP	BARNIV et RAVEH (1989)
AD – RL – PR – BR – LM – ADNP – EGB	BARNIV et McDONALD (1992)
AD – RL – RLM	PEEL et PEEL (1987)
ADRL – SVM	VAN GESTEL et al. (2006)
AD – SC	THEODOSSIOU (1993)
AD – SVM	HÄRDLE et al. (2005)
AD – TC	LINDSAY et CAMPBELL (1996)
RL	OHLSON (1980) – ZAVGREN (1985) – GILBERT et al. (1990) – KEASEY et WATSON (1987) – KEASEY et MCGUINNESS (1990) – PLATT et PLATT (1990) – FLAGG et al. (1991) – PLATT et al. (1994) – CYBINSKI (1995) – LUSSIER (1995) – HWANG et al. (1997) – LAITINEN et LAITINEN (1998) – KANE et al. (1998) – MOSSMAN et al. (1998) – RICHARDSON et al. (1998) – TIRAPAT et NITTAYAGASETWTAT (1999) – NAM et JINN (2000) – LAITINEN et LAITINEN (2000) – HUNTER et ISACHENKOVA (2001) – WESTGAARD et VAN DER WIJST (2001) – BARNIV et al. (2002) – PLATT et PLATT (2002) – FOREMAN (2003) GODLEWSKI (2003) – SHARMA et ISELIN (2003) – LIN et PIESSE (2004) – RAHMAN et al. (2004) – SUN et al. (2005) – WEISS et CAPKUN (2005)
RL – MRJ	LAITINEN et LAITINEN (1995)
RL – MS	LEE et URRUTIA (1996)
RL – PG	LENSBERG et al. (2006)
RL – PR	GRICE et DUGAN (2003)
RL – PR – LPM	THEODOSSIOU (1991)
RL – PR – RSP – RLC – PRC	BARANOFF et al. (2000)
RL – RT	KOLARI et al. (2002) – LANINE et VANDER VENNET (2005)

RL – SE	ELMER et BOROWSKI (1988)
RLM	LAU (1987)
RLMX – RLM	JONES et HENSHER (2004)
RLQI	TSENG (2005)
RN	BACK et al. (1994) – RAHIMIAN et al. (1993) – WALLRAFEN et al. (1996) – PIRAMUTHU et al. (1998) – CHARALAMBOUS et al. (2000) – ATIYA (2001) – BAEK et CHO (2003) – KIM et al. (2003) – SEXTON et al. (2003) – VIEIRA et al. (2003) – TANG et CHI (2005) – WEST et al. (2005)
RN – AC	DORSEY et al. (1995)
RN – AC – AD – RL	KIM et MCLEOD (1999)
RN – AC – AG	KIM et HAN (2003)
RN – AC – SVM – kNN – RS	KOTSIANTIS et al. (2005)
RN – AD	ODOM et SHARDA (1990) – COATS et FANT (1993) – BROCKETT et al. (1994) – WILSON et SHARDA (1994) – LACHER et al. (1995) – LESHNO et SPECTOR (1996) – TYREE et LONG (1996) – YANG et al. (1999) – ANANDARAJAN et al. (2001) – BRABAZON et KEENAN (2004) – POMPE et BILDERBEEK (2005 A) – POMPE et BILDERBEEK (2005 B)
RN – AD – AC	LEE et al. (1996) – POMPE et FEELDERS (1997) – PENDHARKAR (2005)
RN – AD – ADAG – ADQ – PR – RL – RLAG	BERG (2005)
RN – AD – ADQ – LVQ	KIVILUOTO (1998)
RN – AD – AE	JO et al. (1997)
RN – AD – kNN	KUMAR et al. (1997)
RN – AD – QV – SVM	FAN et PALANISWAMI (2000)
RN – AD – RA	AGARWAL (1999)
RN – AD – RL	ALTMAN et al. (1994) – GOSS et RAMCHANDANI (1995) – BACK et al. (1996 A) – BACK et al. (1996 B) – ETHERIDGE et SRIRAM (1997) – SERRANO-CINCA (1997) – YANG et HARRISON (2002) – YIM et MITCHELL (2002)
RN – AD – RL – AC – MS	LAITINEN et KANKAANPAA (1999)
RN – AD – RL – AC – kNN	TAM et KIANG (1992)
RN – AD – RL – kNN	TAM et KIANG (1990) – HUANG et al. (1994)
RN – AD – RL – PR – ADNP – ADQ	BORITZ et KENNEDY (1995)
RN – AD – RL – PR – SVM	WU et al. (2007)
RN – AD – RL – RS	BACK et al. (1997)
RN – AD – RL – RSP	LEE (2004)
RN – AD – RL – SVM	MIN et LEE (2005)
RN – AD – RLM	AGARWAL et al. (2001)
RN – AD – SVM	BOSE et PAL (2006)
RN – ADQ – RL	LEE et al. (2005)
RN – MS	TUNG et al. (2004)
RN – PR	TAN et DIHARDJO (2001)
RN – RL	BELL et al. (1990) – SALCHENBERGER (1992) – WILSON et al. (1995) – ZURADA et al. (1998) – ZHANG et al. (1999) – CHARITOU et al. (2004) – SEN et al. (2004)
RN – RL – AC	MCKEE et GREENSTEIN (2000)
RN – RL – AD – PR	YIM et MITCHELL (2005)
RN – RL – MS – PW	FANNING et COGGER (1994)
RN – RL – PR – RLH – PRH – SM	LI et GUPTA (2002)
RN – RL – SVM	MIN et al. (2006)

RN – RLM – ADNP	BARNIV et al. (1997)
RN – SVM	MIN et LEE (2004) – SHIN et al. (2005)
RN – TF	ALAM et al. (2000)
AC – MB	SARKAR et SRIRAM (2001)
AC – PR	GENTRY et al. (2002)
ADI	RAMSER et FOSTER (1931) – FITZPATRICK (1932) – WINAKOR et SMITH (1935) – MERWIN (1942) – BEAVER (1966) – BEAVER (1968)
ADQ	SINKEY (1975)
AG	SHIN et LEE (2003)
AG – EG	MCKEE et LENSBERG (2002)
EGR	BRABAZON et O'NEILL (2004)
ED – PL – AC	CIELEN et al. (2004)
MB	PHILOSOPHOV et PHILOSOPHOV (2002)
MDS	NEOPHYTOU et MAR-MOLINERO (2004)
MRJ	WILCOX (1973)
MS	CHAVA et JARROW (2004)
PG – EG – SVM	SALCEDO-SANZ et al. (2005)
PR	ZMIJEWSKI (1984) – GENTRY et al. (1987) – SKOGSVIK (1990)
RLB	MEYER et PIFER (1970) – DREZNER et al. (2001)
SVM	SALCEDO-SANZ et al. (2004) – HUA et al. (2007)

AC : Arbre de Classification (ou partitionnement récursif)	MS : Modèle de Survie (Hazard Model)
AD : Analyse Discriminante linéaire	PG : Programmation Génétique (Genetic Programming)
ADAG : Analyse Discriminante Additive Généralisée	PL : Programmation Linéaire (Linear Programming)
ADH : Analyse Discriminante Hiérarchique	PO : Programmation par Objectif (Linear Goal Programming)
ADI : Analyse Dichotomique	PR : Probit
ADNP : Analyse Discriminante Non Paramétrique	PRC : Probit Cubique
ADS : Analyse Discriminante Quadratique	PW : Processus de Wiener (modèle stochastique)
AE : Apprentissage à partir d'Exemples (Case-Based Reasoning)	QV : Quantification Vectorielle (Learning Vector Quantization)
AG : Algorithme Génétique	RA : Réseau Abductif (Abductive Network - Abductive Reasoning)
AP : Ajustement Partiel	RC : Risque de Crédit
BR : Burit	RL : Régression Logistique
ED : Enveloppement de Données (Data Envelopment Analysis)	RLAG : Régression Logistique Additive Généralisée
EG : Ensembles Grossiers (Rough Sets)	RLB : Régression Linéaire Binaire
EGB : Exponentiated Generalized Beta	RLC : Régression Logistique Cubique
EGR : Évolution Grammaticale (mélange d'un algorithme génétique avec de la programmation génétique)	RLM : Régression Logistique Multinomiale
GANN : Generalized Adaptive Neural Network (réseau utilisant un algorithme génétique)	RLMX : Régression Logistique Mixte (Mixed Logit)
GL : Gestion des Liquidités	RLQI : Régression Logistique Quadratique avec Intervalle (Quadratic Interval Logistic Regression)
GSS : Generalized Smoothing Spline	RN : Réseaux de Neurones
IR : Induction de Règles (Rules Induction)	RNL : Réseau de Neurones Logiques (Neural Logic Network)
kNN : k Nearest Neighbour (k plus proches voisins)	RS : Règles Symboliques
LM : Lomit	RSP : Régression Spline
MB : Modèle Bayésien	RT : Reconnaissance de Traits (Trait Recognition)
MDP : Méthode de Désagrégation des Préférences (Preference Disaggregation Analysis)	SC : Somme Cumulative (CUSUM)
MDS : Multidimensional Scaling	SE : Système Expert
MN : Méthode du Noyau (Kernel Method)	SVM : Support Vector Machine (machine à vecteur support)
MPL : Modèle Probabiliste Linéaire (Linear Probability Model)	TC : Théorie du Chaos (modèle élaboré à partir de l'exposant de Lyapunov)
MRJ : Modèle de Ruine du Joueur (Gambler's Ruin Model)	TF : Typologie Floue (Fuzzy Clustering)

Maintenant, un réseau ne s'emploie pas indépendamment d'un paramétrage préalable qui dépend en partie des variables retenues. Or, le réseau doit servir lors des sélections, pour déterminer les variables d'intérêt. Une question s'est donc posée : faut-il que le paramétrage s'effectue pendant le processus de sélection ou en amont ? La première solution, où l'on recherche en même temps une combinaison optimale de paramètres du réseau et des variables est particulièrement longue à mettre en œuvre. La seconde, où l'on fixe *a priori* les paramètres, ne conduit pas nécessairement au choix de la meilleure architecture, mais est bien plus rapide. Dans la littérature, les auteurs fixent parfois, de manière arbitraire, leurs paramètres, comme Back et al. (1997) ou Kim et al. (2003) qui utilisent un nombre constant de neurones cachés. D'autres, introduisent une certaine variabilité, en testant différentes tailles de la couche cachée, mais en nombre réduit, comme Back et al. (1994), Back et al. (1996 B), Sexton et al. (2003) ou Brabazon et Keenan (2004) qui opèrent leur sélection en faisant varier la taille de la couche cachée de quelques unités seulement. D'un point de vue général, lorsqu'un réseau est employé dans un processus de sélection, les paramètres sont fixés *a priori*, pour des raisons liées aux temps de calcul et l'on devine qu'ils l'ont été à la suite de tests dont le contenu est souvent restitué par leurs auteurs de manière édulcorée.

Pour notre part, nous avons aussi préféré fixer *a priori* les paramètres du réseau, car le protocole de sélection retenu repose sur un *bootstrap* des données pour les raisons évoquées au cours du paragraphe 3.3.5, p. 133. Dans un tel contexte, il était en effet inenvisageable de tester de multiples combinaisons de paramètres tout en effectuant les calculs sur un grand nombre d'échantillons. Il a donc été nécessaire de choisir une architecture qui demeurerait fixe tout au long des calculs. Pour y parvenir, nous avons conduit une série de tests préalables pour éviter de déterminer de manière totalement arbitraire des paramètres d'importance comme le pas d'apprentissage ou le terme de régularisation. Nous avons procédé en trois étapes. Tout d'abord, nous avons tiré aléatoirement 50 ensembles de variables parmi celles qui ont été retenues et qui sont décrites plus loin dans ce document. Ensuite, pour chaque ensemble de variables, nous avons fait varier les paramètres suivants : le pas d'apprentissage (entre 0,1 et 0,5 avec un pas de 0,1), le terme de moment (entre 0,5 et 0,9 avec un pas de 0,1), les termes de régularisation des poids (*weight decay*), à raison d'un terme par couche, (entre 10^{-5} et 10^{-2} avec un pas de 10^{-1}) et le nombre de neurones cachés (entre 2 et 10). Le nombre de couches cachées a été limité à 1. Toutes ces valeurs ont été fixées sur la base des pratiques relevées dans notre littérature de

référence³¹, même si rien ne garantit qu'il s'agisse là d'ensembles optimaux ou proches d'un optimum. Elles l'ont également été en fonction du nombre de combinaisons à explorer. Pour chaque combinaison, nous avons estimé l'erreur par validation croisée ($k = 10$)³² sur l'un de nos échantillons (données de 2002)³³ et retenu les paramètres conduisant à la meilleure solution, soit, 4 neurones cachés, 0,4 pour le pas d'apprentissage, 0,4 pour le moment, 10^{-4} et 10^{-3} comme termes de *decay* respectivement entre l'entrée et la couche cachée et entre la couche cachée et la sortie. Enfin, l'apprentissage a été arrêté après 1 000 itérations, l'erreur ne variant quasiment plus au-delà de cette limite et le temps de traitement associé étant acceptable.

4.2 MÉTHODES DE SÉLECTION DES VARIABLES

Les méthodes de sélection des variables ont d'abord été choisies à partir du contenu du tableau 12, p. 38, et du tableau 13, p. 41, qui indiquent les pratiques en la matière des auteurs de modèles de défaillance, ainsi qu'en fonction des méthodes implémentées par SPSS que nous avons utilisées pour certains traitements. Nous avons d'abord retenu une méthode *stepwise*, couplée à un *Lambda de Wilks* comme critère d'évaluation et à un test *F* de *Fisher* comme critère d'arrêt. Nous y avons ajouté deux autres méthodes de type *stepwise* (*forward stepwise* et *backward stepwise*), associées notamment à une statistique de *vraisemblance* comme critère d'évaluation, et à un *Khi²* comme critère d'arrêt. La première méthode est celle généralement retenue en amont d'une analyse discriminante, un test *F* permettant d'ajouter puis d'éliminer des variables tout au long de la procédure, par comparaison à un seuil à définir. Les deux autres le sont avant une régression logistique : la méthode *forward stepwise* procède par ajout de variables avant de revenir en arrière et la méthode *backward stepwise*, opère en sens inverse. L'ajout est déterminé par un test sur les coefficients des variables (score de *Rao*) et l'élimination par un test portant sur un rapport de vraisemblance (*Likelihood Ratio Test*). Nous avons retenu ces deux dernières méthodes, car rien dans la littérature n'indique ce que les auteurs ont privilégié parmi les parcours *stepwise* que l'on peut employer conjointement avec ces derniers critères³⁴.

³¹ Elles ont aussi été fixées en fonction du nombre de combinaisons à explorer pour des raisons tenant à la longueur des temps de traitement. C'est pourquoi nous n'avons retenu que 5 valeurs pour le pas d'apprentissage et 5 pour le moment, alors que des réglages plus fins pouvaient être envisagés.

³² La valeur du « *cut* » retenue pour tous les calculs dont les résultats sont présentés dans ce document est celle qui minimise l'erreur générale, toute classe confondue.

³³ Les échantillons sont décrits un peu plus loin dans ce chapitre.

³⁴ Le détail des algorithmes implémentés dans SPSS auquel nous avons eu recours est indiqué dans SPSS Inc (2004), *SPSS Regressions Models* 13.0, 70 p., et dans SPSS Inc, *Algorithms for Discriminant Analysis*, 18 p. et SPSS Inc, *Algorithms for Logistic Regression*, 14 p.

Dans le registre des méthodes ou critères de sélection employés habituellement avec l'analyse discriminante ou la régression logistique, les trois ensembles retenus correspondent à la quasi-totalité des pratiques. Nous n'en avons donc pas ajouté d'autres.

Ensuite, nous avons choisi pour le réseau une mesure de chacune des deux premières familles présentées, ainsi qu'une mesure faisant intervenir directement une estimation de l'erreur comme critère d'évaluation : les deux premières relèvent des techniques de type « *filter* » et la dernière, de la technique de type « *wrapper* », où le réseau intervient dans l'évaluation :

- mesure d'*ordre zéro* : nous avons privilégié un critère relatif à la structure du réseau plutôt qu'une mesure statistique, car plus simple d'emploi et ayant prouvé son efficacité et nous avons retenu la formule proposée par Yacoub et Bennani (1997), décrite au paragraphe 3.3.2, p. 128 ;
- mesure d'*ordre un* : nous avons utilisé la dérivée moyenne des sorties par rapport aux entrées, décrite au paragraphe 3.3.3, p. 129 ;
- mesure d'*erreur* : l'erreur employée a été calculée à partir d'échantillons de test.

Nous n'avons pas retenu de mesures d'*ordre 2*, d'une part, car nous souhaitions opposer 3 mesures neuronales aux 3 autres décrites précédemment afin de ne pas introduire un déséquilibre dans le nombre de points de comparaison ; d'autre part, car ce type de mesure requiert des calculs extrêmement longs avec les procédures de *bootstrap* que nous avons mises en œuvre.

Avec ces critères, nous avons utilisé, à chaque fois, une recherche *backward* uniquement, plutôt qu'une procédure *forward* ou *stepwise*. En effet, une recherche *forward* présente l'inconvénient de ne pas prendre en compte l'ensemble des relations pouvant exister entre toutes les variables, car se focalisant uniquement sur le sous-ensemble en cours d'analyse, et une *stepwise* pose le problème de la profondeur du retour en arrière et celui des temps de calcul.

La valeur du *cut* ayant servi de critère de classification, c'est-à-dire la valeur permettant de séparer les observations et les affecter à l'une ou l'autre classe, a toujours été la valeur permettant de minimiser l'erreur globale de prévision. Nous aurions pu choisir une autre règle faisant intervenir, par exemple, le coût d'une mauvaise classification. Mais une telle règle aurait complexifié et multiplié grandement les calculs pour un gain vraisemblablement marginal. Pour qu'elle ait un véritable intérêt, il aurait fallu tester différentes hypothèses et donc, toute une série de paramètres de coûts.

La recherche s'effectuant par élimination successive, nous avons opéré un réapprentissage du réseau à chaque étape et la procédure d'élimination a été conduite à son terme, jusqu'à ce qu'il n'y ait plus de variables en jeu. En effet, un arrêt décidé sur la seule base d'une augmentation à un moment donné de l'erreur peut conduire, comme nous l'avons souligné, à un arrêt prématuré. Ainsi, après chaque suppression, l'ensemble des variables restant a été évalué et à la fin, seul celui conduisant à l'erreur la plus faible a été retenu. Le tableau 28 ci-après résume l'ensemble des choix opérés.

Tableau 28 : Techniques de sélection de variables

Critère d'évaluation	Procédure de recherche	Critère d'arrêt
Lambda de Wilks	<i>Stepwise</i>	Fisher
Vraisemblance	<i>Forward Stepwise</i>	Khi ²
Vraisemblance	<i>Backward Stepwise</i>	Khi ²
Mesure Ordre 0	<i>Backward</i>	Erreur
Mesure Ordre 1	<i>Backward</i>	Erreur
Erreur	<i>Backward</i>	Erreur

4.3 PROTOCOLE DE CHOIX DES VARIABLES

Nous avons souligné précédemment le caractère instable des procédures de sélection des variables et l'intérêt de techniques comme le *bootstrap* pour lisser les perturbations pouvant affecter la procédure. Un échantillon *bootstrap* correspond à un échantillon de taille identique à l'échantillon initial et constitué à partir de ce dernier par tirage aléatoire avec remise. Donc, dans chaque échantillon *bootstrap*, certaines données sont reproduites plusieurs fois et d'autres sont absentes. Nous avons donc utilisé cette technique et fait en sorte que chaque échantillon *bootstrap* possède le même nombre d'entreprises saines et d'entreprises défailtantes que dans les échantillons initiaux pour éviter toute distorsion qui pourrait résulter de la sur- ou sous-représentation d'une classe. La sélection des variables a été effectuée sur la base de la procédure décrite par Ferguson et al. (2003) en utilisant 1 000 échantillons *bootstraps*. La procédure a été conduite de la manière suivante : tous les échantillons ont d'abord été parcourus à l'aide de chaque méthode et les occurrences de sélection ont été comptées, puis les variables ont ensuite été classées par ordre de fréquence de sélection et seules celles qui sont apparues dans au moins 70 %³⁵ des cas ont été retenues, dès lors que leurs signes étaient identiques et conformes à ce que l'on pouvait attendre.

³⁵ Ce pourcentage, recommandé par Sauerbrei et Schumacher (1992), est totalement empirique. Nous l'avons conservé en l'état sans faire de tests particuliers.

Cette procédure permet de connaître les variables qui ont *a priori* le plus fort pouvoir de séparation mais elle peut aussi conduire à en éliminer certaines prématurément, s'il existe des paires de variables particulièrement corrélées. De fait, si deux variables qui peuvent être sélectionnées potentiellement sont très corrélées, il se peut que la plupart des ensembles retenus contiennent l'une ou l'autre de ces variables sans pour autant qu'une des deux n'apparaissent dans la majorité des sélections (Ferguson et al., 2003). Pour identifier ces variables en compétition, nous avons calculé les pourcentages de sélection de chaque paire de variables initiales. Les paires dans lesquelles l'une ou l'autre, voire les deux, sont apparues au total dans plus de 90 % des sélections ont alors été considérées comme des paires contenant des variables potentiellement sélectionnables. Pour chaque paire, la variable sélectionnée le plus souvent a été définitivement retenue et ajoutée à celles déterminées précédemment. Et ceci a été effectué pour chaque méthode de sélection retenue et présentée dans le tableau 28, présenté précédemment. Une fois ces ensembles sélectionnés, nous avons réitéré toute la procédure afin de choisir les variables qui seront finalement employées pour les modèles. En effet, les opérations d'ajout post-sélections peuvent conduire à retenir des variables inutiles. C'est pourquoi, il est nécessaire d'opérer une dernière passe. Nous avons introduit une variante dans cette procédure. Ferguson et al. (2003) ne tiennent pas compte du signe des coefficients des variables lors de la comptabilisation des fréquences de choix et il nous est apparu que le fait de sélectionner une variable alors qu'elle possède un signe contraire à ce que l'on en attend n'est pas légitime, car la pertinence d'un prédicteur dépend de sa contribution à l'appréhension du phénomène étudié. Un signe erroné va précisément à l'encontre de cet objectif. Aussi, avons-nous éliminé des décomptes les occurrences de sélection où les variables n'étaient pas pourvues d'un coefficient dont le signe était adéquat.

4.4 PROTOCOLE DE CRÉATION ET DE VALIDATION DES MODÈLES

Une fois les variables retenues, nous avons procédé à la construction des modèles puis à leur test. La procédure employée ici est une procédure de *bagging*. Nous l'avons choisie, car certaines méthodes comme les réseaux sont particulièrement sensibles à la structure de l'échantillon où une petite variation dans ce dernier ne se traduit pas forcément par une petite variation des coefficients du réseau. Il en est de même d'autres méthodes, comme l'analyse discriminante, lorsque l'on emploie un faible nombre de variables. *Bootstraper* les données peut rendre les prévisions plus stables que si elles étaient réalisées, par exemple, par validation croisée.

Afin de ne pas introduire de distorsion dans les résultats de notre étude comparative, nous avons préféré alourdir les calculs pour diminuer autant que faire se peut l'influence de la structure des échantillons sur les résultats. Là encore, les tests préalables que nous avons conduits ont plutôt plaidé en faveur de l'emploi de cette technique. En effet, la distribution des variables financières n'est jamais normale et comporte souvent beaucoup de valeurs aberrantes que l'on ne peut pas éliminer. Comme nous n'avons employé que des ratios financiers, sauf pour une variable et que leur distribution indique la présence de nombreux points atypiques, nous avons retenu le *bagging* pour limiter leur influence, selon les arguments de Grandvallet (2001, 2004) exposés précédemment. La création des modèles s'est effectuée à l'aide d'un premier échantillon de 500 entreprises. La procédure employée, tirée de Breiman (1996), est la suivante :

- l'échantillon a d'abord été divisé aléatoirement en deux : un échantillon d'apprentissage A et un échantillon de test T, le premier contenant 450 entreprises et le second 50 (Breiman utilise au cours de plusieurs expérimentations des échantillons de test dont la taille avoisine les 10 % de l'échantillon d'apprentissage, voire moins) en faisant en sorte que la répartition des entreprises saines et défaillantes soit la même dans chaque échantillon ;
- puis, 25 échantillons *bootstraps* ont été tirés à partir de A et ont servi à calculer, pour chaque ensemble de variables retenues, autant de règles de classification que d'échantillons *bootstraps* ;
- enfin, à partir des 25 classifieurs calculés précédemment, nous avons ensuite déterminé la classe d'appartenance des observations de l'échantillon de test T sur la base d'un vote majoritaire (une observation a été classée dans le groupe ayant obtenu le score le plus élevé) et calculé l'erreur de prévision à l'issue du vote majoritaire.

L'ensemble de ces opérations a été réalisé 100 fois et ce, pour chaque ensemble de variables.

Une fois le processus achevé, nous avons calculé la moyenne des erreurs afin d'obtenir une approximation de l'erreur en généralisation de l'ensemble des modèles élaborés. Comme les observations de l'échantillon initial (500 entreprises) n'ont pas été tirées de manière indépendante et identiquement distribuées, les modèles comportent un biais en direction des échantillons de test *bootstraps*. C'est pourquoi nous avons réutilisé les 25 x 100 modèles créés précédemment pour étudier leur capacité de généralisation et estimer plus précisément l'erreur associée à cette capacité à partir d'un nouvel échantillon, dit de validation, comportant 520 entreprises. Les échantillons sont décrits dans le paragraphe suivant.

Là aussi, les taux d'erreur de classification dépendent de la valeur du *cut* employé pour séparer les groupes. Pour les mêmes raisons que celles évoquées précédemment, les valeurs employées ont toujours été celles permettant de minimiser l'erreur globale de classification.

4.5 ÉCHANTILLONS

La plupart des études sur la défaillance réalisées jusqu'à présent se sont appuyées sur des échantillons « pairés » en fonction de différents paramètres, parmi lesquels on retrouve la plupart du temps le secteur d'activité, la classe d'actifs et l'année des comptes employés. Ceci revient à définir au sein de l'échantillon, un ensemble d'entreprises saines et un ensemble d'entreprises défaillantes qui sont presque similaires selon ces critères. Pour constituer notre échantillon initial, dont sont issus les échantillons d'apprentissage et de test *bootstraps* qui ont servi à paramétrer puis à tester les modèles, nous avons aussi retenu des critères liés au secteur, à la taille des actifs et à l'année des comptes, mais sans procéder à un quelconque « pairage » des observations, puisque l'échantillon constitué selon ces trois dimensions est suffisamment homogène pour éviter d'avoir recours à cette pratique.

Nous avons d'abord sélectionné un secteur d'activité puis un ensemble de structures sociales. Nous avons retenu le secteur du commerce, selon la nomenclature de l'Insee (codes NAF 5xx), car il représente, avec le bâtiment, l'un des secteurs qui traditionnellement compte le plus grand nombre d'entreprises défaillantes et facilite ainsi le choix de données. Seules les entreprises de type SA, SARL, SAS ou EURL ont ensuite été retenues, car représentant les structures juridiques les plus communes. Au sein de cet ensemble, nous avons choisi de retenir des firmes ayant une structure d'actifs aussi homogène que possible du point de vue des ratios que nous allons employer. Comme rien n'indique *a priori* comment effectuer un tel découpage, nous avons réalisé, pour tous les ratios retenus, une série d'analyses de variance et de tests de *Mann-Whitney*³⁶ sur différentes partitions, afin de retenir celle qui apparaissait comme la plus homogène, en étudiant séparément les entreprises saines et les défaillantes. Il fallait donc trouver un seuil capable à la fois de ne pas pénaliser la taille de l'échantillon et de garantir son homogénéité. Nous avons ainsi testé une série de seuils, isolant à chaque fois deux ensembles : un situé au dessus, l'autre en dessous. De toutes les partitions testées, la partition opposant les

³⁶ L'analyse de la variance compare les moyennes d'échantillons, mais elle suppose la normalité des distributions comme l'égalité des variances. Comme les variables employées ici sont très loin d'être distribuées normalement, nous avons aussi employé un test de Mann-Whitney qui lui, abandonne l'hypothèse de normalité et compare la somme des rangs des observations de chaque groupe.

firmes de moins 750 000 € d'actifs de toutes les autres était précisément l'une de celles permettant d'obtenir les différences les plus importantes sur le plus grand nombre de ratios, sans pénaliser la taille de l'échantillon. Ce seuil a un côté aussi arbitraire car, par exemple, le seuil d'un million d'euros aboutissait à des résultats sensiblement identiques. Nous ne l'avons cependant pas retenu, car ce dernier ne permettait d'accroître que marginalement le nombre d'entreprises. Or, l'écart entre 750 000 € et 1 000 000 € nous a semblé trop important par rapport au nombre d'entreprises faisant partie de cet intervalle, relativement au nombre de firmes disposant de moins de 250 000 € ou ayant entre 250 000 et 500 000 €. Le seuil retenu correspond donc à une frontière délimitant deux groupes relativement homogènes de chaque catégorie d'entreprises, au regard des principaux indicateurs de défaillance permettant de garantir que les modèles créés ne seront pas biaisés par un effet de taille.

Nous avons ensuite sélectionné des données et retenu uniquement des données de nature comptable, pour les raisons évoquées au chapitre 1, en les manipulant sous la forme de ratios. Nous n'avons pas souhaité effectuer un recueil sur plusieurs années pour les motifs présentés aux paragraphes 1.4.2, p. 66, et 1.4.3, p. 76, relatifs à la stabilité et à l'horizon de prévision des modèles. C'est pourquoi, nous avons choisi une seule année d'arrêté des comptes, 2002, avec néanmoins une donnée de l'année précédente (capitaux propres). De plus, nous n'avons conservé, pour les entreprises saines, que les entreprises « réellement » saines, comme le suggérait Malecot (1991). Nous avons donc éliminé toutes les entreprises qui, en 2005, n'existaient plus pour cause de liquidation ou de dissolution. En complément, seules les entreprises ayant au moins quatre ans d'activité ont été retenues, car souvent les entreprises, au cours des toutes premières années, ont une structure financière proche d'entreprises en mauvaise santé. Nous avons aussi pris en compte l'historique des firmes afin de ne conserver dans le groupe des saines que celles qui, en plus des critères précédents, n'ont pas connu de défaillance au cours des quatre années qui ont précédé la date retenue pour la sélection des comptes. En effet, les entreprises qui ont connu un redressement, puis un moratoire de façon à poursuivre leur activité, peuvent pendant quelques années porter dans leurs comptes la trace de cet événement et donc disposer d'un profil pouvant les laisser confondre avec des défaillantes. Ainsi, Rapinel (1994), dans une étude publiée par la Banque de France, montre que très souvent un tel moratoire n'est pas accompagné d'un effort en fonds propres suffisant de la part des repreneurs. Ceci peut alors provoquer un déséquilibre du passif tel que ces entreprises, bien qu'étant aptes à poursuivre leur activité, finissent par présenter une situation financière proche

de certaines firmes défailtantes et conserver ce profil pendant un délai qui peut aller jusqu'à cinq ans. Le choix de 2002 s'est opéré par rapport à celui de 2005, année pour laquelle on peut disposer des comptes de toutes les entreprises encore en activité, pour avoir non seulement les données les plus récentes, donc proches de la conjoncture actuelle, mais aussi pour disposer d'un délai suffisamment long entre ces deux années pour juger de la solidité d'une entreprise.

Quant aux défailtantes, nous avons retenu toutes celles pour cause de redressement ou de liquidation en 2003 et dont la date de clôture des derniers comptes utilisés est postérieure d'au moins 12 mois à la date du jugement de redressement ou de liquidation, de façon à ce que les comptes retenus ne soient pas des comptes calculés juste avant ou juste après le jugement de règlement judiciaire et éliminer ainsi toute distorsion dont nous avons décrit les effets au paragraphe 1.4.3, p. 76. Malecot (1991) suggérait de ne retenir que les défailtantes pour cause financière et d'exclure les entreprises qui sombrent dans la défailtance à cause d'une grève, d'une technologie de production dépassée... Mais la complexité de la tâche ne nous a pas permis d'en tenir compte.

Afin que l'échantillon ne soit pas composé majoritairement d'entreprises défailtantes de création récente, donc que les modèles créés ne soient pas que des modèles de prévision de défailtance des jeunes entreprises (c'est dans ce segment que l'on trouve le taux de liquidation et de redressement le plus important), les firmes ont été sélectionnées en tentant d'équilibrer les effectifs des tranches « 6 – 10 ans », « 11 – 15 ans » et « 16 ans et plus »³⁷. Quant à la tranche « 5 ans et moins », elle est sous représentée, car nous n'avons retenu que des entreprises saines d'au moins 4 ans d'activité. Les entreprises saines ont été choisies de manière à ce que la distribution des dates de création des deux groupes d'entreprises soit similaire, donc que l'on ait autant de jeunes que de moins jeunes dans les deux groupes (seule la première tranche diffère car les saines ont au moins quatre ans d'existence).

Enfin, nous avons retenu les défailtantes pour lesquelles on dispose de données comptables en 2002 et des capitaux propres en 2001, et dont la défailtance s'est matérialisée par une décision de justice en 2003 (redressement ou liquidation). Ce premier échantillon, composé de 250 entreprises saines et de 250 entreprises défailtantes, a servi à la sélection des variables et à une évaluation de l'erreur en généralisation des modèles créés. Les firmes défailtantes ont été

³⁷ Parmi les entreprises qui disparaissent, il y a en moyenne deux fois plus d'entreprises de moins de 5 ans que d'entreprises ayant entre 6 et 10 ans d'existence. En revanche, dans les tranches « 11 - 15 ans », et « 16 ans et plus », le taux est à peu près égal à la moitié du taux de la tranche « 6-10 ans » (sources : www.insee.fr et Les défailtances d'entreprises en France au premier trimestre 2006, Altares, Les études trimestrielles, 7 p.).

choisies parmi les 1 548 entreprises appartenant au secteur du commerce et enregistrées par Diane, ce dernier nombre représentant environ 15,5% du total des défailtantes de ce secteur comptabilisées par l’Insee en France cette année-là. Tous les comptes ont été vérifiés, un à un, à partir des données publiées par societe.com, et seules 15 entreprises ont été éliminées parce que non référencées par cette dernière base.

Nous avons, par ailleurs, constitué un deuxième échantillon de firmes du même secteur et de la même classe d’actifs, dit de validation, mais les données retenues sont celles de l’exercice 2003, avec là aussi les capitaux propres de l’année précédente, soit 2002. Les entreprises défailtantes retenues ont été déclarées en redressement ou en liquidation par une décision de justice en 2004, toujours avec un décalage d’au moins 12 mois entre la date de clôture des comptes utilisés et la date du jugement. Les entreprises saines ont été choisies aléatoirement parmi toutes celles toujours en activité fin 2005. Ce deuxième échantillon est composé de 260 défailtantes et de 260 entreprises saines tirées aléatoirement parmi l’ensemble des entreprises répondant aux critères précédents et présentes dans la base Diane. Nous avons fait en sorte que ces entreprises ne fassent pas partie de l’échantillon précédent, contrairement à certaines études qui considèrent que les comptes d’une année d’une entreprise donnée, associés au statut de l’entreprise pour la dite année, peuvent être considérés comme formant une entreprise à part entière ; ce qui revient à dire que deux occurrences de comptes d’une même firme peuvent représenter deux firmes. Cet échantillon a servi à valider l’erreur en généralisation des modèles, c’est-à-dire l’erreur que l’on observera lorsque les modèles seront confrontés à des données qui n’ont jamais été employées pour leur paramétrage et qui reflète leur véritable performance.

Tableau 29 : Structures des échantillons

Caractéristiques générales		
Secteur	Commerce	
Statuts juridiques	SA – SARL – SAS – EURL	
Taille des actifs	35 000 € – 750 000 €	
Critères de défaillance	Redressement et Liquidation	
Années de prélèvement des comptes	Statut	
	Saines	Défailtantes
Échantillons Apprentissage et Test	2002	2002
Échantillon Validation	2003	2003

Une donnée de l’année 2001, pour les échantillons d’apprentissage et de test, et une de l’année 2002, pour l’échantillon de validation, ont aussi été prélevées (capitaux propres)

Nombre d'entreprises	Statut	
	Saines	Défaillantes
Échantillons Apprentissage et Test	250	250
Échantillon Validation	260	260

Distribution de l'âge des entreprises appartenant aux échantillons d'apprentissage et de test	Statut	
	Saines	Défaillantes
5 ou moins de 5 ans	13	29
6 – 10 ans	77	76
11 – 15 ans	71	72
16 ans et plus	89	73

Distribution du type de défaillance ayant affecté les entreprises appartenant aux échantillons d'apprentissage et de test	Statut	
	Saines	Défaillantes
Liquidation		170
Redressement		80

4.6 VARIABLES

L'analyse de la littérature que nous avons conduite a fait ressortir le fait que plus de 500 ratios différents ont été employés jusqu'à présent dans des modèles de prévision de la défaillance. Ce nombre ne correspond qu'aux ratios retenus après sélection, laissant augurer du nombre particulièrement important de ratios qui ont dû être choisis initialement. Tous ces ratios sont issus de combinaisons diverses et variées des postes dont la liste est donnée dans le tableau 30 indiqué ci-après, qui distingue ceux qu'on retrouve le plus souvent au numérateur et au dénominateur.

Tableau 30 : Postes comptables inclus dans les ratios des modèles de défaillance

Numérateur		Dénominateur
Actif Circulant	Résultat d'Exploitation	Actif Circulant
Actif Net	Stocks	Actif Net
Capitaux Propres	Total Actif (Passif)	Capitaux Propres
Chiffre d'Affaires	Total Dette	Chiffre d'Affaires
Dette Court Terme	Trésorerie	Dette Court Terme
Dette Long Terme		Total Actif (Passif)
Intérêts sur Dette		Total Dette
Résultat Net		

On retrouve bien dans ce tableau les principaux indicateurs de l'état de santé des firmes, depuis la structure capitalistique jusqu'à la rentabilité en passant par les grands équilibres entre les éléments de long terme et de court terme, donc par les relations entre les engagements et leurs modes de

financement, sans oublier des variables d'activité. Tous ces postes sont employés pour construire des ratios tentant d'évaluer aussi bien un éventuel déséquilibre financier que son ampleur.

Pour déterminer les ratios à inclure dans notre étude, nous nous sommes d'abord référé à l'analyse financière plutôt qu'à des pratiques expérimentales, afin de tenir compte de toutes les dimensions susceptibles d'expliquer l'état de défaillance au travers des mécanismes pouvant induire un déséquilibre de la structure financière des firmes. Nous en avons retenu sept : liquidité-solvabilité, structure financière, rentabilité, efficacité, rotation, prélèvement et apport. Pour chacune des sept dimensions retenues³⁸, nous avons choisi les ratios et variables présentés dans le tableau 31 ci-après.

Tableau 31 : Ratios et variables retenus

Liquidité – Solvabilité	LI	Rentabilité	RE	Rotation	RO
Actif Circulant / Dette Court Terme		EBE / Capitaux Permanents		Actif Circulant / CA	
Actif Circulant / Total Actif		EBE / Total Actif		BFR / CA	
(Actif Circulant - Stocks) / Total Actif		Résultat Courant Av. I. / Capitaux Propres		Clients / CA	
(Actif Circulant - Stocks) / Dette Court Terme		Résultat Exploitation / Total Actif		Fournisseurs / CA	
Dette Court Terme / Total Actif		Résultat Net / Capitaux Propres		Stocks / CA	
Dette Financière / CAF		Résultat Net / Total Actif		Trésorerie / CA	
(Disponibilités + VMP) / CA					
(Disponibilités + VMP) / Total Actif		Efficacité	EF	Prélèvement	PR
EBE / CA		CA / Capitaux Propres		Frais Financiers / CA	
Trésorerie / Dette Court Terme		CA / Total Actif		Personnel / CA	
Trésorerie / Total Actif		ETE / Total Actif			
Trésorerie / Total Dette		ETE / CA		Apport	AP
		Marge Commerciale / CA		Variation Autres Dettes	
Structure Financière	SF	Résultat Exploitation / CA		Variation Capitaux Propres Hors Résultat	
BFR / Total Actif		Valeur Ajoutée / CA			
Capitaux Propres / Total Actif					
Dette Long Terme / Capitaux Propres					
Dette Long Terme / Total Actif					
Total Dette / Capitaux Propres					
Total Dette / Total Actif					

BFR :	Besoin en Fond de Roulement	EBE :	Excédent Brut d'Exploitation
CA :	Chiffre d'Affaires	ETE :	Excédent de Trésorerie d'Exploitation
CAF :	Capacité d'Autofinancement	VMP :	Valeurs Mobilières de Placement

³⁸ Dans tous les tableaux qui vont suivre, dès lors qu'ils concerneront les variables, nous précisons, en regard de chacune d'elles, un code de deux caractères correspondant à leur catégorie d'appartenance correspondant à une des dimensions décrites dans le tableau 31.

Le premier élément d'équilibre s'inscrit dans la capacité de l'entreprise à faire face à ses engagements à court et à long terme et renvoie à sa liquidité et à sa solvabilité. Il s'agit ici, de savoir dans quelle mesure une entreprise est capable d'abord de couvrir ses dettes à court terme à partir de ses actifs exigibles, tout en maintenant une marge de sécurité pour faire face à une détérioration de leur rotation et ainsi maintenir durablement sa solvabilité. Le rôle de la trésorerie est alors essentiel car elle permet d'absorber des retards dans les encaissements de créances à court terme ou bien une accélération dans le paiement des dettes de même horizon.

Le deuxième élément d'équilibre renvoie à la structure financière de l'entreprise, qui participe aussi du processus qui conduit à la solvabilité. Dans cette optique, l'appréciation de la structure financière peut alors être effectuée au travers de l'aptitude de la firme à engendrer des liquidités servant à honorer ses échéances, mais aussi au travers de son degré d'autonomie vis-à-vis de débiteurs externes et de sa capacité de remboursement des dettes à long terme.

Le troisième élément d'équilibre s'inscrit dans la rentabilité qui sanctionne la façon dont les ressources ont été affectées à l'exercice de l'activité et permet d'apprécier la performance industrielle et commerciale d'une entreprise. On s'intéresse alors, à l'activité en elle-même et à son économie, donc à la capacité de la firme à mobiliser ses ressources pour dégager un résultat, mais aussi à l'incidence de l'endettement sur la rentabilité des capitaux. En effet, l'activité économique peut profiter d'un endettement, dès lors que le coût de ce dernier est inférieur au gain de la première, ou bien être desservie par celui-ci dans l'hypothèse d'un ralentissement de l'activité, la firme endettée pouvant rencontrer des difficultés à servir les intérêts de sa dette.

Le quatrième a trait à l'efficacité de l'entreprise, son efficience, et son aptitude à vendre et à dégager de la valeur ajoutée. Elle est révélatrice d'une position plus ou moins forte et stable sur un marché et constitue un bon indicateur de pérennité.

Le cinquième est lié à la rotation des éléments du besoin en fond de roulement (BFR), *i.e.* à l'aptitude des stocks ou des créances clients à se transformer plus ou moins vite en encaissements, ou à celles des dettes aux fournisseurs à se transformer en décaissements. La dégradation du BFR est généralement un indicateur des difficultés de la firme et accélère le processus de par les charges financières qu'elle engendre.

Le sixième concerne la capacité du chiffre d'affaires à faire face à la croissance de certaines charges pouvant gréver la liquidité de l'entreprise.

Enfin, le septième, vise à mesurer, au travers d'apports, l'évolution dans le temps des ressources stables auxquelles une firme peut faire appel, soit pour alimenter et soutenir son activité dans une stratégie de développement, soit pour pallier des insuffisances, voire se maintenir à flot. De toutes les variables employées, seules celles de cette catégorie ne sont pas exprimées sous la forme de ratios. Elles correspondent simplement à une mesure de variation effectuée par la différence entre deux valeurs.

Tous ces ratios et variables ont été définis indépendamment des caractéristiques des échantillons, et notamment de la taille des entreprises et du secteur et sans qu'aucun test statistique n'ait été réalisé *a priori*. Comme la plupart des auteurs qui ont souhaité disposer d'un large spectre de choix, seules des considérations relatives aux dimensions pouvant affecter l'équilibre financier des firmes ont été privilégiées.

4.7 RETRAITEMENT DES DONNÉES

L'examen des variables calculées avec les données des échantillons d'apprentissage et de test a montré un étirement extrême des queues de distribution et une forte absence de normalité. Nous n'avons pas choisi de couper les queues avant le premier centile et après le dernier centile. En effet, l'examen des distributions au-delà du premier et du troisième quartile montre que de nombreuses variables possèdent des queues extrêmement longues et que toute coupure ne change quasiment rien à leur moyenne, leur écart type, leur mode ou médiane, et rien non plus aux tests de différences de moyenne entre les groupes, qu'il soit paramétriques ou non. C'est la raison pour laquelle, il nous est apparu qu'une coupure ne permettant pas de modifier les distributions des variables, il n'était pas légitime d'opérer ainsi. Nous avons donc conservé toutes les variables en l'état. Ensuite, nous avons centré et réduit les variables, car elles ne sont pas toutes mesurées sur la même échelle. En effet, les variables de la catégorie Apport correspondent à des différences et non à des rapports.

Le tableau 32 présenté à la page suivante indique, dans sa partie centrale, les quartiles des distributions de chaque variable. Il illustre ainsi l'amplitude des écarts qui existent entre les deux groupes d'entreprises : les défaillantes et les saines. Dans sa partie droite, il précise les résultats d'un test de *Shapiro-Wilks* mesurant l'écart à la normalité des distributions des variables, ainsi que les résultats de deux tests d'évaluation des différences entre les saines et les défaillantes et ce, pour chaque variable. Les tests de normalité montrent clairement qu'aucune

variable n'est distribuée selon une courbe de *Gauss*, au seuil habituel de signification, ce qui est généralement le cas de tous les ratios financiers. Aussi, des deux tests de différences effectués, le test non paramétrique *U* de *Mann-Whitney*, est-il plus fiable que le test *t*. Il montre qu'au seuil de 5 %, toutes les variables, sauf CA / Total Actif, Actif Circulant / CA et Personnel / CA, présentent bien des différences significatives entre chaque groupe.

Tableau 32 : Caractéristiques des variables des échantillons d'apprentissage et de test
Quartiles – Tests de normalité et de différences entre les groupes

	Quartiles											
	Défaillantes			Saines			Défaillantes		Saines		t	U
	25%	50%	75%	25%	50%	75%	S-W	S-W				
Liquidité												
Actif Circulant / Dette Court Terme	-0,86	-0,54	-0,27	-0,16	0,20	0,77	0,000	0,000	0,000	0,000		
Actif Circulant / Total Actif	-0,41	0,46	0,91	-0,68	0,04	0,70	0,000	0,000	0,008	0,000		
(Actif Circulant - Stocks) / Dette Court Terme	-0,85	-0,55	-0,22	-0,31	0,25	0,95	0,000	0,000	0,000	0,000		
(Actif Circulant - Stocks) / Total Actif	-0,95	-0,11	0,63	-0,69	0,21	0,81	0,000	0,000	0,063	0,039		
Dette Court Terme / Total Actif	-0,15	0,29	0,70	-0,79	-0,53	-0,28	0,000	0,024	0,000	0,000		
Dette Financière / CAF	-0,07	-0,05	-0,03	-0,04	-0,04	-0,02	0,000	0,000	0,669	0,000		
(Disponibilités + VMP) / CA	-0,78	-0,68	-0,38	-0,39	0,13	1,13	0,000	0,000	0,000	0,000		
(Disponibilités + VMP) / Total Actif	-0,85	-0,73	-0,35	-0,43	0,28	1,28	0,000	0,000	0,000	0,000		
EBE / CA	-0,96	-0,30	0,02	0,15	0,35	0,72	0,000	0,000	0,000	0,000		
Trésorerie / Dette Court Terme	-0,70	-0,47	-0,36	-0,21	0,23	0,95	0,000	0,000	0,000	0,000		
Trésorerie / Total Actif	-0,94	-0,42	-0,13	-0,10	0,44	1,19	0,000	0,095	0,000	0,000		
Trésorerie / Total Dette	-0,68	-0,44	-0,33	-0,21	0,18	0,89	0,000	0,000	0,000	0,000		
Structure Financière												
BFR / Total Actif	-0,54	0,05	0,50	-0,10	0,21	0,53	0,000	0,473	0,000	0,000		
Capitaux Propres / Total Actif	-0,69	-0,15	0,07	0,34	0,49	0,73	0,000	0,000	0,000	0,000		
Dette Long Terme / Capitaux Propres	-0,10	-0,04	0,02	-0,03	-0,01	0,02	0,000	0,000	0,681	0,000		
Dette Long Terme / Total Actif	-0,57	-0,14	0,50	-0,63	-0,33	0,10	0,000	0,000	0,000	0,000		
Total Dette / Capitaux Propres	-0,77	-0,49	-0,33	-0,05	0,18	0,73	0,000	0,000	0,000	0,000		
Total Dette / Total Actif	-0,08	0,15	0,72	-0,74	-0,48	-0,32	0,000	0,000	0,000	0,000		
Rentabilité												
EBE / Capitaux Permanents	-0,11	-0,06	-0,02	-0,04	-0,03	-0,01	0,000	0,000	0,743	0,000		
EBE / Total Actif	-0,81	-0,30	0,03	0,20	0,39	0,68	0,000	0,000	0,000	0,000		
Résultat Courant Av. I. / Capitaux Propres	0,01	0,06	0,08	0,08	0,09	0,11	0,000	0,000	0,003	0,000		
Résultat Exploitation / Total Actif	-0,87	-0,23	0,11	0,27	0,43	0,70	0,000	0,000	0,000	0,000		
Résultat Net / Capitaux Propres	-0,09	0,00	0,10	-0,01	0,01	0,04	0,000	0,000	0,651	0,026		
Résultat Net / Total Actif	-0,86	-0,24	0,11	0,29	0,41	0,70	0,000	0,000	0,000	0,000		

Efficienc										
CA / Capitaux Propres	-0,11	-0,05	0,14	0,01	0,04	0,09	0,000	0,000	0,142	0,000
CA / Total Actif	-0,59	-0,29	0,31	-0,61	-0,29	0,24	0,000	0,000	0,878	0,994
ETE / CA	-0,64	-0,28	0,18	-0,19	0,12	0,55	0,000	0,000	0,000	0,000
ETE / Total Actif	-0,67	-0,28	0,26	-0,18	0,16	0,50	0,000	0,000	0,000	0,000
Marge Commerciale / CA	-0,64	-0,10	0,47	-0,55	0,08	0,71	0,000	0,005	0,001	0,006
Résultat Exploitation / CA	-0,95	-0,24	0,09	0,21	0,40	0,69	0,000	0,000	0,000	0,000
Valeur Ajoutée / CA	-0,89	-0,46	0,24	-0,32	0,24	0,99	0,000	0,001	0,000	0,000
Rotation										
Actif Circulant / CA	-0,70	-0,04	0,60	-0,78	-0,16	0,39	0,000	0,000	0,028	0,121
BFR / CA	-0,67	-0,04	0,44	-0,18	0,13	0,48	0,000	0,000	0,000	0,000
Clients / CA	-0,70	-0,08	0,74	-0,86	-0,44	0,25	0,000	0,000	0,000	0,001
Fournisseurs / CA	-0,34	0,14	0,76	-0,76	-0,51	-0,11	0,000	0,000	0,000	0,000
Stocks / CA	-0,64	-0,13	0,65	-0,76	-0,40	0,17	0,000	0,000	0,000	0,001
Trésorerie / CA	-0,88	-0,40	-0,15	-0,11	0,30	1,08	0,000	0,000	0,000	0,000
Prélèvement										
Frais Financiers / CA	-0,26	-0,19	-0,03	-0,30	-0,27	-0,22	0,000	0,000	0,046	0,000
Personnel / CA	-0,73	-0,17	0,47	-0,67	-0,11	0,51	0,000	0,000	0,431	0,995
Apport										
Variation Autres Dettes	-0,09	-0,05	0,18	-0,09	-0,05	-0,01	0,000	0,000	0,139	0,026
Variation Capitaux Propres Hors Résultat	0,25	0,25	0,25	-0,44	0,15	0,25	0,000	0,000	0,000	0,000

S-W : Seuil de significativité – p-valeur – d'un test de Shapiro-Wilks portant sur la normalité des distributions

t : Seuil de significativité – p-valeur – d'un test t de Student portant sur l'égalité des moyennes des 2 groupes

U : Seuil de significativité – p-valeur – d'un test U de Mann-Whitney portant sur l'égalité de la somme des rangs des 2 groupes

Les modèles construits à l'aide de l'analyse discriminante et de la régression logistique ont été évalués à partir des données centrées-réduites. En revanche, avec le réseau, les données ont été discrétisées en classes de taille égale (30 classes) puis converties de manière à ce que l'intervalle des valeurs soit compris entre -1 et +1, de façon à être compatible avec la sortie de la fonction d'activation des neurones qui, rappelons-le, emploie une tangente hyperbolique. Pourquoi avoir discrétisé ? Dans la littérature consacrée aux réseaux de neurones, on trouve de nombreux travaux consacrés à la façon de les paramétrer et aux types de retraitement à opérer sur les données. Parmi ces derniers, la transformation des données initiales lorsqu'elles sont exprimées sous la forme d'échelles d'intervalle ou de rapport, en données ordinales, apparaît comme une pratique assez courante. Nous avons conduit une série de tests avec et sans transformation pour étudier le pouvoir discriminant d'ensembles de variables aléatoires.

Ces tests ont montré que la discrétisation était favorable à une meilleure discrimination. C'est pourquoi nous avons procédé ainsi. Nous avons aussi effectué des tests pour tenter de déterminer le nombre de classes à conserver. Nous en avons retenu 30 de manière assez arbitraire, car sur les différentes expérimentations conduites, nous n'avons pas observé de relation ténue entre les résultats des classements et la finesse d'un tel découpage.

4.8 LOGICIELS EMPLOYÉS ET PROCÉDURES DE CALCUL

Pour effectuer nos calculs, nous avons eu recours à SPSS pour une partie des traitements, à Excel pour toutes les opérations de consolidation, et nous avons développé toute une série de procédures en VBA avec Excel.

SPSS a servi pour les traitements statistiques de base (certains recodages de variables, statistiques descriptives, tests d'hypothèses...), les analyses factorielles, la sélection des variables à l'aide du critère *Lambda de Wilks* et du critère de *vraisemblance* et l'estimation des paramètres des fonctions discriminantes et logistiques.

Tous les autres traitements ont été développés par nos soins. Ainsi, les procédures de *bootstrap*, celles propres au réseau de neurones employé ici pour effectuer les classifications (*Perceptron multicouches*), celles relatives aux sélections de variables adaptées au réseau et celles permettant de définir les variables apparaissant le plus souvent au cours des sélections, celles implémentant l'algorithme de *Kohonen* et celui de *Sammon*, pour ne citer que les principales, ont été réalisées en VBA sur Excel. En tout, ce sont plus de 20 000 lignes de codes qui ont été écrites pour effectuer tous ces calculs.

La sélection des variables à l'aide du réseau a nécessité l'emploi de 30 machines fonctionnant en parallèle pendant de nombreuses semaines compte tenu du nombre de combinaisons à explorer, et les autres sélections effectuées avec SPSS ont été en partie automatisées à l'aide de macros.

Les classifications ont été opérées en deux temps. Les coefficients des fonctions ont d'abord été calculés à l'aide de procédures écrites en VBA mais aussi à l'aide de macro avec SPSS, par ensemble de 25 fonctions, pour faciliter les vérifications, puis les statistiques de classements ont été réalisées pour chaque ensemble de 25 fonctions avec Excel. Enfin, les calculs de consolidation pour obtenir les résultats définitifs ont été opérés là aussi avec Excel, ainsi que le codage des variables appartenant aux échantillons de validation afin que les valeurs employées soient compatibles avec les valeurs des échantillons d'apprentissage et de test.

4.9 CONCLUSION DU CHAPITRE

Nous avons présenté dans ce chapitre nos choix méthodologiques au travers des méthodes de construction de modèles et des techniques de sélection de variables que nous souhaitons comparer et nous avons présenté les modalités de paramétrage du réseau de neurones. Les critères de choix ont été dictés essentiellement par la littérature sur la défaillance, puisque l'un de nos objectifs consiste précisément à montrer les limites de pratiques qui présentent un caractère assez récurrent, sauf lorsqu'il s'est agi de retenir les méthodes de sélection de variables propres au réseau ; quasiment aucun modèle de défaillance n'ayant été construit par leur intermédiaire.

Nous avons aussi présenté les protocoles de sélection des variables et de validation des modèles fondés sur une technique de rééchantillonnage, le *bootstrap*, afin de limiter l'impact des particularités de nos échantillons sur les résultats. Nous avons ensuite décrit les échantillons retenus et les modalités de sélection des observations, à partir des principales conclusions tirées au chapitre 1 et relatives notamment à la stationnarité des modèles.

Enfin, nous avons présenté la stratégie de choix des variables explicatives puis décrit la structure des données appartenant aux échantillons qui ont servi à l'élaboration des modèles et expliqué les modalités de réalisation de tous les calculs.

CHAPITRE 5 : STRUCTURES ET PERFORMANCES DES MODÈLES

Les résultats que nous allons présenter dans ce chapitre concernent à la fois la structure des modèles et leurs performances. Nous avons souhaité faire ressortir plusieurs points. Tout d'abord, comme on peut le constater dans le tableau 12, p. 38, certains auteurs se sont appuyés sur des tests notamment de différences de moyennes (test t ou test F) pour effectuer les sélections ou les présélections des entrées d'un réseau, ou ont choisi directement des ensembles de variables déjà retenus par d'autres à l'aide de ces mêmes tests. Ces choix sont-ils judicieux ? Une hiérarchie de variables opérée à l'aide de telles mesures de distance se retrouve-t-elle dans celle qui peut être obtenue à l'aide d'autres mesures ? Il paraît légitime de penser que plus les classes sont homogènes et bien séparées les unes des autres, plus cette réalité est à même de jouer sur tout résultat d'une sélection. Certes, ce n'est pas la seule, mais nous analyserons plus particulièrement la nature du rapport pouvant exister entre la capacité discriminante d'une variable au sens de ces tests paramétriques et cette même capacité mesurée sous d'autres angles.

Ensuite, nous étudierons, au travers des variables retenues, ce qui distingue chaque famille de critère. Il s'agira notamment d'appréhender les différences de représentations de la défaillance qu'elles donnent. Puis, nous analyserons les points de convergence de tous les modèles au travers des variables qu'ils partagent et de leur importance respective dans chacun d'eux, qui permettraient de mettre en lumière une structure ou des éléments de structure qui sont recherchés depuis longtemps. En effet, bon nombre d'études ont tenté de trouver des constantes dans les modèles, la plupart du temps en examinant le comportement d'un même modèle une fois appliqué à des situations pour lesquelles il n'a pas été conçu. Ce fut le cas notamment de celui d'Altman (1968) et celui d'Ohlson (1980), au travers de processus de validation souvent *ex post*. Or, comme nous l'avons souligné, peu de constantes ont pu être soulignées ainsi, car les modèles sont très dépendants de paramètres conjoncturels qu'ils ne maîtrisent pas. C'est pourquoi, nous avons choisi d'explorer une voie qui ne l'a quasiment jamais été. Seuls Back et al. (1996 A) ont comparé des ensembles de variables sélectionnés spécifiquement pour une analyse discriminante (méthode *stepwise* et test F), une régression logistique (méthode *stepwise*) et un réseau de neurones (algorithme génétique), mais seulement pour étudier les différences de performance de modèles selon plusieurs horizons de prévision (1, 2 et 3 ans).

Enfin, nous détaillerons les performances de chaque modèle et nous analyserons les rapports entre leur degré de précision et les techniques mises en œuvre pour les concevoir.

5.1 VARIABLES SÉLECTIONNÉES

Les tableaux présentés ci-après indiquent pour chaque couple « méthode de recherche – critère dévaluation », les ratios retenus lors de la phase de sélection finale et leur fréquence de choix. Ils ont été classés par ordre décroissant de fréquence et répartis à chaque fois en deux groupes. Le premier correspond aux ensembles retenus pour l'élaboration des modèles de prévision, donc aux ratios ayant été retenus dans plus de 70 % des cas, et le deuxième aux ratios dont la fréquence de sélection est inférieure à 70%.

Tableau 33 : Variables sélectionnées

Critère d'évaluation : Lambda de Wilks Recherche : Stepwise		Fréquence de sélection
Trésorerie / Total Actif	LI	93,4%
Total Dette / Capitaux Propres	SF	91,1%
Trésorerie / Total Dette	LI	88,7%
(Disponibilités + VMP) / Total Actif	LI	87,5%
Résultat Exploitation / Total Actif	RE	81,2%
EBE / Total Actif	RE	76,8%
Capitaux Propres / Total Actif	SF	72,2%
<hr/>		
EBE / CA	LI	61,9%
Actif Circulant / Dette Court Terme	LI	60,4%
Dette Court Terme / Total Actif	LI	49,7%
Trésorerie / Dette Court Terme	LI	37,1%

Critère d'évaluation : Vraisemblance Recherche : Backward Stepwise		Fréquence de sélection
Capitaux Propres / Total Actif	SF	94,0%
Résultat Courant Av. I. / Capitaux Propres	RE	89,3%
Variation Capitaux Propres Hors Résultat	AP	87,6%
(Disponibilités + VMP) / Total Actif	LI	86,1%
(Disponibilités + VMP) / CA	LI	81,5%
EBE / Total Actif	RE	73,9%
Trésorerie / CA	RO	70,2%
<hr/>		
BFR / CA	RO	68,7%
Total Dette / Capitaux Propres	SF	62,2%
Résultat Exploitation / Total Actif	RE	54,5%
Actif Circulant / Total Actif	LI	49,4%
Total Dette / Total Actif	SF	42,3%

Critère d'évaluation : Vraisemblance Recherche : Forward Stepwise		Fréquence de sélection
Variation Capitaux Propres Hors Résultat	AP	83,6%
Capitaux Propres / Total Actif	SF	81,2%
Trésorerie / Total Dette	LI	77,3%
EBE / Total Actif	RE	72,1%
Résultat Exploitation / CA	EF	70,8%
<hr/>		
Résultat Net / Total Actif	RE	65,3%
Trésorerie / CA	RO	57,0%
Résultat Courant Av. I. / Capitaux Propres	RE	48,7%
Dette Court Terme / Total Actif	LI	35,2%
(Disponibilités + VMP) / Total Actif	LI	27,2%

Critère d'évaluation : Erreur		Fréquence
Recherche : Backward		de sélection
Capitaux Propres / Total Actif	SF	91,8%
Résultat Exploitation / Total Actif	RE	86,2%
Trésorerie / Dette Court Terme	LI	83,1%
Variation Capitaux Propres Hors Résultat	AP	81,8%
EBE / Total Actif	RE	76,6%
Résultat Exploitation / CA	EF	76,1%
(Disponibilités + VMP) / Total Actif	LI	74,9%
Clients / CA	RO	70,6%
Résultat Net / Capitaux Propres	RE	66,8%
Résultat Courant Av. I. / Capitaux Propres	RE	61,4%
BFR / CA	RO	52,1%
Total Dette / Total Actif	SF	46,7%
Dette Court Terme / Total Actif	LI	41,2%
Trésorerie / CA	RO	37,1%
Trésorerie / Total Actif	LI	26,9%

Critère d'évaluation : Mesure Ordre 0		Fréquence
Recherche : Backward		de sélection
Résultat Net / Total Actif	RE	86,7%
(Disponibilités + VMP) / Total Actif	LI	84,3%
Capitaux Propres / Total Actif	SF	83,8%
EBE / Total Actif	RE	80,9%
Trésorerie / Dette Court Terme	LI	78,2%
Total Dette / Total Actif	SF	74,5%
Variation Capitaux Propres Hors Résultat	AP	73,9%
Trésorerie / CA	RO	71,5%
BFR / CA	RO	68,5%
Total Dette / Capitaux Propres	SF	62,3%
CA / Capitaux Propres	EF	54,1%

Critère d'évaluation : Mesure Ordre 1		Fréquence
Recherche : Backward		de sélection
Total Dette / Capitaux Propres	SF	91,2%
Actif Circulant / Dette Court Terme	LI	84,7%
Variation Capitaux Propres Hors Résultat	AP	81,9%
Résultat Exploitation / Total Actif	RE	77,1%
EBE / Total Actif	RE	76,8%
ETE / CA	EF	70,5%
Trésorerie / Total Dette	LI1	68,4%
Capitaux Propres / Total Actif	SF	62,9%
BFR / CA	RO	45,0%
Dette Long Terme / Total Actif	SF	32,6%
Valeur Ajoutée / CA	EF	27,4%
(Disponibilités + VMP) / Total Actif	LI	24,1%

Parmi les 41 variables initiales, 18 apparaissent dans au moins 70 % des sélections, mais avec des fréquences très variables. Ainsi, seules 5 variables font partie d'au moins 3 modèles sur les 6, les 13 autres étant incluses dans 1 ou 2 modèles seulement.

On trouve alors en tête des sélections et par ordre décroissant de fréquence, les variables suivantes qui sont incluses dans au moins 3 modèles : EBE / Total Actif, Capitaux Propres / Total Actif,

Variation Capitaux Propres Hors Résultat, (Disponibilités + VMP) / Total Actif, Résultat Exploitation / Total Actif. Elles représentent 4 dimensions : apport, rentabilité, structure financière et liquidité. Le tableau ci-après donne l'ensemble des résultats, les variables étant classées en fonction du nombre d'occurrences de sélection. La partie droite précise leur rang d'apparition dans les modèles.

Tableau 34 : Fréquences de sélection des variables

Nombre de sélections			Rangs d'apparition dans les modèles					
EBE / Total Actif	RE	6	4	4	5	5	6	6
Capitaux Propres / Total Actif	SF	5	1	1	2	3	7	
Variation Capitaux Propres Hors Résultat	AP	5	1	3	3	4	7	
(Disponibilités + VMP) / Total Actif	LI	4	2	4	4	7		
Résultat Exploitation / Total Actif	RE	3	2	4	5			
Total Dette / Capitaux Propres	SF	2	1	2				
Trésorerie / Total Dette	LI	2	3	3				
Trésorerie / Dette Court Terme	LI	2	3	5				
Résultat Exploitation / CA	EF	2	5	6				
Trésorerie / CA	RO	2	7	8				
Résultat Net / Total Actif	RE	1	1					
Trésorerie / Total Actif	LI	1	1					
Actif Circulant / Dette Court Terme	LI	1	2					
Résultat Courant Av. I. / Capitaux Propres	RE	1	2					
(Disponibilités + VMP) / CA	LI	1	5					
ETE / CA	EF	1	6					
Total Dette / Total Actif	SF	1	6					
Clients / CA	RO	1	8					

Si l'on analyse plus précisément les variables retenues à l'aide de techniques propres au réseau, on obtient la hiérarchie indiquée ci-après.

Tableau 35 : Fréquences de sélection des variables à l'aide de techniques propres au réseau de neurones

Rang	Nombre de sélections		Rang	Nombre de sélections	
1	EBE / Total Actif	RE 3	7	Clients / CA	RO 1
1	Variation Capitaux Propres Hors Résultat	AP 3	7	ETE / CA	EF 1
3	Capitaux Propres / Total Actif	SF 2	7	Résultat Exploitation / CA	EF 1
3	(Disponibilités + VMP) / Total Actif	LI 2	7	Résultat Net / Total Actif	RE 1
3	Résultat Exploitation / Total Actif	RE 2	7	Trésorerie / CA	RO 1
3	Trésorerie / Dette Court Terme	LI 2	7	Total Dette / Capitaux Propres	SF 1
7	Actif Circulant / Dette Court Terme	LI 1	7	Total Dette / Total Actif	SF 1

Nous avons comparé cette dernière hiérarchie à une autre qui classe les variables selon leur capacité de discrimination mesurée au travers d'un F de *Fisher* exprimant le rapport entre la variance à l'intérieur des groupes et la variance entre les groupes. Ce comparatif est indiqué dans le tableau ci-après. Les variables y sont classées par ordre décroissant selon la valeur du F et nous avons indiqué pour chaque variable sélectionnée, le rang qu'elle occupe dans le tableau précédent.

Tableau 36 : Classement des variables selon un F de *Fisher*

	F	p-val.	Rang		F	p-val.	Rang
Résultat Exploitation / CA	220,15	0,000	7	BFR / Total Actif	27,21	0,000	
EBE / CA	219,49	0,000		BFR / CA	21,10	0,000	
Résultat Exploitation / Total Actif	218,96	0,000	3	ETE / Total Actif	19,40	0,000	
EBE / Total Actif	213,91	0,000	1	Dette Long Terme / Total Actif	19,32	0,000	
Résultat Net / Total Actif	210,01	0,000	7	Stocks / CA	16,00	0,000	
Capitaux Propres / Total Actif	207,59	0,000	3	Clients / CA	13,38	0,000	7
Total Dette / Total Actif	202,20	0,000	7	Marge Commerciale / CA	10,53	0,001	
Total Dette / Capitaux Propres	201,14	0,000	7	Résultat Courant Av. I. / Capitaux Propres	8,97	0,003	
Trésorerie / Total Actif	195,01	0,000		Actif Circulant / Total Actif	7,13	0,008	
Trésorerie / CA	179,60	0,000	7	Actif Circulant / CA	4,83	0,028	
Dette Court Terme / Total Actif	179,32	0,000		Frais Financiers / CA	4,04	0,045	
(Disponibilités + VMP) / Total Actif	171,62	0,000	3	(Actif Circulant - Stocks) / Total Actif	3,47	0,063	
Trésorerie / Dette Court Terme	168,19	0,000	3	Variation Autres Dettes	2,20	0,139	
Trésorerie / Total Dette	150,50	0,000		CA / Capitaux Propres	2,16	0,142	
(Disponibilités + VMP) / CA	145,63	0,000		Personnel / CA	0,62	0,431	
Actif Circulant / Dette Court Terme	133,77	0,000	7	Résultat Net / Capitaux Propres	0,20	0,651	
(Actif Circulant - Stocks) / Dette C. Terme	131,30	0,000		Dette Financière / CAF	0,18	0,669	
Fournisseurs / CA	85,95	0,000		Dette Long Terme / Capitaux Propres	0,17	0,681	
Valeur Ajoutée / CA	68,37	0,000		EBE / Capitaux Permanents	0,11	0,743	
Variation Capitaux Propres Hors Résultat	44,29	0,000	1	CA / Total Actif	0,02	0,878	
ETE / CA	28,57	0,000	7				

Les rangs indiqués dans ce tableau correspondent à ceux du tableau précédent et permettent de mettre en relation la capacité discriminante de chaque variable vue au travers d'un F de *Fisher* avec leur fréquence de sélection par un réseau de neurones. Les 6 variables qui apparaissent en tête du tableau 35 ont été surlignées en jaune dans celui-ci.

La première moitié du tableau ci-dessus comporte les variables au plus fort pouvoir de séparation, au sens du F et l'on y retrouve 13 variables sur les 14 sélectionnées grâce au réseau. Il existe donc bien une relation entre cette mesure de la capacité de discrimination et toutes les autres que nous avons employées. Cette relation est cependant assez grossière, car les deux hiérarchies obtenues sont très différentes. Par exemple, les 6 variables qui apparaissent dans le plus grand nombre de modèles neuronaux, EBE / Total Actif, Variation Capitaux Propres Hors Résultat, Capitaux Propres /

Total Actif, (Disponibilités + VMP) / Total Actif, Résultat Exploitation / Total Actif et Trésorerie / Dette Court Terme occupent respectivement les 4^{ème}, 20^{ème}, 6^{ème}, 12^{ème}, 3^{ème} et 13^{ème} position dans ce tableau. Inversement, certaines variables au fort pouvoir de séparation, comme EBE / CA, Trésorerie / Total Actif, Dette Court Terme / Total Actif, ou Trésorerie / Total Dette ne se retrouvent pas dans nos sélections finales. Seules Trésorerie / Total Actif, Dette Court Terme / Total Actif et Trésorerie / Total Dette apparaissent dans la liste des variables dont la fréquence de sélection est inférieure à 70 %. En conséquence, les présélections opérées par certains auteurs à l'aide de test F ou de test t , pour déterminer notamment les entrées d'un réseau, ont pu conduire à retenir des variables inutiles, mais surtout à éliminer des variables essentielles. Ce qui aurait pu être le cas, ici, de la variable Variation Capitaux Propres Hors Résultat, qui dispose d'une valeur de F assez faible, alors qu'il s'agit d'une variable essentielle pour le réseau. D'ailleurs, la sélection à l'aide d'un *Lambda de Wilks* a conduit à l'éliminer. Cependant, on constate aussi qu'en deçà d'un certain niveau de F , plus aucune variable ne réussit à se maintenir dans les sélections. Il n'y a d'ailleurs que Clients / CA qui apparaît une seule fois parmi les sélection finales.

Ces résultats sont à mettre en regard de ceux obtenus avec un *Lambda de Wilks*, décrits dans le tableau 33, p. 159, où les 11 variables qui font partie des sélections finales occupent des places comprises entre la 1^{ère} et la 16^{ème} dans le tableau 36.

5.2 STRUCTURE DES MODÈLES

Maintenant, lorsque l'on examine chaque ensemble de variables décrit dans le tableau 33, p. 159, on peut constater que le critère de sélection fondé sur un calcul de variance, traditionnellement utilisé conjointement avec l'analyse discriminante (*Lambda de Wilks*), donne une image plus édulcorée de la défaillance que les deux autres ensembles de critères (*vraisemblance* et critères spécifiques au réseau : *erreur*, mesures d'ordre 0 et 1) car fondée uniquement sur trois dimensions : liquidité, rentabilité et structure financière. Ces dimensions, et notamment la liquidité, sont d'ailleurs celles que l'on retrouve avec bon nombre de modèles créés à l'aide de ce critère, ce qui n'est pas étonnant puisque la définition juridique de la défaillance est une définition qui par construction privilégie la liquidité. Avec les deux autres ensembles, la situation est un peu différente. Le deuxième, qui s'appuie sur la *vraisemblance*, intègre la liquidité, la rentabilité et la structure financière de manière identique, mais les complète par trois autres dimensions : apport, efficacité et rotation. Le troisième, qui se fonde sur les critères employés avec le réseau, présente les mêmes caractéristiques que le deuxième, mais

les hiérarchise différemment. En effet, lorsque l'on dénombre les variables de chaque modèle selon leur appartenance à une dimension et que l'on tient compte de leur rang, on s'aperçoit que le deuxième ensemble (qui s'appuie sur le critère de *vraisemblance*) conduit à créer des modèles qui privilégient par ordre décroissant, la structure financière, les apports, la rentabilité et la liquidité, alors que le troisième (critères spécifiques au réseau), s'il place aussi en tête la structure financière, complète la liste par la rentabilité, la liquidité, puis les apports. Suivent ensuite, loin derrière, l'efficience et la rotation. Ces différences de hiérarchie tiennent peut-être en partie au fait que le nombre de modèles ne soit pas le même selon la famille du critère d'évaluation. Mais d'autres facteurs pourraient aussi expliquer dans ces différences. Tout se passe comme si les deux derniers ensembles de critères permettaient de « capter » des relations qui ont échappé au premier. Il est, à cet égard, symptomatique de constater que la variable Variation Capitaux Propres Hors Résultat, qui est particulièrement importante pour nombre d'entreprises de notre échantillon ayant subi une défaillance, ait complètement échappé au tout premier critère. C'est donc moins son importance intrinsèque qui explique sa présence, que le type de relation qu'elle peut entretenir avec d'autres variables et que le critère *Lambda de Wilks* a appréhendé de manière radicalement différente des autres. On voit bien ainsi, comment diverses méthodes de sélection, selon leur manière d'appréhender des linéarités et des non linéarités, façonnent ou révèlent des différences de structure. On voit aussi comment le critère de variance s'oppose aux deux autres qui proposent, eux, des modèles en définitive conceptuellement peu éloignés, car présentant sensiblement les mêmes dimensions. Peut-être que les résultats des modèles construits à l'aide de ces trois ensembles de critères confirmeront leur « proximité », mais cette fois en termes de résultats.

Si l'on regarde la façon dont les variables se répartissent entre les modèles, on peut discerner des points communs entre les trois ensembles de critères. Certains ratios se retrouvent en effet dans presque tous les modèles et sont communs à chaque famille de critères, et la plupart des autres sont partagés par au moins deux ensembles. Trois ratios reviennent donc avec les trois critères : $\text{EBE} / \text{Total Actif}$, $\text{Capitaux Propres} / \text{Total Actif}$ et $(\text{Disponibilités} + \text{VMP}) / \text{Total Actif}$. Ils montrent la place importante de la structure de capitalisation, mais aussi de l'état de la trésorerie parmi les facteurs de risque de défaillance. Ces ratios sont complétés différemment selon le critère employé : le premier (*Lambda de Wilks*) a tendance à privilégier essentiellement la place de la dette parmi les facteurs d'importance, alors que le deuxième et le troisième ont plutôt tendance à faire intervenir des ratios de rotation et d'évolution de la structure de capitalisation.

Le tableau 37 ci-après regroupe par catégorie de critères, les variables qui sont partagées par les différents modèles. Nous avons indiqué, entre parenthèses, le nombre d’occurrences de sélection de chacune d’entre elles.

Tableau 37 : Classement des variables selon le critère d’évaluation

Critère d’évaluation : Lambda de Wilks	Critère d’évaluation : Vraisemblance	Critère d’évaluation : Ordre 0-1 + Erreur
	AP Variation Capitaux Propres H. Ré. – (2)	AP Variation Capitaux Propres H. Ré. – (3)
RE EBE / Total Actif – (1)	RE EBE / Total Actif – (2)	RE EBE / Total Actif – (3)
SF Capitaux Propres / Total Actif – (1)	SF Capitaux Propres / Total Actif – (2)	SF Capitaux Propres / Total Actif – (2)
LI (Disponibilités + VMP) / Total Actif – (1)	LI (Disponibilités + VMP) / Total Actif – (1)	LI (Disponibilités + VMP) / Total Actif – (2)
RE Résultat Exploitation / Total Actif – (1)		RE Résultat Exploitation / Total Actif – (2)
SF Total Dette / Capitaux Propres – (1)		SF Total Dette / Capitaux Propres – (1)
LI Trésorerie / Total Dette – (1)	LI Trésorerie / Total Dette – (1)	
		LI Trésorerie / Dette Court Terme – (2)
	RO Trésorerie / CA – (1)	RO Trésorerie / CA – (1)
	EF Résultat Exploitation / CA – (1)	EF Résultat Exploitation / CA – (1)

Le socle commun à tous ces modèles est représenté par les trois ratios surlignés en jaune. Le ratio EBE / Actif est présent dans 6 modèles sur les 6 calculés. Il représente, par son numérateur, un indicateur de résultat qui ne tient compte ni de la manière dont l’entreprise est financée, ni de sa politique d’amortissement, ni des éléments exceptionnels de son activité. Il incarne en fait la part de création de valeur qui revient vraiment à l’exercice de l’activité. Ce ratio exprime donc la rentabilité économique de la firme.

Le ratio Capitaux Propres / Total Actif apparaît dans 5 modèles. Il désigne le rapport entre une des ressources, les capitaux propres, et leur utilisation, et mesure la capacité de l’entreprise à autofinancer ses investissements. On peut aussi le voir comme la capacité de résistance d’une firme aux difficultés survenant d’une concurrence âpre ou d’un retournement défavorable du marché.

Enfin, le ratio (Disponibilités + VMP) / Total Actif, qui revient dans 4 cas sur 6, incarne l’aisance de trésorerie et mesure l’aptitude de la firme à honorer ses engagements à court terme. On retrouve bien là, les caractéristiques fondamentales des PME, où la survie dépend de la rentabilité qui permet souvent d’assurer les engagements de court et de long terme, et de la structure capitalistique, qui permet aussi bien de financer le développement que les engagements de court terme lorsque le BFR est problématique. De fait, ce type d’entreprise ne peut compter que sur des ressources propres, les concours bancaires étant en général limités et le recours aux marchés financiers impensable.

Ce socle de 3 ratios peut être étendu à la variable Variation Capitaux Propres Hors Résultat qui indique dans quelle mesure la firme a fait appel à ses actionnaires pour compléter ses ressources, car elle est présente dans 5 modèles sur les 6, bien qu'elle ait échappé à la sélection via le *Lambda de Wilks*. Elle complète très bien les autres, puisque le rationnement du crédit que connaissent les petites entreprises ne peut être compensé, en cas de difficulté, que par une augmentation des fonds propres. Cette variable peut donc jouer dans deux directions : sur la capacité d'une entreprise à honorer ses dettes à court terme, si d'aventure elle connaît des problèmes de liquidité et de trésorerie et sur son aptitude à financer sa croissance.

À côté de ce socle, on trouve un pilier important, le ratio Résultat Exploitation / Total Actif, présent dans 3 modèles sur 6, mais il fait d'une certaine manière double emploi avec le ratio EBE / Total Actif, car il mesure sensiblement la même chose, les deux ratios étant très corrélés. Malgré ceci, il permet de pondérer le poids important de la dimension liée à la liquidité : sur les 41 variables que contiennent les 6 modèles, 11 sont des variables de liquidité et 11 de rentabilité. À la marge, on trouve enfin le ratio Total Dette / Capitaux Propres, présent dans 2 modèles, qui indique le poids relatif du financement de l'actif par l'emprunt et qui est dans ce registre, le pendant du ratio Capitaux Propres / Total Actif.

Toutes ces variables forment ainsi le tronc commun des indicateurs de défaillance de nos échantillons qui s'articulent autour de quatre dimensions : apport, rentabilité, structure financière et liquidité. D'autres variables de ces catégories, et notamment de liquidité, comme Total Dette Capitaux Propres, Trésorerie / Total Dette, Trésorerie / Dette Court Terme, Trésorerie / CA ou Résultats Exploitation / CA, sont aussi présentes dans les modèles, mais ne sont jamais compris dans plus de deux d'entre eux.

En marge des grandes dimensions très bien représentées (apport, rentabilité, structure financière et liquidité), on trouve des ratios de rotation et d'efficience. Ces deux dimensions apparaissent dans les modèles autres que celui sélectionné par un *Lambda de Wilks* et les ratios qui les représentent sont toujours placés, dans la hiérarchie des variables de chaque modèle, plutôt vers la fin. Leur fréquence de sélection, dans tous les cas, est légèrement supérieure à 70 % (et comprise entre 70,2 % et 76,1 %). Ce ne sont donc pas des variables stratégiques, comme les précédentes, mais elles ne sont pas pour autant totalement secondaires. En effet, 1 des 5 modèles concernés – celui calculé à l'aide du critère d'*erreur* – comporte 1 variable issue de chacune de ces catégories, sur les 8 qui le composent.

On retrouve ainsi, parmi les ratios de rotation, Trésorerie / CA et Clients / CA, et parmi les ratios d'efficacité, Résultat Exploitation / CA et ETE / CA. Ces indicateurs renforcent la place du cycle d'exploitation, évoquée précédemment, comme élément central dans l'analyse de la défaillance des PME. La rotation ici, met l'accent sur la capacité des entreprises à transformer leurs créances clients en liquidité, et donc sur les difficultés qui peuvent résulter de délais de transformation trop longs eu égard aux échéances des dettes, difficultés qui peuvent être renforcées par l'insuffisance de capitalisation. On voit alors, se profiler la fonction de la variable Variation Capitaux Propres Hors Résultat, qui permet d'ajuster ce décalage, l'entreprise pouvant se trouver dans une situation telle que seule une augmentation de ses fonds propres peut l'aider à financer ses dettes à court terme, à cause d'un BFR déséquilibré. L'efficacité, quant à elle, s'incarne ici dans l'aptitude de la firme à dégager une marge sur les ventes : elle renvoie à des indicateurs de rentabilité, mais exprimés comme une mesure de l'efficacité de l'exploitation, qui rejoignent les mesures effectuées à l'aide des ratios de rotation.

Plusieurs conclusions générales peuvent être tirées des résultats que nous venons de présenter.

Tout d'abord, les différences que l'on constate entre les sélections obtenues avec un critère de variance et celles issues des autres critères, montrent que les variables issues des premières constituent manifestement un sous ensemble des secondes. Ce constat s'applique aussi bien aux variables individuelles qu'aux dimensions qui leur sont sous jacentes.

Ensuite, les modèles construits à l'aide du critère de *vraisemblance* sont très proches de ceux élaborés avec les critères adaptés au réseau et il ne semble pas y avoir de relations d'inclusion d'un ensemble dans un autre, à tout le moins du point de vue des variables. Tout se passe comme si on était en présence d'une sorte de continuum qui a comme point de départ, une modélisation linéaire de relations entre ratios financiers et à laquelle certaines méthodes viennent y greffer des dépendances plus ou moins non linéaires, mais qui ne sont pas de nature à remettre en cause la structure de base.

Enfin, tous les modèles présentés montrent qu'il existe bien des indicateurs fondamentaux de la défaillance qui ne sont pas que conjoncturels et dépendant des seules spécificités d'une méthode, ce que la littérature jusqu'à présent n'avait pas fait apparaître.

La problématique qui se pose maintenant concerne la relation entre la performance des modèles et leur structure. Les différences comme les points communs que nous avons relevés se retrouvent-ils dans les résultats et sous quelles formes ? En particulier, la relation d'inclusion

que nous avons illustrée précédemment se traduit-elle par une hiérarchie des résultats qui suit le sens de cette relation ? Et l'apparente similarité de certains modèles conduit-elle à des performances identiques ou relativement proches ? C'est à toutes ces questions que nous allons maintenant tenter d'apporter des réponses.

5.3 PERFORMANCE DES MODÈLES

Afin d'apprécier le mieux possible la capacité des modèles à discriminer correctement les entreprises de nos échantillons et à illustrer l'apport d'un processus de sélection, nous avons évalué leur performance de différentes manières.

Pour disposer de points de repères, nous avons, dans un premier temps, analysé le degré de séparabilité des échantillons à l'aide de ratios tirés de manière aléatoire. Pour chaque ensemble ainsi constitué, nous avons évalué les modèles issus d'une analyse discriminante, d'une régression logistique et d'un réseau de neurones, puis moyenné les résultats. Ceci permet de mesurer la distance entre le hasard et un processus déterministe et d'en déduire l'économie de ce dernier. En effet, si l'écart est faible, on peut penser qu'il n'a pas de raison d'être, et plus il est important, plus sa valeur ajoutée est forte.

Dans un deuxième temps, nous avons calculé ce même type de performance, mais cette fois en incluant toutes les variables présélectionnées, c'est-à-dire les 41 prédicteurs décrits dans le tableau 31, p. 151. Cette mesure permet d'apprécier l'intérêt d'un élagage, et donc d'une diminution du nombre de dimensions prises en compte dans un modèle.

Dans un troisième et dernier temps, nous avons calculé la performance des 6 ensembles de variables décrits dans le tableau 33, p. 159, à la fois sur les échantillons de tests et les échantillons de validation. Et cette performance a été évaluée au travers de modèles issus, là aussi, d'une analyse discriminante, d'une régression logistique et d'un réseau de neurones. Nous avons procédé ainsi, pour étudier les comportements des modèles selon les techniques de sélection de variables employées et analyser les points de compatibilité. Par exemple, quelle différence existe-t-il entre deux modèles neuronaux, l'un composé de variables retenues grâce à un *Lambda de Wilks*, l'autre grâce à un ensemble choisi à l'aide d'une mesure d'ordre 0 ou 1 ? Inversement, qu'en est-il d'un modèle « logistique » par rapport à un modèle « discriminant » créés à partir des mêmes variables ?

5.3.1 PERFORMANCE MESURÉE À L'AIDE DE VARIABLES TIRÉES AU HASARD

Les données sont-elles aisément séparables ? Pour répondre à cette question, nous avons tiré au hasard 50 ensembles de ratios et calculé les taux de classement à partir d'échantillons *bootstrapés* comme indiqué au cours du paragraphe 4.4, p. 145, mais avec une procédure simplifiée. Pour des raisons liées aux temps de calcul, nous avons procédé à 25 *bootstraps* répliqués 40 fois, contre 25 *bootstraps* répliqués 100 fois. En limitant le nombre de répliqués, l'estimation de l'erreur de chaque ensemble voit sa variance augmenter, mais en même temps ceci nous a permis de tester plus de combinaisons de variables que si nous avions conservé les paramètres initiaux et donc de diminuer la variance de l'erreur entre les différents ensembles de variables. Les résultats du tableau 38 correspondent aux moyennes obtenues. Dans ce tableau, et dans ceux qui suivent, les entreprises saines seront désignées par l'intitulé « S. », et les défaillantes par « Déf. ».

Tableau 38 : Résultats des prévisions obtenus à partir de ratios tirés aléatoirement et calculés à l'aide des échantillons de test

		Analyse discriminante		Régression logistique			Réseau de neurones				
		Prévu		Prévu			Prévu				
		S.	Déf.				S.	Déf.			
Actuel	S.	83,22%	16,78%	Actuel	S.	82,39%	17,61%	Actuel	S.	83,99%	16,01%
	Déf.	26,38%	73,62%		Déf.	23,12%	76,88%		Déf.	21,55%	78,45%
		Cl. Corrects ³⁹		Cl. Corrects		79,64%		Cl. Corrects		81,22%	

Ces résultats montrent que les échantillons ne sont pas facilement séparables, puisque les taux de mauvais classements sont voisins de 20 %. Pour autant, ces taux ne sont pas fondamentalement mauvais pour des ensembles aléatoires, montrant par là, que les ratios initiaux que nous avons choisis possèdent dans l'ensemble d'assez bonnes capacités de discrimination.

5.3.2 PERFORMANCE MESURÉE À L'AIDE DE TOUTES LES VARIABLES INITIALES

Quel écart existe-t-il entre un ensemble de variables aléatoires et l'ensemble de toutes les variables ? Les résultats sont indiqués ci-après. L'inclusion dans les modèles des 41 variables initiales permet d'améliorer sensiblement les résultats, mais au prix d'une complexité nettement accrue. Dans les tableaux 38 et 39, le réseau présente des performances supérieures aux deux autres méthodes, mais nous n'avons pas effectué de tests de différences de moyennes comparant les différents résultats, car ils ne servent pas directement notre argumentation.

³⁹ Les tableaux de résultats comportent tous une ligne intitulée « Cl. Corrects » qui indique le taux global de classements corrects.

Nous reviendrons sur les résultats du tableau 39 ultérieurement pour analyser dans quelle mesure une sélection a pu ou non améliorer ces pourcentages tout en réduisant la dimension des modèles.

Tableau 39 : Résultats des prévisions obtenus à partir de tous les ratios et calculés à l'aide des échantillons de test

Analyse discriminante				Régression logistique			Réseau de neurones								
Prévu				Prévu			Prévu								
	S.	Déf.		S.	Déf.	S.	Déf.								
Actuel	S.	93,56%	6,44%	Actuel	S.	91,18%	8,82%	Actuel	S.	93,60%	6,40%				
	Déf.	22,28%	77,72%		Déf.	18,24%	81,76%		Déf.	13,06%	86,94%				
Cl. Corrects				85,64%			Cl. Corrects			86,47%					
										Cl. Corrects			90,27%		

5.3.3 PERFORMANCE MESURÉE SELON LES COUPLES « MÉTHODE DE CLASSIFICATION – TECHNIQUE DE SÉLECTION »

Après avoir mesuré les résultats obtenus à partir de combinaisons aléatoires de variables, puis de toutes les variables, nous avons ensuite étudié les relations entre les techniques de sélection et les méthodes de classification. Il s'agit de déterminer s'il existe des couples qui « performant » mieux que d'autres et d'étudier notamment, comment se comporte un réseau avec des variables qui ont été choisies grâce à des critères plutôt optimisés pour d'autres méthodes.

Nous avons d'abord mesuré les performances pour les couples « méthode de classification – technique de sélection », mais uniquement pour les combinaisons où le critère d'évaluation est optimisé pour le classifieur. Nous avons comparé les résultats des couples « analyse discriminante – *Lambda de Wilks* », « régression logistique – critère de *vraisemblance* (en employant 2 procédures de recherche) » et « réseau – mesures d'*ordre 0* et *1* et critère d'*erreur* », soit 6 combinaisons. Les résultats sont indiqués ci-après, en commençant par l'analyse discriminante, suivie de la régression logistique puis du réseau. Nous conserverons cet ordre par la suite.

Tableau 40 : Résultats des prévisions obtenus pour différents couples « méthodes de classification – technique de sélection » et calculés à l'aide des échantillons de test

Analyse discriminante			
Prévu			
	S.	Déf.	
Actuel	S.	91,20%	8,80%
	Déf.	16,80%	83,20%
Cl. Corrects		87,20%	
Critère : Lambda de Wilks			
Recherche : Stepwise			

Régression logistique

		Prévu		Prévu	
		S.	Déf.	S.	Déf.
Actuel	S.	93,60%	6,40%	S.	89,56%
	Déf.	9,58%	90,42%	Déf.	11,16%
Cl. Corrects		92,01%		89,20%	
Critère : Vraisemblance Recherche : Backward Step.					

Réseau de neurones

		Prévu		Prévu		Prévu	
		S.	Déf.	S.	Déf.	S.	Déf.
Actuel	S.	92,78%	7,22%	S.	91,96%	S.	92,82%
	Déf.	4,72%	95,28%	Déf.	4,78%	Déf.	7,18%
Cl. Corrects		94,03%		93,59%		92,82%	
Critère : Erreur Recherche : Backward		Critère : Mesure Ordre 0 Recherche : Backward		Critère : Mesure Ordre 1 Recherche : Backward			

Tests de différence des taux de classements corrects

Méthodes	RL VB	RL VF	RN E	RN 0	RN 1
AD W	0,0000	0,0000	0,0000	0,0000	0,0000
RL VB		0,0000	0,0000	0,0000	0,0000
RL VF			0,0000	0,0000	0,0000
RN E				0,1221	0,0000
RN 0					0,0080

Les valeurs de ce tableau correspondent aux p-valeurs de tests de différence des taux de classements corrects effectués à l'aide d'un *t* de *Student*. Les acronymes ont les significations suivantes :

AD W :	Analyse Discriminante – Lambda de Wilks
RL VB :	Régression Logistique – Vraisemblance – Backward Step.
RL VF :	Régression Logistique – Vraisemblance – Forward Step.
RN E :	Réseau de Neurones – Erreur
RN 0 :	Réseau de Neurones – Ordre 0
RN 1 :	Réseau de Neurones – Ordre 1

L'analyse discriminante a tendance à surclasser les entreprises saines au détriment des défaillantes, montrant pas là que, du point de vue du critère de séparation employé, il y a plus de firmes défaillantes qui ont un profil similaire à celui des saines, qu'il n'y a d'entreprises saines qui ressemblent à des défaillantes. La frontière entre les classes définies par la discriminante est donc telle qu'un sous-groupe important d'entreprises défaillantes se situe du même côté que la majorité des saines, alors que dans le cas inverse, le sous-groupe de firmes saines qui se situe du côté de la majorité des défaillantes est plus restreint. Cette situation se rencontre souvent dans

la littérature, notamment lorsque le critère retenu pour définir le *cut* est celui qui maximise le taux général de bons classements, l'asymétrie frappant plutôt le groupe des défaillantes. On peut penser que les firmes saines présentent des profils plus variés que les défaillantes et qu'en ce sens elles sont moins homogènes : il est en effet plus plausible de considérer que des entreprises en mauvaise santé ont plus de chance d'exister que d'autres en bonne santé de disparaître. C'est vraisemblablement cette propriété qui explique la façon dont ces deux groupes se chevauchent. Nous analyserons cette question plus en détail avec nos échantillons dans le chapitre 6.

Avec l'analyse discriminante, le taux moyen de classements correctement effectués, qui est de 87,2 %, est à peine inférieur au taux moyen que l'on relève dans la littérature, ce qui montre que notre échantillon est aussi difficilement séparable avec cette méthode que dans les autres études. La régression logistique, elle, donne de meilleurs résultats, avec une légère tendance à sur-classer les saines, mais d'une amplitude nettement inférieure à ce que l'on observe avec la discriminante : 92,01 % et 89,20 % selon le modèle. Quant au réseau de neurones, il obtient des taux classements corrects légèrement supérieurs à la régression logistique ; 94,03 %, 93,59 % et 92,82 % selon le modèle. Ces derniers résultats sont plutôt bons, compte tenu de ce que l'on retrouve dans la littérature, mais ici le réseau est quelque peu moins performant comparativement à la régression logistique, l'écart entre les deux étant généralement plus important.

Pour chaque paire de méthodes, nous avons effectué un test de différence de moyenne à l'aide d'un *t* de *Student*, afin de caractériser l'amplitude des écarts observés entre les différents taux de bons classements. Les résultats des tests sont présentés à la fin du tableau 40. Dans tous les cas, sauf un, ces écarts sont particulièrement significatifs, montrant par là que les modèles ne disposent pas d'une précision statistiquement identique.

Les résultats du tableau 40 ayant été calculés à partir des échantillons de test, nous les avons recalculés sur la base des échantillons de validation, selon une même procédure de calcul, afin de tester la réelle capacité de généralisation des modèles. Les résultats obtenus, indiqués dans le tableau 41, sont à peu près similaires à ceux du tableau 40, même si l'on note une diminution générale des taux de bons classements, sauf dans le cas d'un des réseaux, dont le taux chute légèrement. Ceci montre que la procédure de sélection des entreprises que nous avons adoptée et qui visait à construire deux profils volontairement différenciés d'entreprises, opposant des entreprises vraiment saines aux défaillantes, a été plutôt favorable au réseau en lui permettant de bien discriminer les firmes de chaque catégorie, ce qui n'a pas été le cas de l'analyse discriminante.

Entre les deux, la régression logistique obtient de très bonnes performances, à tel point que l'un des deux modèles dont elle est à l'origine obtient de meilleurs résultats que l'un des modèles neuronaux. Ici, les écarts entre les taux de classements de chaque paire de méthodes ne sont pas tous significatifs, surtout lorsque les comparaisons ne se font pas avec l'analyse discriminante.

Tableau 41 : Résultats des prévisions obtenus pour différents couples « méthodes de classification – technique de sélection » et calculés à l'aide des échantillons de validation

Analyse discriminante											
Prévu											
S. Déf.											
Actuel	S.	89,62%	10,38%								
	Déf.	19,23%	80,77%								
Cl. Corrects		85,19%									
Critère : Lambda de Wilks Recherche : Stepwise											
Régression logistique											
Prévu				Prévu							
S. Déf.				S. Déf.							
Actuel	S.	91,15%	8,85%	Actuel	S.	88,85%	11,15%				
	Déf.	9,62%	90,38%		Déf.	11,54%	88,46%				
Cl. Corrects			90,77%	Cl. Corrects		88,65%					
Critère : Vraisemblance Recherche : Backward Step.				Critère : Vraisemblance Recherche : Forward Step.							
Réseau de neurones											
Prévu			Prévu			Prévu					
S. Déf.			S. Déf.			S. Déf.					
Actuel	S.	93,08%	6,92%	Actuel	S.	92,69%	7,31%	Actuel	S.	91,15%	8,85%
	Déf.	5,38%	94,62%		Déf.	8,08%	91,92%		Déf.	11,15%	88,85%
Cl. Corrects			93,85%	Cl. Corrects		92,31%		Cl. Corrects			90,00%
Critère : Erreur Recherche : Backward				Critère : Mesure Ordre 0 Recherche : Backward				Critère : Mesure Ordre 1 Recherche : Backward			
Test de différence des taux de classements corrects											
Méthodes	RL VB	RL VF	RN E	RN 0	RN 1						
AD W	0,0028	0,0489	0,0000	0,0001	0,0093						
RL VB		0,1308	0,0313	0,1864	0,3370						
RL VF			0,0015	0,0223	0,2410						
RN E				0,1642	0,0114						
RN 0					0,0950						

Les valeurs de ce tableau correspondent aux p-valeurs de tests de différence des taux de classements corrects effectués à l'aide d'un test de différence de proportion. Les acronymes ont les significations suivantes :

AD W :	Analyse Discriminante – Lambda de Wilks
RL VB :	Régression Logistique – Vraisemblance – Backward Step.
RL VF :	Régression Logistique – Vraisemblance – Forward Step.
RN E :	Réseau de Neurones – Erreur
RN 0 :	Réseau de Neurones – Ordre 0
RN 1 :	Réseau de Neurones – Ordre 1

Les procédures de choix des échantillons employées dans la littérature ne s'intéressent en général pas au profil des entreprises⁴⁰ et conduisent souvent à des écarts moins grands entre les réseaux et l'analyse discriminante. Manifestement, la procédure que nous avons adoptée, et qui visait à s'assurer que parmi les entreprises saines choisies il n'en existait pas présentant une situation financière difficile, a été plutôt favorable au réseau. Avec les échantillons de validation, l'analyse discriminante a donné des prévisions correctes dans 85,19 % des cas, la régression logistique, entre 88,65 % et 90,77 % des cas, et le réseau, entre 90,00 % et 93,85 %. La performance de chaque méthode s'interprète bien évidemment par rapport à celle des autres, lorsque mesurée sur des échantillons de validation, mais aussi par rapport aux gains de précision obtenus sur des ensembles de variables aléatoires ou des ensembles les comprenant toutes. Ainsi, l'analyse discriminante, avec un ensemble dûment sélectionné, permet d'augmenter de presque 9 % le taux de prévisions correctes par rapport à un ensemble aléatoire, et seulement d'un peu plus de 1,5 % par rapport à un ensemble contenant les 41 variables initiales (ces taux ont été calculés sur les échantillons de test uniquement). La régression logistique, dans le premier cas, permet de gagner entre presque 10 % et un peu plus de 12%, et dans le second, entre plus de 2,5 % et plus de 5,5 %. Quant au réseau, ses résultats progressent entre presque 12 % et presque 13 %, dans la première situation, et entre un peu plus de 2,5 % et un peu plus de 3,5 %, dans la seconde. Globalement, toutes les méthodes tirent un réel bénéfice des processus de sélection, mais aussi de réduction de l'espace des variables. En effet, la taille des modèles passe de 41 variables à 8 ou moins de 8 et leurs performances dans le même temps augmentent légèrement par rapport à un modèle complet. Lorsque l'on détaille les améliorations par type de méthode, on voit que leurs ordres de grandeur sont à peu près les mêmes. Les performances des modèles sont en définitive liées d'abord aux caractéristiques des méthodes, et aucune d'entre elles ne semblent mieux tirer partie qu'une autre, d'un processus de réduction.

⁴⁰ Très peu d'études ont effectivement pris en compte la recommandation de Malecot (1991) qui suggère de sélectionner des entreprises réellement saines. Hormis dans le cas des travaux qui visent à discriminer plusieurs catégories d'entreprises saines, dès lors qu'il s'est agi simplement de séparer des firmes en bonne santé de firmes défaillantes, à notre connaissance, seul Taffler (1983) a appliqué cette règle en tenant un raisonnement identique à celui de Malecot (1991).

À présent, venons-en à l'une des questions initiales que nous posions et qui est relative à la pertinence des procédures de sélection de variables généralement employées avec les réseaux dans un grand nombre d'études expérimentales. Pour étudier cette question, nous avons employé chacun des 6 ensembles de ratios, rappelons-le plutôt optimisé pour une méthode en particulier, avec les méthodes pour lesquelles le critère de sélection n'est pas optimum. Ceci revient à se demander, par exemple, si un ensemble de ratios choisi grâce à un *Lambda de Wilks* peut conduire à un modèle neuronal performant et inversement, si un ensemble de ratios optimisé pour un réseau peut donner de bons résultats avec une analyse discriminante. Les résultats de ces modèles sont donnés dans les tableaux qui suivent. Le tableau 42 décrit les résultats obtenus à l'aide des variables sélectionnées avec un *Lambda de Wilks*, le tableau 43 fait de même mais avec les 2 ensembles de variables choisies grâce au critère de *vraisemblance* et le tableau 44 complète ces résultats par ceux obtenus avec les 3 ensembles de variables optimisés pour le réseau.

Le tableau 42 montre que les variables choisies par un *Lambda de Wilks* conduisent à des résultats qui ne sont pas très bons, car avec l'analyse discriminante, le taux de bons classements qui est de 87,20 %, chute largement avec les deux autres méthodes. Les tests d'égalité des classements montrent d'ailleurs que la discriminante obtient des résultats qui sont significativement différents des autres. Le critère employé ici est fondé sur des hypothèses qui correspondent parfaitement à la logique de calcul de l'analyse discriminante. En ce sens, il n'est pas réellement étonnant de constater que des variables qui ne peuvent produire qu'un modèle discriminant aux performances moyennes, même si elles présentent individuellement un fort pouvoir de séparation au sens de ce critère, ne sont pas en mesure de donner de bons résultats avec d'autres méthodes de classification. Et c'est bien le cas puisque la régression logistique, comme le réseau, obtiennent un taux presque identique. Ce critère est donc loin d'être propice au réseau.

Tableau 42 : Résultats des prévisions obtenus par chaque méthode de classification et calculés à l'aide des échantillons de test avec les variables sélectionnées par :
Critère : Lambda de Wilks – Procédure de recherche : Stepwise

		Analyse discriminante		Régression logistique			Réseau de neurones		
		Prévu		Prévu			Prévu		
		S.	Déf.	S.		Déf.	S.		Déf.
Actuel	S.	91,20%	8,80%	S.	88,06%	11,94%	S.	90,02%	9,98%
	Déf.	16,80%	83,20%	Déf.	20,82%	79,18%	Déf.	22,80%	77,20%
Cl. Corrects		87,20%		Cl. Corrects		83,62%	Cl. Corrects		83,61%

Test de différence des taux de classements corrects		
Méthodes	RL	RN
AD	0,0000	0,0000
RL		0,9831

Les valeurs numériques correspondent aux p-valeurs d'un test *t* de Student

Tableau 43 : Résultats des prévisions obtenus par chaque méthode de classification et calculés à l'aide des échantillons de test avec les variables sélectionnées par :
Critère : Vraisemblance – Procédure de recherche : Backward Stepwise

Analyse discriminante				Régression logistique				Réseau de neurones			
Prévu				Prévu				Prévu			
	S.	Déf.		S.	Déf.		S.	Déf.		S.	Déf.
Actuel	S.	87,28%	12,72%	Actuel	S.	93,60%	6,40%	Actuel	S.	89,68%	10,32%
	Déf.	15,16%	84,84%		Déf.	9,58%	90,42%		Déf.	7,26%	92,74%
Cl. Corrects			86,06%	Cl. Corrects			92,01%	Cl. Corrects			91,21%

Test de différence des taux de classements		
Méthodes	RL	RN
AD	0,0000	0,0000
RL		0,0015

Les valeurs numériques correspondent aux p-valeurs d'un test *t* de Student

Critère : Vraisemblance – Procédure de recherche : Forward Stepwise

Analyse discriminante				Régression logistique				Réseau de neurones			
Prévu				Prévu				Prévu			
	S.	Déf.		S.	Déf.		S.	Déf.		S.	Déf.
Actuel	S.	87,98%	12,02%	Actuel	S.	89,56%	10,44%	Actuel	S.	88,08%	11,92%
	Déf.	17,58%	82,42%		Déf.	11,16%	88,84%		Déf.	8,86%	91,14%
Cl. Corrects			85,20%	Cl. Corrects			89,20%	Cl. Corrects			89,61%

Test de différence des taux de classements		
Méthodes	RL	RN
AD	0,0000	0,0000
RL		0,1711

Les valeurs numériques correspondent aux p-valeurs d'un test *t* de Student

Tableau 44 : Résultats des prévisions obtenus par chaque méthode de classification et calculés à l'aide des échantillons de test avec les variables sélectionnées par :

Critère : Erreur – Procédure de recherche : Backward

		Analyse discriminante		Régression logistique			Réseau de neurones		
		Prévu		Prévu			Prévu		
		S.	Déf.	S.		Déf.	S.		Déf.
Actuel	S.	83,38%	16,62%	S.	90,44%	9,56%	S.	92,78%	7,22%
	Déf.	14,62%	85,38%	Déf.	10,44%	89,56%	Déf.	4,72%	95,28%
Cl. Corrects		84,38%		Cl. Corrects		90,00%	Cl. Corrects		94,03%

Test de différence des taux de classements

Méthodes	RL	RN
AD	0,0000	0,0000
RL		0,0000

Les valeurs numériques correspondent aux p-valeurs d'un test *t* de *Student*

Critère : Mesure Ordre 0 – Procédure de recherche : Backward

		Analyse discriminante		Régression logistique			Réseau de neurones		
		Prévu		Prévu			Prévu		
		S.	Déf.	S.		Déf.	S.		Déf.
Actuel	S.	83,20%	16,80%	S.	86,38%	13,62%	S.	91,96%	8,04%
	Déf.	16,00%	84,00%	Déf.	10,36%	89,64%	Déf.	4,78%	95,22%
Cl. Corrects		83,60%		Cl. Corrects		88,01%	Cl. Corrects		93,59%

Test de différence des taux de classements

Méthodes	RL	RN
AD	0,0000	0,0000
RL		0,0000

Les valeurs numériques correspondent aux p-valeurs d'un test *t* de *Student*

Critère : Mesure Ordre 1 – Procédure de recherche : Backward

		Analyse discriminante		Régression logistique			Réseau de neurones		
		Prévu		Prévu			Prévu		
		S.	Déf.	S.		Déf.	S.		Déf.
Actuel	S.	87,06%	12,94%	S.	88,16%	11,84%	S.	92,82%	7,18%
	Déf.	18,16%	81,84%	Déf.	9,78%	90,22%	Déf.	7,18%	92,82%
Cl. Corrects		84,45%		Cl. Corrects		89,19%	Cl. Corrects		92,82%

Test de différence des taux de classements		
Méthodes	RL	RN
AD	0,0000	0,0000
RL		0,0000

Les valeurs numériques correspondent aux p-valeurs d'un test *t* de Student

Les ensembles de variables créés par le critère de *vraisemblance* donnent, avec l'analyse discriminante, des résultats sensiblement moins bons qu'avec la régression logistique ; respectivement 86,06 % contre 92,01 %, avec une recherche de type *backward stepwise*, et 85,20 % contre 89,20 % avec une recherche *forward stepwise*. En revanche, avec le réseau, ces deux ensembles donnent d'assez bons résultats, respectivement 91,21 % et 89,61% – tableau 43 –, qui sont très proches de ceux issus de la régression logistique. D'ailleurs, dans un des deux cas, le réseau fait mieux que la logistique.

Les résultats que l'on observe avec le critère de *vraisemblance*, où les modèles logistiques et neuronaux sont assez proches en termes de performances, on ne les retrouve pas avec les critères propres au réseau. Ainsi, les variables sélectionnées avec le critère d'*erreur*, donnent 94,03 % de bons classements avec un réseau, 90,00 % avec une régression logistique, et seulement 84,38 % avec une analyse discriminante – tableau 44. L'écart entre les trois méthodes de classification est sensiblement le même avec une mesure d'*ordre 0*, avec respectivement 93,59 %, 88,01 % et 83,60 % de bons classements, mais se réduit un peu avec un mesure d'*ordre 1*, avec respectivement, 92,82 %, 89,19 % et 84,45 % de bons classements.

Le réseau présente donc des performances nettement supérieures aux autres méthodes, surtout avec le critère d'*erreur* et la mesure d'*ordre 0*. Il est somme toute assez peu étonnant de constater que le critère d'*erreur* produise les meilleurs résultats, puisqu'il incarne à la fois le critère d'évaluation de la pertinence des variables et la mesure de cette pertinence, ce qui est le propre des *wrappers*, où l'algorithme de classification intervient directement dans la sélection des variables. Ceci est donc conforme à ce que l'on peut attendre. Quant à la hiérarchie obtenue avec les deux autres critères, la mesure d'*ordre 0* conduisant à des résultats qui dépassent ceux issus de la mesure d'*ordre 1*, elle semble être purement conjoncturelle, rien *a priori* ne pouvant laisser supposer que tel critère est susceptible d'être plus performant que tel autre. Les modèles neuronaux avec des variables appropriées dominant donc tous les résultats. Ceci étant, ceux issus d'une régression logistique arrivent à mieux tirer partie des informations contenues dans

les données que ne le font ceux issus d'une discriminante, pour les 3 ensembles de variables choisis par le réseau. D'ailleurs, les variables retenues grâce au critère d'*erreur* conduisent à un modèle logistique qui classe correctement 90,00 % des observations, contre 94,03 %, pour le modèle neuronal, laissant loin derrière le modèle discriminant avec un taux de succès de 84,38 %.

Tous les calculs de performance confirment d'une certaine façon les conclusions que nous tirions de la structure des modèles au cours du paragraphe 5.2. D'un côté, l'image de la défaillance renvoyée par l'ensemble de variables sélectionné par un *Lambda de Wilks* apparaissait, du point de vue des dimensions sous-jacentes, moins complexe et moins multiforme que celles renvoyées par les deux autres ensembles. L'on retrouve bien la traduction de ceci dans les résultats des prévisions ; non seulement cet ensemble, lorsqu'il est utilisé avec une analyse discriminante, est peu performant, mais en plus, il ne donne pas de résultats particulièrement probants avec les autres méthodes. De l'autre, les similarités entre les ensembles de variables obtenus avec le critère de *vraisemblance* et ceux créés à l'aide des critères adaptés au réseau semblent aussi se retrouver dans des taux de prévisions finalement assez proches : les meilleurs modèles de chaque famille de critères n'ont qu'un peu plus de 3 points d'écart avec les données de validation et tous ces ensembles de variables donnent de bons résultats, aussi bien avec la régression logistique qu'avec le réseau. Néanmoins, les écarts constatés tiennent vraisemblablement plus aux spécificités des méthodes qu'au seul contenu des variables. On voit bien comment le réseau peut faire jeu égal avec la régression logistique en présence de variables qui ne procèdent pas d'un choix optimum pour lui, et en même temps comment il peut la dépasser lorsque ce choix s'appuie sur un critère qui lui est spécifique.

La précision d'un modèle vaut donc pour partie en raison des qualités intrinsèques de la méthode de classification qui a permis de l'élaborer et pour partie, en raison de l'adéquation entre les propriétés de cette méthode et celles de la technique de sélection de variables mise en œuvre. Dans le domaine de la prévision de la défaillance, toutes les études portant sur des échantillons un tant soit peu conséquents montrent que les ratios entretiennent bien des relations non linéaires et qu'il est difficilement envisageable de vouloir construire des modèles précis dès lors que cette propriété fondamentale ne peut être prise en compte. Si les techniques de sélection généralement utilisées avec la régression logistique ne procèdent pas d'un mauvais choix pour construire des modèles neuronaux, il n'en est donc pas de même de celles optimisées pour l'analyse discriminante. À tout le moins faut-il envisager conjointement d'autres solutions.

5.4 CONCLUSION DU CHAPITRE

Ce chapitre a permis de faire ressortir plusieurs caractéristiques des modèles de défaillance que nous avons élaborés. En premier lieu, les sélections ou pré-sélections de variables destinées à alimenter un réseau, dès lors qu'elles sont effectuées avec des tests paramétriques de différences de moyennes, risquent de conduire à l'élimination de variables importantes, si elles ne sont pas complétées par d'autres types de tests. Bien que ces tests, dans l'ensemble, donnent une idée assez générale des capacités de discrimination d'une variable, notamment lorsqu'elle présente un très fort ou un très faible pouvoir de séparation au sens du test, ils demeurent imprécis pour toutes celles qui se situe entre ces deux extrémités. On arrive aux mêmes conclusions lorsque l'on observe les résultats d'une procédure de sélection de type *stepwise* qui s'appuie sur un *Lambda de Wilks*, où les points communs entre ces derniers et ceux que procurent des méthodes spécifiques au réseau n'apparaissent qu'en très petit nombre. Ceci montre bien les limites d'un critère fondé sur une analyse de la variance et son incompatibilité avec les propriétés d'une méthode neuronale. En revanche, on trouve une réelle ressemblance entre les ensembles de variables choisis avec un critère de *vraisemblance* et ceux choisis par les critères dédiés au réseau.

En deuxième lieu, les différences et les similarités relevées précédemment se retrouvent entre les mêmes ensembles de variables, mais cette fois dans l'image qui est restituée du phénomène de la défaillance. Le premier ensemble, créé via un *Lambda de Wilks*, donne corps à une vision de la défaillance conceptuellement plus pauvre que celles renvoyées par les deux autres qui eux, en revanche, permettent de dresser une sorte de portrait robot de ce phénomène, bien sûr propres aux firmes présentes dans les échantillons utilisés ; ainsi, ils mettent particulièrement en évidence la place centrale du cycle d'exploitation et du BFR dans le processus qui nous intéresse. Une entreprise qui se maintient en activité est une entreprise rentable économiquement et liquide. Cette rentabilité est stratégique en cas de difficulté car les sources de financement externes sont, pour ce type d'entreprises, limitées. L'on voit bien, au travers de ces modèles, comment les firmes qui ne peuvent pas assurer correctement leurs engagements, procèdent pour tenter de s'en sortir, prioritairement via leurs actionnaires. Avec le premier ensemble de variables, la place du cycle d'exploitation, qui est au centre de ce portrait robot, n'est pas aussi prononcée, et la fonction stratégique des actionnaires comme recours en cas de difficultés, cette fois est passée complètement sous silence.

En troisième lieu, les modèles élaborés sur la base de ces ensembles de variables voient leurs performances tributaires en définitive à la fois de la complétude de l'image qu'ils incarnent de la défaillance et de la technique de classification en jeu. En effet, quelle que soit la méthode de classification utilisée, l'ensemble de variables créé par un *Lambda de Wilks* donne les moins bons résultats. Ceci explique bien cela. Quant aux autres, ils fonctionnent bien avec la régression logistique et le réseau, mais avec un petit avantage pour le réseau lorsqu'il est couplé avec un ensemble optimisé pour son utilisation.

CHAPITRE 6 : APPORT DES MODÈLES À LA COMPRÉHENSION DE LA DÉFAILLANCE

La défaillance ne participant pas d'une logique de dégradation qui s'impose de façon identique à toute entreprise, quelques auteurs ont essayé d'isoler les processus pouvant y conduire. Ainsi, de la même manière que nous avons cherché des constantes parmi des ensembles de prédicteurs choisis de différentes manières, nous avons aussi étudié dans quelle mesure ces ensembles permettaient d'appréhender des constantes dans les chemins empruntés par les entreprises lorsqu'elles sont conduites vers la défaillance, mais aussi dans ceux qu'elles utilisent pour demeurer en activité. Nous avons d'abord classé les entreprises, au travers de profils et ce, pour chacun des 6 ensembles de variables. Puis nous avons recherché, dans tous les sous groupes ainsi obtenus, si des points communs dans la logique financière déployée d'un côté par les catégories de firmes saines et de l'autre, par les catégories de défaillantes, pouvaient être dégagés. Nous avons vu, par exemple, l'importance de la variable Variation Capitaux Propres Hors Résultat comme facteur d'ajustement. La question qui se pose à présent consiste notamment à se demander comment celle-ci est utilisée, s'il y a une ou plusieurs logiques qui président à son emploi, et si cette logique est dépendante d'un profil donné. Nous avons fait de même pour les principales variables des modèles. Enfin, de manière à compléter l'étude des catégories, nous avons souhaité étudier les rapports de proximité qu'elles entretiennent. En effet, lorsque l'on dresse un profil, on fournit une vision moyenne d'un sous-ensemble d'entreprises qui est souvent trompeuse en raison de la très grande variabilité des ratios, notamment lorsqu'ils décrivent des firmes en mauvaise santé, laquelle ne permet pas de comprendre comment ce profil se situe par rapport aux autres. On ne sait donc pas de quel groupe celui-ci est-il le plus proche ou de quel autre est-il plutôt éloigné, et où se situe-t-il par rapport aux frontières entre des saines et des défaillantes. Nous avons donc analysé la topologie des données de chaque ensemble de variables afin de comprendre comment s'organisent les groupes autour des frontières délimitant des territoires de défaillance et regarder s'il existe des points communs entre tous ces territoires qui permettraient de compléter ce que les profils définis précédemment ont mis en évidence.

6.1 CLASSEMENT DES ENTREPRISES

Pour classer les entreprises, nous avons réalisé une analyse en composantes principales avec chaque ensemble de variables – selon la procédure employée par Laitinen (1991) – et les données des échantillons d'apprentissage et de test réunis. Nous avons ensuite retenu les facteurs expliquant la plus grande partie de la variance, dès lors que les dimensions sous-jacentes à chacun

d'eux permettaient d'isoler des profils d'entreprises assez différents. Lorsque deux facteurs présentaient des corrélations similaires, celui expliquant le moins de variance a été éliminé immédiatement, ainsi que les suivants. Nous avons donc privilégié la capacité d'un facteur à délimiter un profil différent des autres sur plusieurs variables, même si, dans certains cas, la part de variance portée par ce facteur pouvait être faible. Les corrélations des variables aux axes retenus sont indiquées ci-après. On constate que les deux premiers facteurs incarnent des profils bien distincts alors que le troisième n'a pas toujours cette même qualité de représentation. Nous l'avons quand même conservé dans tous les cas de figure, car il permet néanmoins de décrire un profil à part entière.

Tableau 45 : Corrélations des variables aux axes factoriels

Critère d'évaluation : **Lambda de Wilks**
Recherche : **Stepwise**

Variables		Axes		
		1	2	3
(Dispo. + VMP) / Total Actif	LI	0,93	0,21	0,11
Trésorerie / Total Actif	LI	0,93	0,20	0,12
Trésorerie / Total Dette	LI	0,85	0,11	0,44
EBE / Total Actif	RE	0,23	0,91	0,18
Résultat Exploi. / Total Actif	RE	0,17	0,90	0,30
Capitaux Propres / Total Actif	SF	0,06	0,43	0,77
Total Dette / Cap. Propres	SF	0,37	0,16	0,86
Variance expliquée		59,18%	19,90%	11,01%
Variance totale				90,09%

Critère d'évaluation : **Vraisemblance**
Recherche : **Backward Stepwise**

Variables		Axes		
		1	2	3
Variation Cap. Pr. H. Résultat	AP	-0,13	-0,71	0,25
(Dispo. + VMP) / Total Actif	LI	0,90	0,22	0,00
(Disponibilités + VMP) / CA	LI	0,94	0,21	-0,01
Rés. C.. Av. I. / Cap. Propres	RE	0,00	0,10	0,94
EBE / Total Actif	RE	0,24	0,78	0,22
Trésorerie / CA	RO	0,93	0,18	0,03
Capitaux Propres / Total Actif	SF	0,18	0,69	0,21
Variance expliquée		47,09%	17,57%	12,98%
Variance totale				77,64%

Critère d'évaluation : **Vraisemblance**
Recherche : **Forward Stepwise**

Variables		Axes		
		1	2	3
Variation Cap. Pr. H. Résultat	AP	-0,19	0,91	-0,17
Résultat Exploitation / CA	EF	0,85	-0,35	0,12
Trésorerie / Total Dette	LI	0,20	-0,22	0,92
EBE / Total Actif	RE	0,87	-0,26	0,13
Capitaux Propres / Total Actif	SF	0,71	0,19	0,43
Variance expliquée		55,08%	16,41%	13,83%
Variance totale				85,32%

Critère d'évaluation : Erreur				
Recherche : Backward				
Variables		Axes		
		1	2	3
Variation Cap. Pr. H. Résultat	AP	-0,45	-0,17	0,39
Résultat Exploitation / CA	EF	0,90	0,17	-0,11
(Dispo. + VMP) / Total Actif	LI	0,21	0,92	-0,13
Tréso. / Dette Court Terme	LI	0,25	0,92	-0,10
EBE / Total Actif	RE	0,88	0,19	-0,10
Résultat Exploi. / Total Actif	RE	0,94	0,17	-0,01
Clients / CA	RO	0,05	-0,11	0,91
Capitaux Propres / Total Actif	SF	0,64	0,28	0,22
Variance expliquée		49,03%	16,70%	11,46%
Variance totale				77,18%

Critère d'évaluation : Mesure Ordre 0				
Recherche : Backward				
Variables		Axes		
		1	2	3
Variation Cap. Pr. H. Résultat	AP	0,01	-0,17	-0,84
(Dispo. + VMP) / Total Actif	LI	0,14	0,91	0,17
Tréso. / Dette Court Terme	LI	0,20	0,89	0,18
EBE / Total Actif	RE	0,54	0,21	0,64
Résultat Net / Total Actif	RE	0,66	0,19	0,55
Trésorerie / CA	RO	0,15	0,90	0,13
Capitaux Propres / Total Actif	SF	0,95	0,17	0,07
Total Dette / Total Actif	SF	-0,94	-0,17	-0,06
Variance expliquée		52,34%	20,16%	11,66%
Variance totale				84,16%

Critère d'évaluation : **Mesure Ordre 1**
Recherche : **Backward**

Variables		Axes		
		1	2	3
Variation Cap. Pr. H. Résultat	AP	-0,09	-0,33	-0,78
ETE / CA	EF	0,24	-0,08	0,84
Actif Circ. / Dette Crt Terme	LI	0,20	0,91	0,03
EBE / Total Actif	RE	0,92	0,19	0,25
Résultat Exploi. / Total Actif	RE	0,91	0,28	0,13
Total Dette / Cap. Propres	SF	0,25	0,86	0,18
Variance expliquée		50,84%	19,04%	14,61%
Variance totale				84,49%

Une fois les analyses factorielles réalisées, nous avons créé des groupes à partir des scores factoriels calculés pour toutes les entreprises, en affectant une firme au facteur présentant la corrélation la plus importante, à l'image de Laitinen (1991). Ainsi, pour chaque ensemble de variables, avons-nous retenu 3 groupes. Nous avons ensuite calculé, par groupe, en différenciant les entreprises saines des entreprises défaillantes, les moyennes de chaque variable. Elles permettent de dresser des profils d'entreprises à partir des dimensions sous-jacentes aux données. Les résultats ont été obtenus sur la base de variables centrées et réduites. Les résultats associés au modèle sélectionné grâce au critère *Lambda de Wilks* sont donnés dans le tableau 46 présenté ci-après.

Tableau 46 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales
Critère : Lambda de Wilks – Procédure de recherche : Stepwise

Entreprises saines				Entreprises défailtantes					
Variables		S. 1	S. 2	S. 3	Variables	Déf. 1	Déf. 2	Déf. 3	
(Dispo. + VMP) / Total Actif	LI	1,63	-0,10	0,01	(Dispo. + VMP) / Total Actif	LI	-0,06	-0,61	-0,73
Trésorerie / Total Actif	LI	1,44	0,02	0,15	Trésorerie / Total Actif	LI	0,07	-0,69	-0,80
Trésorerie / Total Dette	LI	1,20	-0,11	0,43	Trésorerie / Total Dette	LI	-0,25	-0,54	-0,58
EBE / Total Actif	RE	0,55	0,65	0,41	EBE / Total Actif	RE	-1,03	0,10	-1,01
Résultat Exploi. / Total Actif	RE	0,57	0,60	0,47	Résultat Exploi. / Total Actif	RE	-1,20	0,09	-0,89
Capitaux Propres / Total Actif	SF	0,56	0,36	0,76	Capitaux Propres / Total Actif	SF	-1,43	-0,33	-0,12
Total Dette / Cap. Propres	SF	0,49	0,04	1,24	Total Dette / Cap. Propres	SF	-0,82	-0,50	-0,35
Nombre d'entreprises		83	95	72	Nombre d'entreprises		64	106	80

Dans le tableau 46 ci-dessus, les entreprises saines se répartissent en 3 groupes, opposant des entreprises particulièrement liquides (rapport très important entre la trésorerie et les fonds propres) disposant d'une bonne structure financière (S. 1), à des entreprises (S. 3) ayant une excellente structure financière (bonne capitalisation, bon rapport de la dette sur fonds propres, notamment) mais beaucoup moins liquides que le premier groupe. Entre les deux, on trouve un groupe d'entreprises (S. 2) présentant un niveau de rentabilité sensiblement identique aux deux autres groupes, bien que légèrement supérieur, possédant un meilleur rapport des dettes aux fonds propres, mais un faible niveau de liquidité, se rapprochant de ce point de vue de certaines entreprises appartenant au groupe des défailtantes.

Les entreprises défailtantes opposent de la même manière deux groupes, mais de manière moins marquée. Ainsi le premier groupe (Déf. 1) se caractérise-t-il par des entreprises ayant une très mauvaise rentabilité d'exploitation, la plus mauvaise structure financière de toutes les entreprises de l'échantillon, mais une liquidité moyenne, presque similaire à celle du groupe 2 des entreprises saines. Ce groupe s'oppose plus particulièrement au deuxième (Déf. 2) qui présente de mauvais indicateurs de liquidité et de structure financière, mais une rentabilité moyenne et même très légèrement supérieure à la moyenne. Quant au troisième groupe (Déf. 3), il est constitué d'entreprises en difficulté sur toutes les dimensions présentes dans les données.

Le critère employé ici a donc mis en relief deux éléments importants : tout d'abord, la rentabilité d'exploitation, telle que mesurée par EBE / Total Actif et Résultat Exploi. / Total Actif est une propriété commune à toutes les entreprises saines de l'échantillon qui permet de pallier,

lorsqu'ils existent, certains problèmes d'endettement ou certaines insuffisances de liquidité. Ensuite, de la même manière que la rentabilité est un critère partagé par les entreprises en bonne santé, la faiblesse importante de la structure financière et plus précisément, le manque de capitalisation et le poids de la dette sont des caractéristiques communes des entreprises défaillantes. On retrouve bien ici, les caractéristiques des petites entreprises où la rentabilité seule permet souvent de faire face aux engagements à court et long terme, mais aussi où le manque de fonds propres ne peut pas toujours être pallié par une bonne rentabilité, le premier conditionnant fréquemment la seconde.

Si l'on examine maintenant les résultats obtenus avec les variables sélectionnées grâce au critère de *vraisemblance*, on obtient une image qui diffère de la précédente sur quelques points. Ces résultats sont donnés dans les tableaux ci-après.

Tableau 47 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales
Critère : Vraisemblance – Procédure de recherche : Backward Stepwise

Entreprises saines				Entreprises défaillantes					
Variables		S. 1	S. 2	S. 3	Variables	Déf. 1	Déf. 2	Déf. 3	
Variation Cap. Pr. H. Résultat	AP	-0,16	-0,62	0,28	Variation Cap. Pr. H. Résultat	AP	0,29	-0,14	0,39
(Dispo. + VMP) / Total Actif	LI	1,42	-0,06	-0,11	(Dispo. + VMP) / Total Actif	LI	0,07	-0,72	-0,67
(Disponibilités + VMP) / CA	LI	1,45	-0,10	-0,25	(Disponibilités + VMP) / CA	LI	0,07	-0,65	-0,63
Rés. C.. Av. I. / Cap. Propres	RE	0,10	0,12	0,24	Rés. C.. Av. I. / Cap. Propres	RE	-0,41	-0,40	0,04
EBE / Total Actif	RE	0,55	0,68	0,21	EBE / Total Actif	RE	-1,12	0,06	-0,49
Trésorerie / CA	RO	1,32	0,02	-0,05	Trésorerie / CA	RO	0,18	-1,05	-0,64
Capitaux Propres / Total Actif	SF	0,59	0,57	0,36	Capitaux Propres / Total Actif	SF	-1,45	-0,13	-0,31
Nombre d'entreprises		97	110	43	Nombre d'entreprises		57	39	154

Critère : Vraisemblance – Procédure de recherche : Forward Stepwise

Entreprises saines				Entreprises défaillantes					
Variables		S. 1	S. 2	S. 3	Variables	Déf. 1	Déf. 2	Déf. 3	
Variation Cap. Pr. H. Résultat	AP	-0,27	0,55	-0,53	Variation Cap. Pr. H. Résultat	AP	0,11	0,37	-0,02
Résultat Exploitation / CA	EF	0,63	0,07	0,58	Résultat Exploitation / CA	EF	0,28	-0,61	-1,52
Trésorerie / Total Dette	LI	-0,02	-0,08	1,61	Trésorerie / Total Dette	LI	-0,51	-0,52	-0,19
EBE / Total Actif	RE	0,63	0,02	0,60	EBE / Total Actif	RE	0,31	-0,59	-1,64
Capitaux Propres / Total Actif	SF	0,46	0,47	0,71	Capitaux Propres / Total Actif	SF	-0,36	-0,28	-1,70
Nombre d'entreprises		141	33	76	Nombre d'entreprises		62	146	42

Dans le premier tableau, on constate, qu'ici aussi, les entreprises saines sont caractérisées par un bon niveau de rentabilité, mais la magnitude de ce niveau est bien moins importante que précédemment, surtout pour les entreprises du groupe 3 (S. 3) qui, pour compenser leur faible rentabilité relative et surtout leur manque de liquidité, doivent s'appuyer sur une bonne structure de capitalisation et avoir recours de temps à autre à des apports en capitaux, comme le montre la variable Variation Capitaux Propres Hors Résultat. On retrouve donc globalement les mêmes profils que ceux présentés dans le tableau 46, mais avec quelques variantes, dues à la façon dont les fonds propres évoluent d'un exercice à l'autre. En ce qui concerne les entreprises défailtantes, les profils varient selon la capacité des entreprises à jouer sur des leviers susceptibles de limiter les problèmes de rentabilité et de structure financière. En effet, certaines entreprises (Déf. 1) tentent de compenser leurs faiblesses en jouant à la fois sur leur trésorerie et leur structure capitalistique. Mais, bien qu'elles parviennent à améliorer leur liquidité, celle-ci est loin d'être suffisante pour redresser correctement la rentabilité et la structure financière. D'autres, qui ne réussissent pas cet exercice lié à la gestion de trésorerie, ont tendance à faire bien plus appel que les précédentes à des apports en fonds propres (Déf. 3), leur permettant ainsi de limiter la dégradation de leur rentabilité, mais pas suffisamment là aussi. Entre les deux, on trouve un ensemble d'entreprises (Déf. 2) qui n'a ni la capacité de recourir à des apports en fonds propres, ni celle de jouer sur la trésorerie pour pouvoir essayer de se maintenir à flot.

Le deuxième tableau, lui, dit sensiblement la même chose que le précédent. Toutes les entreprises saines présentent un bon niveau de rentabilité et une bonne structure financière, et celles qui ont des problèmes de liquidité ont plutôt tendance à recourir à des augmentations de fonds propres pour compenser leurs faiblesses. Quant aux défailtantes, on constate qu'un groupe (Déf. 1) tente de subsister en faisant tourner son cycle d'exploitation plus que les autres, qui constitue une façon d'alimenter la trésorerie, tout en ayant recours à des augmentations de fonds propres. Un deuxième groupe (Déf. 2), qui ne possède pas les mêmes ressorts dans l'exercice de son activité que le précédent, est obligé de faire bien plus appel à ses actionnaires pour tenter de compenser fortement cette faiblesse. Et le troisième (Déf. 3), lui, n'a ni les ressources internes, ni l'appui de ses actionnaires pour essayer de se maintenir en activité.

On voit donc ici comment la prise en compte de cette variation des capitaux propres permet d'éclairer les profils discernés précédemment et illustrés par le tableau 46, p. 185 : elle permet à certaines entreprises saines de maintenir un bon niveau de rentabilité et explique comment

des défailtantes arrivent à éviter l'effondrement de cette rentabilité. Là encore, il s'agit d'une caractéristique des petites entreprises qui, ne pouvant pas faire appel à l'emprunt obligataire, et étant limitées dans leur capacité à recourir aux emprunts bancaires, peuvent être amenées à financer leur cycle d'exploitation par l'intermédiaire de capitaux propres.

Enfin, lorsque l'on analyse les résultats obtenus avec les variables sélectionnées grâce au réseau de neurones, on obtient des images assez voisines, dans les grandes lignes, des précédentes. Ces résultats sont donnés dans les tableaux 48 et 49 ci-après, le premier tableau ci-dessous correspondant à ceux obtenus à l'aide du critère d'erreur.

Tableau 48 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales
Critère : Erreur – Procédure de recherche : Backward

Entreprises saines				Entreprises défailtantes					
Variables		S. 1	S. 2	S. 3	Variables		Déf. 1	Déf. 2	Déf. 3
Variation Cap. Pr. H. Résultat	AP	-0,38	-0,34	0,18	Variation Cap. Pr. H. Résultat	AP	0,11	0,14	0,46
Résultat Exploitation / CA	EF	0,59	0,59	0,34	Résultat Exploitation / CA	EF	0,06	-1,13	-0,66
(Dispo. + VMP) / Total Actif	LI	-0,15	1,57	0,00	(Dispo. + VMP) / Total Actif	LI	-0,64	-0,08	-0,62
Tréso. / Dette Court Terme	LI	-0,06	1,45	0,00	Tréso. / Dette Court Terme	LI	-0,62	-0,28	-0,53
EBE / Total Actif	RE	0,61	0,59	0,22	EBE / Total Actif	RE	0,08	-1,24	-0,61
Résultat Exploi. / Total Actif	RE	0,60	0,59	0,28	Résultat Exploi. / Total Actif	RE	0,09	-1,45	-0,53
Clients / CA	RO	-0,36	-0,44	1,22	Clients / CA	RO	-0,53	-0,62	0,93
Capitaux Propres / Total Actif	SF	0,51	0,63	0,44	Capitaux Propres / Total Actif	SF	-0,38	-1,33	-0,28
Nombre d'entreprises		122	92	36	Nombre d'entreprises		73	55	122

Comme précédemment, les entreprises saines présentent toutes une bonne rentabilité et une structure de capitalisation forte, alors que les défailtantes sont dans une situation totalement opposée qui se double d'une mauvaise liquidité. Parmi les entreprises saines, on trouve trois groupes. Le premier (S. 1) est composé d'entreprises disposant d'une bonne structure de capitalisation, leur permettant de ne pas avoir recours à des apports dans ce domaine, présentant une excellente rentabilité et réalisant une bonne gestion de leurs créances clients, mais n'offrant pas un bon niveau de liquidité. Le deuxième (S. 2) est le plus solide des trois car caractérisé par de bons indicateurs sur toutes les dimensions. Et le troisième groupe (S. 3) est le plus problématique : en effet, il représente des entreprises qui ont une rentabilité assez moyenne et un niveau de liquidité faible et qui ne réussissent à conserver une bonne structure de fonds propres qu'en raison d'apports non négligeables dans ce domaine. Par ailleurs, on

voit que les difficultés de trésorerie des entreprises de ce groupe sont vraisemblablement aplanies grâce à une forte rotation des comptes clients, car leur ratio Clients / CA est particulièrement élevé comparativement aux deux autres groupes.

Les entreprises défaillantes, elles, présentent des profils aussi contrastés que les saines. Le premier sous ensemble (Déf. 1) possède une structure capitalistique déficiente et un niveau de liquidité dans l'ensemble assez mauvais, par manque de trésorerie, qui explique les difficultés de paiement de la dette à court terme, mais finalement une rentabilité d'exploitation moyenne, en tout cas bien meilleure que celle des deux autres groupes. Le deuxième sous ensemble (Déf. 2) est plus fragile que le premier si l'on se réfère à sa structure de capitalisation, et dispose d'un meilleur niveau de trésorerie, compte tenu de sa structure d'actifs et de ses dettes à court terme, mais d'une très mauvaise rentabilité d'exploitation : c'est d'ailleurs ce groupe qui possède le ratio BFR / Total Actif le plus faible. Le troisième sous-ensemble (Déf. 3) combine les difficultés des deux premiers, et seul un apport important en fonds propres l'empêche de présenter une structure capitalistique trop mauvaise.

Les grandes différences à l'intérieur de chaque groupe s'incarnent donc essentiellement autour de la structure de liquidité, à laquelle s'ajoute, mais uniquement pour les défaillantes, la structure de rentabilité. Les autres variables viennent, soit confirmer la primauté de ces facteurs, soit indiquer comment les entreprises tentent de compenser les difficultés qui en résultent, *via* le recours aux actionnaires et/ou *via* des mesures visant à agir sur l'exploitation.

Si l'on compare les caractéristiques générales des saines et des défaillantes précédentes à celles obtenues avec les critères d'*ordre 0* et d'*ordre 1*, décrits dans le tableau 49 indiqué plus loin, on obtient des résultats voisins. Encore une fois, toutes les entreprises saines sont toutes rentables, au sens où les ratios moyens mesurés ici, et décrivant la rentabilité d'exploitation, sont dans l'ensemble bien meilleurs que ceux des défaillantes, avec une structure financière plus solide que celle des autres, mais jouent différemment avec le cycle d'exploitation et les augmentations de fonds propres, selon qu'il s'agit de pallier des difficultés ou de préparer l'avenir. Quant aux défaillantes, elles ont toutes des sérieux problèmes liés à la rentabilité et à la trésorerie et effectuent, pour la majeure partie d'entre elles, des recours massifs aux augmentations de capitaux pour reculer l'échéance fatidique. Maintenant, si l'on retrouve bien ces points communs, la structure de chaque groupe diffère selon le critère, avec d'un côté, une représentation propre au critère d'*erreur* et de l'autre, une représentation partagée par les critères d'*ordre 0* et d'*ordre 1*.

Ainsi, le premier groupe d'entreprises saines (S. 1) du tableau 49, et pour les deux critères, est composé de firmes qui présentent le taux de liquidité le plus faible, qui sont plutôt en retrait des deux autres quant à la rentabilité, et qui doivent faire appel à leurs actionnaires pour conserver une bonne capitalisation compte tenu de leurs actifs (Capitaux Propres / Total Actif), qui n'en demeure pas moins légèrement plus faible que celle des deux autres. Ce dernier ratio ne faisant pas partie des variables sélectionnées avec le critère d'*ordre 1*, nous l'avons alors calculé pour ce groupe : le ratio moyen de cet ensemble est de 0,42, contre 0,77 pour le deuxième et 0,54 pour le troisième.

Tableau 49 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales

Critère : Mesure Ordre 0 – Procédure de recherche : Backward

Entreprises saines				Entreprises défaillantes					
Variables		S. 1	S. 2	S. 3	Variables	Déf. 1	Déf. 2	Déf. 3	
Variation Cap. Pr. H. Résultat	AP	0,15	-0,17	-1,54	Variation Cap. Pr. H. Résultat	AP	0,42	0,33	0,00
(Dispo. + VMP) / Total Actif	LI	-0,28	1,53	0,36	(Dispo. + VMP) / Total Actif	LI	-0,69	-0,15	-0,58
Tréso. / Dette Court Terme	LI	-0,18	1,38	0,40	Tréso. / Dette Court Terme	LI	-0,60	-0,28	-0,58
EBE / Total Actif	RE	0,32	0,58	1,02	EBE / Total Actif	RE	-0,25	-1,40	-0,10
Résultat Net / Total Actif	RE	0,39	0,57	0,87	Résultat Net / Total Actif	RE	-0,17	-1,40	-0,23
Trésorerie / CA	RO	-0,07	1,32	0,31	Trésorerie / CA	RO	-0,72	-0,02	-0,71
Capitaux Propres / Total Actif	SF	0,51	0,59	0,53	Capitaux Propres / Total Actif	SF	0,07	-0,89	-1,23
Total Dette / Total Actif	SF	-0,50	-0,59	-0,52	Total Dette / Total Actif	SF	-0,08	0,85	1,26
Nombre d'entreprises		111	92	47	Nombre d'entreprises		113	73	64

Critère : Mesure Ordre 1 – Procédure de recherche : Backward

Entreprises saines				Entreprises défaillantes					
Variables		S. 1	S. 2	S. 3	Variables	Déf. 1	Déf. 2	Déf. 3	
Variation Cap. Pr. H. Résultat	AP	0,03	-0,27	-1,28	Variation Cap. Pr. H. Résultat	AP	0,45	0,37	0,03
ETE / CA	EF	0,14	-0,03	0,97	ETE / CA	EF	-0,29	-1,21	0,53
Actif Circ. / Dette Crt Terme	LI	0,01	1,43	0,21	Actif Circ. / Dette Crt Terme	LI	-0,48	0,09	-0,83
EBE / Total Actif	RE	0,58	0,36	0,76	EBE / Total Actif	RE	0,07	-1,03	-0,92
Résultat Exploi. / Total Actif	RE	0,55	0,44	0,74	Résultat Exploi. / Total Actif	RE	0,13	-0,86	-1,12
Total Dette / Cap. Propres	SF	0,14	1,31	0,45	Total Dette / Cap. Propres	SF	-0,45	-0,36	-0,76
Nombre d'entreprises		133	73	44	Nombre d'entreprises		101	62	87

Le deuxième groupe (S. 2), lui, est le pendant du groupe S. 2 du tableau 48. Dans ce dernier tableau, S. 2 apparaissait comme le plus solide : meilleure liquidité, excellente rentabilité,

meilleur rapport des fonds propres sur l'actif, et faible recours à des augmentations de capital. On constate bien la même chose dans le tableau 49 (dans la partie relative au critère d'*ordre 1*), qui fait aussi ressortir que ce groupe à le plus fort ratio moyen Total Dette / Cap. Propres. Nous avons recalculé ce ratio pour le groupe S. 2 propre au critère d'*erreur* et au critère d'*ordre 0*, et l'on constate que le S. 2 ici aussi présente un ratio très élevé comparativement aux autres groupes (avec le critère d'*erreur*, il est de 0,85 pour S. 2, contre 0,40 pour S. 1 et 0,18 pour S. 3 ; et avec le critère d'*ordre 0*, il est de 0,69 pour S. 2, contre 0,34 pour S. 1 et 0,71 pour S. 3). Il s'agit vraisemblablement d'entreprises qui ont recours à la dette plutôt pour financer leur développement que pour pallier des insuffisances.

Le troisième et dernier groupe d'entreprises saines (S. 3) ne se retourne quasiment pas vers ses actionnaires et présente des indicateurs de rentabilité très élevés, une bonne liquidité ainsi qu'une bonne structure financière.

En ce qui concerne les défailtantes, le groupe Déf. 1 recourt de manière importante à des augmentations de fonds propres pour pallier son manque de liquidité et surtout, son incapacité à financer ses dettes de court terme. Le groupe Déf. 2, lui, développe une stratégie qui s'appuie à la fois sur une augmentation des capitaux propres et sur l'accélération du cycle d'exploitation pour supporter sa très mauvaise rentabilité. C'est d'ailleurs le groupe de défailtantes qui présente les meilleurs indicateurs de rotation. Enfin, Déf. 3, est composé des entreprises les plus endettées et les moins bien capitalisées, ce qui explique qu'elles ne réussissent pas à obtenir de l'argent frais de leurs actionnaires.

Tous ces résultats ont permis de préciser les conclusions que l'on pouvait déjà tirer du seul examen des ratios issus des sélections et de dresser une liste de constantes caractérisant plusieurs profils de firmes saines et défailtantes. Mais, les profils isolés par tel groupe de variables présentent néanmoins des variantes par rapport à tel l'autre, car les combinaisons de valeurs associées à des ratios qui reviennent simultanément ne sont pas stables. On voit bien les différences entre les profils des entreprises déterminées par le critère d'*erreur* et ceux qui proviennent du critère d'*ordre 0*. Ceci est d'ailleurs dû au fait que les partitions opérées sur la base des analyses factorielles divergent sensiblement : seules un peu plus de 10 % des entreprises se répartissent de manière identique dans les 6 partitions.

On n'obtient en définitive qu'une vision très générale de différents processus qui vont s'incarner ensuite avec des variantes selon les groupes, voire même selon les entreprises. On peut alors

se demander ce que les visions proposées par le réseau de neurones apportent par rapport à celles données à l'aide du critère de *vraisemblance*. Fondamentalement, les dimensions prises en compte sont les mêmes et les différents profils conduisent à des conclusions identiques sur ce qui fait la spécificité des saines et des défaillantes. De plus, toutes ces méthodes ont mis en évidence le rôle de la rentabilité, du niveau d'endettement, de l'efficacité du cycle d'exploitation, de la structure de capitalisation et de son évolution comme facteur essentiel de survie des entreprises. Il faut plutôt voir cet apport comme une confirmation du fait qu'il s'agit là de dimensions d'importance et que les variables sélectionnées sont des variables « fortes », donc plus attachées à la défaillance elle-même qu'aux aléas des techniques de sélection, et que les processus mis en exergue sont bien les processus fondamentaux expliquant notamment la façon dont les entreprises procèdent pour éviter ou tenter d'éviter la défaillance. Il faut aussi voir cet apport comme la démonstration confirmée que certains choix expérimentaux opérés jusqu'à présent, où les variables sélectionnées pour la construction de modèles à base de réseaux de neurones l'étaient à partir d'un critère de variance comme un *Lambda de Wilks*, représentaient des choix pas forcément judicieux.

Peut-on aller plus loin dans l'analyse ? Si l'on observe, non plus la structure moyenne des profils, mais leur contenu et leur articulation, ceci peut-il nous apprendre quelque chose de supplémentaire sur la défaillance ? De ce point de vue, les ensembles sélectionnés grâce à un réseau présentent-ils des spécificités ou, comme précédemment, ne permettent-ils que de renforcer des constats effectués par ailleurs ? Par exemple, y a-t-il des ensembles de variables permettant de créer des profils plus homogènes que d'autres, donc bien distincts, au sein d'un même ensemble de saines ou défaillantes ? Ou alors, peut-on mettre en évidence un lien entre la façon dont des profils se répartissent sur un territoire de la défaillance et leur hiérarchie vis-à-vis du risque de défaillance ? C'est à ces questions que nous allons maintenant tenter de répondre.

6.2 TOPOLOGIE DES DONNÉES

Pour analyser la topologie des différentes partitions calculées à la suite des procédures de sélection de variables, nous avons construit des cartes⁴¹ représentant des territoires qui peuvent être interprétées comme des territoires de défaillance et qui permettent de visualiser

⁴¹ Les cartes qui vont être présentées ont été calculées à partir des données décrivant les 500 entreprises de l'échantillon ayant servi aux apprentissages et aux tests. Elles ne doivent donc pas être interprétées dans l'absolu, mais les unes par rapport aux autres, puisqu'elles sont tributaires de la structure de cet échantillon.

sur un plan, les proximités entre les entreprises de nos échantillons ainsi qu’entre les profils auxquelles elles appartiennent. Pour ce faire, nous avons employé un type de réseau de neurones particulier : les « cartes de *Kohonen* ». Les cartes de *Kohonen* constituent une classe spécifique de réseaux permettant de projeter l’espace des variables initiales sur un espace de plus faible dimension : deux, la plupart du temps. Elles possèdent deux propriétés intéressantes : d’abord, elles permettent de résumer « au mieux » les données initiales, en limitant les pertes d’information ; ensuite, elles permettent de conserver la topologie des données et donc les rapports de proximité initiaux. De telles cartes présentent un avantage notable par rapport aux cartes factorielles, issues par exemple d’une analyse en composantes principales : la notion d’axe et de variance captée par les axes n’existe pas et l’on n’est pas contraint d’utiliser plusieurs projections pour obtenir une représentation fidèle de ce que contiennent les données, par exemple en croisant plusieurs axes deux à deux. Une seule et unique carte permet de résumer au mieux les données, sans qu’il soit nécessaire d’en utiliser plusieurs. Le fonctionnement des cartes de *Kohonen* est décrit en annexe 1, p. 213.

Nous avons donc créé une carte par ensemble de variables à partir de toutes les entreprises des échantillons d’apprentissage et de test. L’élaboration d’une carte se faisant au travers d’un processus de quantification vectorielle, dont l’algorithme a été largement décrit notamment par Kohonen (1989), chaque neurone peut être vu comme un prototype ou un individu type d’un certain nombre d’entreprises qui sont similaires au sens de la métrique employée : ici, une distance euclidienne. Deux neurones proches sur la carte décrivent des entreprises proches dans l’espace initial des données. Nous avons complété les cartes en associant une classe à chaque neurone, par l’intermédiaire d’un vote majoritaire. À la fin du processus de quantification, la classe attribuée au neurone correspond à la classe d’appartenance du plus grand nombre d’entreprises dont il se trouve être le plus proche. En cas d’égalité entre deux classes, celle des deux qui est affectée le plus souvent aux neurones contigus est alors attribuée au neurone en question.

Pour déterminer la forme de cartes, donc le nombre de lignes et de colonnes, nous avons d’abord procédé, pour chaque ensemble de variables, à une étude de leur topologie à l’aide d’une projection de *Sammon* (*Sammon Mapping*), dont nous décrivons l’algorithme en annexe 2, p. 217. Cette méthode cherche à trouver une représentation de chaque observation dans un espace à 2 dimensions où les distances entre paires d’observations mesurées dans l’espace initial correspondent au mieux aux distances entre les mêmes paires d’observations, mais mesurées dans

l'espace de sortie. Il s'agit d'une méthode d'analyse multidimensionnelle comme MDS (*Multi-dimensional Scaling*). Ainsi, en observant la structure projetée, on obtient une idée générale de la forme des données, qui renseigne utilement sur celle qu'il est judicieux d'attribuer à la carte, donc si l'on a plutôt intérêt à disposer d'une carte de forme rectangulaire ou carrée. Nous avons opté pour cette dernière architecture au vu des différentes représentations, car rien ne plaidait particulièrement pour une forme différente. Une fois ceci défini, nous avons choisi un nombre de neurones. Ici, la fixation de ce nombre ne peut qu'être assez arbitraire, même si elle dépend des objectifs que l'on se fixe. Nous avons souhaité une topologie plutôt restreinte et qui permette en même temps de détailler chaque profil. Nous avons donc choisi des cartes composées de 100 neurones⁴², avec 10 neurones en ligne et 10 en colonne, soit un nombre de neurones représentant $1/5^{\text{ème}}$ du nombre d'entreprises dont les données seront employées.

Afin de garantir le respect des relations d'ordre de l'espace d'entrée, nous avons contrôlé à la fin de l'apprentissage de chaque carte qu'aucun repli ne s'était produit. En effet, une carte décrit des rapports de proximité dans l'espace des variables. Donc, il faut pouvoir s'assurer que les neurones obtenus respectent ces relations et que deux neurones contigus sont bien deux neurones proches l'un de l'autre au sens de la distance employée. Pour cela, nous avons effectué une analyse des cartes à l'aide de la méthode de *Sammon* : une projection de *Sammon* possède la particularité de restituer les distances les plus faibles, donc les structures locales, plus fidèlement que toutes les autres méthodes de sa famille, bien que ceci se fasse au détriment des distances les plus grandes. C'est précisément parce que cette propriété de conservation des structures locales est aussi une propriété essentielle des cartes de *Kohonen* que nous avons employé cette méthode.

Avec la méthode de *Sammon*, nous avons projeté sur un plan l'ensemble des neurones d'une carte, puis nous avons observé si la projection obtenue restituait correctement le positionnement des neurones sur la carte. Afin de matérialiser les proximités sur la carte, nous avons fait en sorte de relier tous les neurones contigus par un trait. Ainsi, on doit donc obtenir, *in fine*, une grille qui ressemble à celle formée par les neurones de la carte. Si l'on observe des replis sur cette grille, ceci signifie que les dimensions sous-jacentes n'ont pas été correctement appréhendées et qu'elles ne se satisfont pas d'une réduction dans un espace à deux dimensions seulement ; donc que la projection ne respecte pas les distances de l'espace des variables.

⁴² Pour calculer les cartes, nous avons utilisé une distance euclidienne, la première fonction de fonction de voisinage décrite en annexe 1, p. 215, parmi les 4 citées, et la quatrième fonction de décroissance du pas d'apprentissage, parmi les 4 cités. Le pas d'apprentissage initial a été fixé à 0,05.

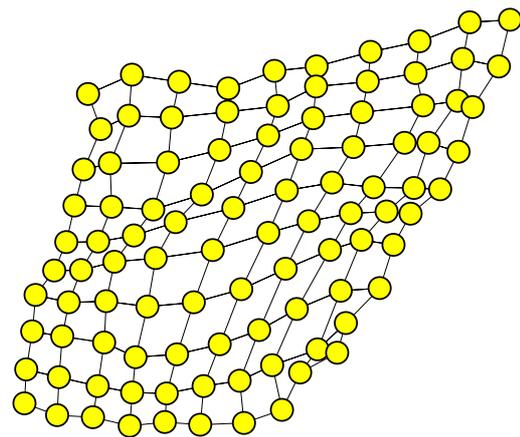
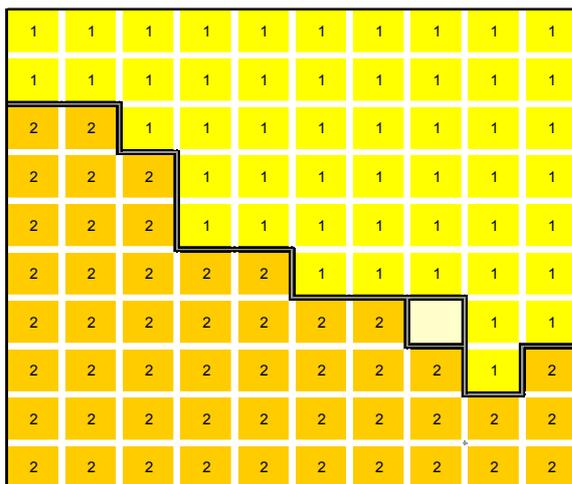
Nous avons tout d'abord étudié la répartition des neurones par groupe d'entreprises, selon qu'ils représentent des saines ou des défaillantes. Chaque carte a été réalisée à l'aide de l'un des 6 ensembles de variables définis précédemment et nous avons joint, à chaque fois, la projection de *Sammon* qui lui est associée ; sur la figure 5, on trouve la carte à gauche, et sa projection de *Sammon* à droite. Pour disposer de points de comparaison, nous avons ajouté en fin de liste, une dernière carte réalisée à l'aide de huit variables tirées au hasard, de façon à ce que les vecteurs aient une dimension similaire à ceux des autres ensembles de variables. Les neurones colorés en jaune et numérotés avec un 1 représentent les entreprises saines, ceux colorés en orange, avec un 2, représentent les défaillantes, et ceux colorés en blanc ne correspondent à aucune des deux classes.

Figure 5 : Cartes des classes et projection de Sammon

Critère : Lambda de Wilks – Procédure de recherche : Backward

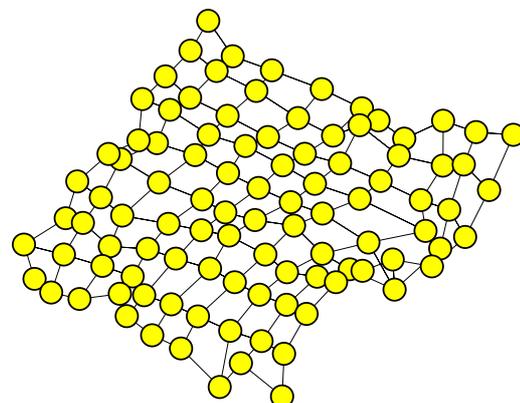
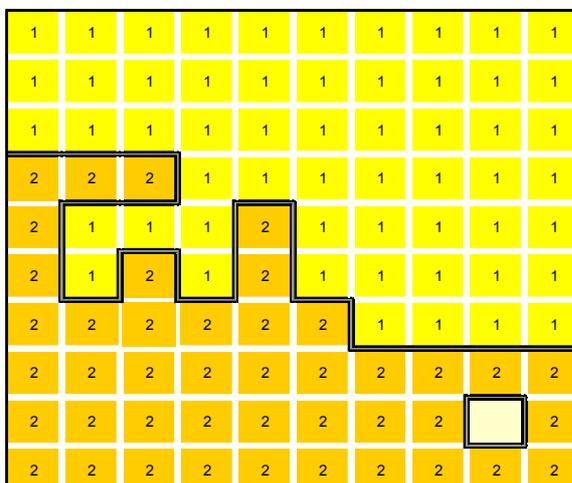
Carte de Kohonen

Projection de Sammon



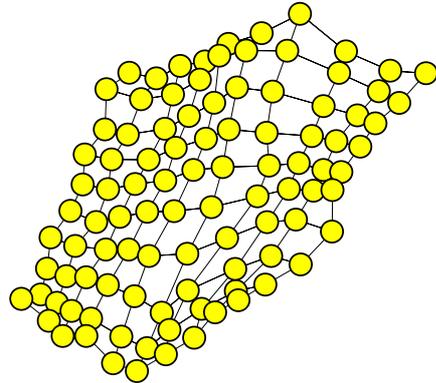
1 Entreprises saines – 2 Entreprises défaillantes

Critère : Vraisemblance – Procédure de recherche : Backward



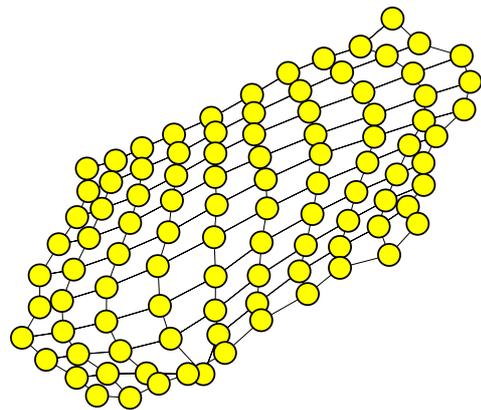
Critère : Vraisemblance – Procédure de recherche : Forward

1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
2	2	1	1	1	1	1	1	1	1
2	2	1	1	1	1		1	1	1
2	2	2	1	1	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1
2	2	2	2	2	2	2	1	1	1
2	2	2	2		2		2	2	1
2	2	2	2	2	2	2	2	1	1
2	2	2	2	2	2	2	2	2	2



Critère : Erreur – Procédure de recherche : Backward

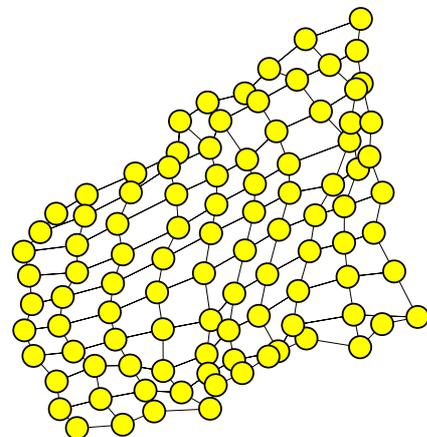
2	1	1	1	1	1	1	1	1	1
2	1	2	1	1	1	1	1	1	1
2		2	2	1	1		1	1	1
2	2	2	1	1	1	1	1	1	1
2	2	2	1	1	1	1	1	1	1
2	2	2	2	2	1	1	1	1	1
2	2	2	2	2	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1
2	2	2	2		2	1	1	1	
z	2	2	2	2	2	1	1	1	1

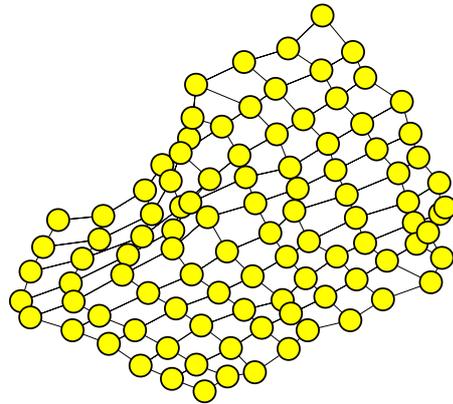
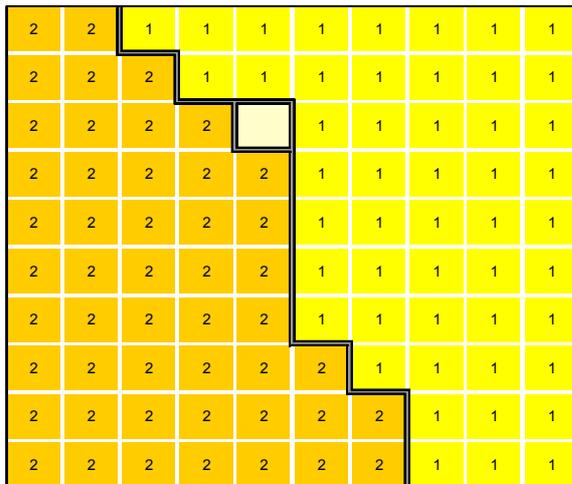
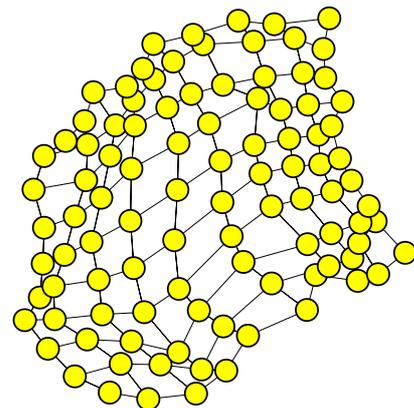
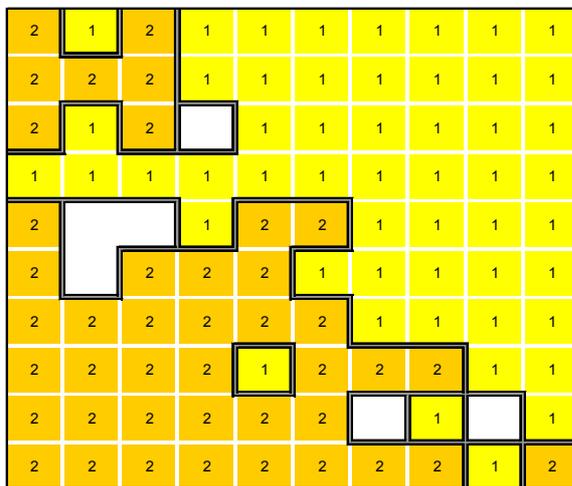


1 Entreprises saines – 2 Entreprises défailiantes

Critère : Mesure Ordre 0 – Procédure de recherche : Backward

2	2	1	1	1	1	1	1	1	1
2	2	2	1	1	1	1	1	1	1
2	2	2	1	1	1	1	1	1	
2	2	2	1	1	1	1	1	1	1
2	2	2	2	1	1	1	1	1	1
2	2	2	2	2	1	1	1	1	1
2	2	2	2	2	1	1	1		1
2	2	2	2	2	2	2	1	1	1
2			2	2	2	2	2	1	1
2	2	2	2	2	2	2	1	1	1



Critère : Mesure Ordre 1 – Procédure de recherche : Backward**Variables tirées au hasard**

1 Entreprises saines – 2 Entreprises défailiantes

La quantification réalisée à l'aide de variables choisies aléatoirement montre que la frontière entre les classes est mal délimitée, puisque certains neurones de la classe 1 se retrouvent entourés de neurones de la classe 2, et qu'il en est de même pour un groupe de neurones de la classe 2 (en haut à gauche de la carte ci-dessus), complètement entourés de neurones de l'autre classe.

En revanche, les quantifications opérées avec les 6 ensembles de variables sélectionnées pour notre étude montrent que les deux groupes d'entreprises occupent chacun un espace homogène, donc que les profils identifiés par les variables ont été bien distingués par les cartes. La forme de la frontière est plus ou moins chaotique selon l'ensemble considéré, montrant la difficulté d'obtenir une ligne de séparation claire en les deux groupes

On retrouve bien, dans d'autres études, une telle répartition des entreprises dans l'espace. Ainsi, Back. et al (1994), Serrano-Cinca (1996) et Kiviluoto (1998) ont-ils effectués des représentations

similaires à l'aide d'une carte de *Kohonen* indiquant que la délimitation entre les classes est assez erratique, mais qu'avec des ensembles de variables bien choisies pour leur pouvoir de discrimination, on obtient deux territoires assez distincts.

Les caractéristiques des cartes sont données dans le tableau 50, ci-après. Ce dernier indique, pour chacune d'elles, donc chaque ensemble de variables sélectionnées sur la base d'un critère d'évaluation, le nombre de neurones par groupe, le nombre de variables (pour mémoire), l'erreur de quantification de chaque groupe, c'est-à-dire la perte d'information due à la réduction de l'espace d'entrée sur l'espace de la carte, l'écart type de l'erreur et le seuil de signification d'un test de *t* de différence de moyenne de l'erreur de quantification. L'erreur correspond à la moyenne des distances euclidiennes séparant tous les neurones d'une même classe des entreprises qui leur sont les plus proches.

Tableau 50 : Caractéristiques des groupes d'entreprises selon leur représentation sur les cartes

Critères	Nombre de neurones par groupe		Nombre de variables	Erreur de quantification		Ecart type de l'erreur		p-val.
	S.	Déf.		S.	Déf.	S.	Déf.	
Lambda de Wilks	50	49	7	0,37	0,43	0,20	0,26	0,002
Vraisemblance – Back.	56	43	7	0,47	0,52	0,17	0,23	0,004
Vraisemblance – For.	51	46	5	0,40	0,43	0,18	0,22	0,164
Erreur	56	40	8	0,77	0,79	0,26	0,30	0,554
Ordre 0	51	45	8	0,51	0,56	0,21	0,27	0,019
Ordre 1	50	49	6	0,48	0,52	0,20	0,23	0,056

S. : entreprises saines – Déf. : entreprises défailtantes

p-val. : seuil de signification – p-valeur – d'un test *t* de *Student* de différence de moyenne de l'erreur de quantification entre les deux groupes

Ce tableau fait ressortir plusieurs éléments. Tout d'abord, on distingue que chaque groupe est représenté dans l'ensemble par un nombre de neurones assez différent, puisqu'en moyenne les saines sont codées par 52 neurones, contre 45 pour les défailtantes. Le plus petit écart, qui apparaît à deux reprises, est dû aux variables sélectionnées par un *Lambda de Wilks* et par le critère d'*ordre 1*, où les saines sont représentées par 50 neurones, et les défailtantes par 49, et le plus grand est dû aux variables sélectionnées par le critère d'*erreur*, où les saines sont associées à 56 neurones, et les défailtantes, à 40. Ceci conforte l'idée que nous évoquons en analysant les résultats des classifications que l'on retrouve dans la littérature. En effet, on observe le plus souvent que les entreprises saines sont mieux classées que les défailtantes, comme si les saines présentaient un spectre de profils plus large que les autres, dont une

partie est plutôt semblable à des profils de défaillantes. On voit ici que les saines sont réparties sur un plus grand nombre de neurones, sans que l'espace entre les neurones ne présentent d'irrégularités, comme le montre les projections de *Sammon* de chaque carte. Incidemment, Pérez (2002) note, en réalisant des typologies d'entreprises à l'aide de cartes de *Kohonen* et de ratios financiers, que les firmes saines présenteraient des profils plus variés que les autres, sans aller plus loin dans l'analyse. Ensuite, on voit que l'erreur de quantification des défaillantes est plus élevée que celle des saines, et la différence est significative dans 3 cas sur 6, au seuil de 5 %, et presque significative dans un 4^{ème} cas, avec le critère d'ordre 1. L'erreur est due en partie à la différence du nombre de neurones propres à chaque classe, lorsque cette différence existe. Dès lors qu'un groupe est représenté par un nombre de prototypes plus faible qu'un autre, il est légitime de considérer que la qualité de sa quantification sera moindre, ce qui est vraisemblablement le cas ici. Mais, un autre facteur joue quand le nombre de neurones référents de chaque groupe est voisin : si l'on analyse de plus près l'erreur, on constate que les différences entre les saines et les défaillantes sont dues à certains profils de défaillantes, comme indiqué sur le tableau qui suit.

Tableau 51 : Caractéristiques des profils d'entreprises selon leur représentation sur les cartes

Critères	Nombre de neurones par groupe		Nombre de variables	Erreur de quantification par profil					
	S.	Déf.		S. 1	S. 2	S. 3	Déf. 1	Déf. 2	Déf. 3
Lambda de Wilks	50	49	7	0,32	0,39	0,39	0,41	0,47	0,39
Vraisemblance – Back.	56	43	7	0,43	0,51	0,48	0,61	0,55	0,49
Vraisemblance – For.	51	46	5	0,42	0,41	0,38	0,53	0,38	0,44
Erreur	56	40	8	0,81	0,75	0,71	0,85	0,86	0,72
Ordre 0	51	45	8	0,51	0,48	0,57	0,50	0,55	0,68
Ordre 1	50	49	6	0,46	0,51	0,48	0,52	0,52	0,52

S. : Entreprises saines – Déf. : Entreprises défaillantes

Seuls quelques profils (7 sur les 21, surlignés ci-dessus en jaune) possèdent une erreur beaucoup plus importante que les autres, suggérant que l'on est certainement en présence de groupes assez hétérogènes. Maintenant, si l'on observe la structure des cartes à la lumière de la performance des modèles, on constate, pour les critères propres au réseau, que la performance est d'autant plus élevée que le nombre de neurones codant les saines est important, comme indiqué ci-après. Et il en est de même pour les 2 modèles élaborés avec un critère de *vraisemblance*. Ceci étant, la hiérarchie n'est pas respectée si l'on ne différencie plus les familles de critères.

Tableau 52 : Structure des cartes et performances des modèles

Critères	Nombre de neurones par groupe		Ecart	Erreur de prévision (échantillons de test)
	S.	Déf.		
Erreur	56	40	16	94,03%
Ordre 0	51	45	6	93,59%
Ordre 1	50	49	1	92,82%
Vraisemblance – Back.	56	43	13	92,01%
Vraisemblance – For.	51	46	5	89,20%
Lambda de Wilks	50	49	1	87,20%

Si l'on analyse ces mêmes performances, mais cette fois en détaillant les taux de prévision par classe et par modèle, on voit dans le tableau 53, ci-après, qu'il y a un lien entre la capacité de prévision des entreprises défaillantes et la structure des cartes. En effet, les modèles qui obtiennent des taux de classement des défaillantes bien supérieurs aux taux des saines sont plutôt bâtis sur des profils où les entreprises saines peuvent exprimer toute leur diversité, ce qui se traduit sur les cartes par une différence importante dans le nombre de neurones codant chacun des 2 groupes d'entreprises. Les trois premières lignes du tableau ci-dessous résument bien la situation, où 2 modèles sur les 3, effectuent de meilleures prévisions des défaillantes que des saines. En bas du tableau, on trouve les modèles qui donnent des taux de classements des défaillantes comparables voire inférieurs à ceux des saines, et pour lesquels les profils sous-jacents ont été codés par les cartes avec un nombre voisin de neurones.

Tableau 53 : Structure des cartes et performances des modèles par groupe d'entreprises

Critères	Nombre de neurones par groupe		Ecart	Taux de bons classements par groupe	
	S.	Déf.		S.	Déf.
Erreur	56	40	16	92,78%	95,29%
Vraisemblance – Back.	56	43	13	93,60%	90,42%
Ordre 0	51	45	6	91,96%	95,21%
Vraisemblance – For.	51	46	5	89,56%	88,85%
Ordre 1	50	49	1	92,81%	92,83%
Lambda de Wilks	50	49	1	91,20%	83,19%

Ces résultats doivent être pris avec précaution car, comme nous l'avons souligné, ils sont fondés sur un seul échantillon et ils expriment seulement une tendance, car la relation présentée ci-dessus ne se vérifie pas à chaque fois. Ils demandent donc à être étudiés de près et confirmés, et s'ils l'étaient, ils permettraient de montrer qu'il existe des profils, donc des ratios, qui intrinsèquement

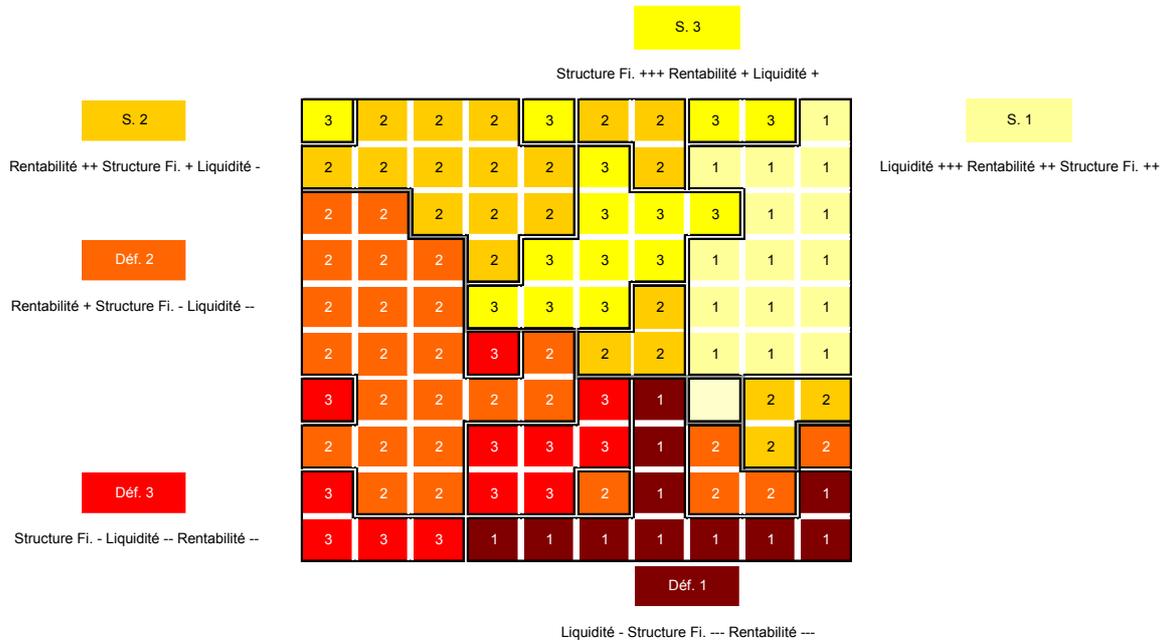
conduisent à discriminer mieux que d'autres les entreprises défaillantes des saines. Les seules techniques dont on dispose aujourd'hui pour maximiser les prévisions correctes des défaillantes consistent à jouer sur le *cut* qui délimite les classes, donc sur un paramètre conjoncturel. L'intérêt de posséder des variables ayant cette aptitude réside dans la simplification des calculs relatifs à la prise en compte, dans la fonction de coût employée pour construire des modèles, de l'asymétrie du coût d'un mauvais classement des saines et des défaillantes. On pourrait alors s'en passer.

On voit donc que les modèles les plus précis sont assis sur une représentation du profil financier des firmes qui aurait plutôt tendance à bien discerner les spécificités des saines. Les profils qui ont été définis ne représentent en somme que des moyennes et donc n'expriment pas toute la dispersion qui peut exister autour de celles-ci. On a vu précédemment, que ces moyennes délimitent bien des structures financières assez différentes et on a vu avec les cartes que les deux classes d'entreprises occupent des positions dans l'espace de la défaillance plutôt bien distinctes. Qu'en est-il alors de la répartition des profils dans cet espace ? Sont-ils répartis chacun sur un territoire distinct de celui des autres ? Quelles relations de proximité entretiennent-ils ? Peut-on distinguer un *continuum* de régions allant des entreprises les « plus saines » aux « moins saines » ? Y a-t-il en dernier ressort, une structure de l'espace de la défaillance ? Pour répondre à ces questions, nous avons élaboré, pour chaque carte décrite précédemment, une répartition des neurones par profil à l'aide d'une procédure similaire à celle qui a permis de répartir les neurones entre les deux groupes de firmes. Afin d'y parvenir, nous avons isolé tout d'abord, par neurone, toutes les entreprises utilisées précédemment pour déterminer leur classe d'appartenance, donc toutes celles qui leur sont le plus proche compte tenu de la distance utilisée ici. Ensuite, nous avons repéré, parmi ces entreprises, le profil dominant, *i.e.* celui auquel elles ont été majoritairement associées à la suite de l'analyse factorielle, calcul réalisé par un simple dénombrement. Enfin, nous avons attribué à chaque neurone, le profil en question. Les cartes indiquées ci-après illustrent ces répartitions pour les 6 ensembles de variables de notre étude.

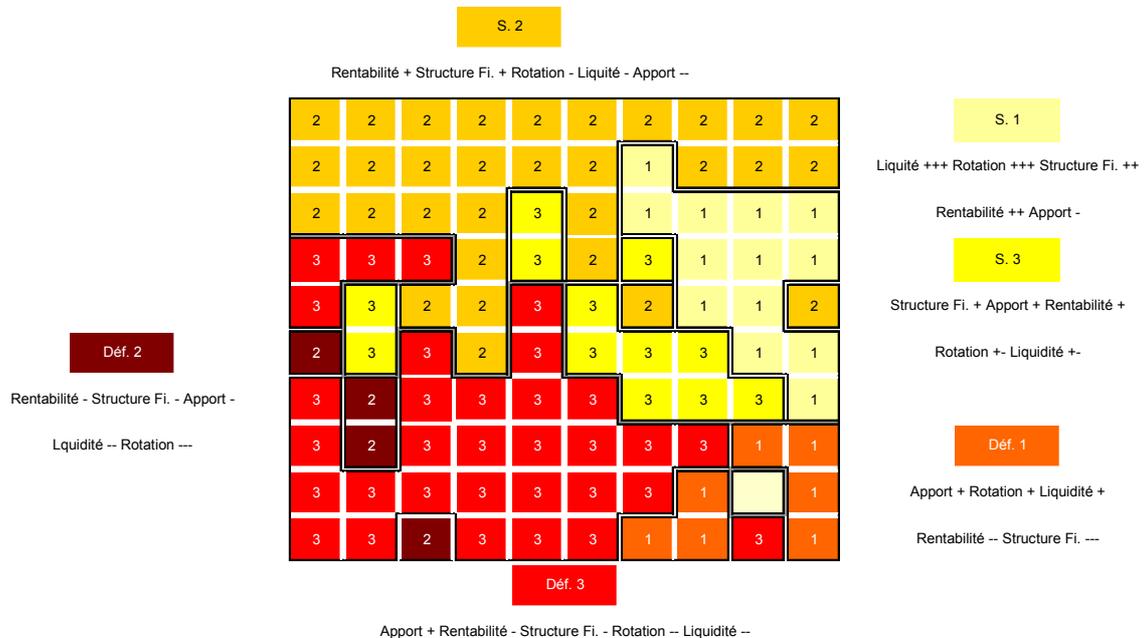
Afin de rendre les cartes les plus lisibles possible, nous avons affecté une couleur en fonction de leur statut vis-à-vis de la défaillance mesuré par les profils, statut apprécié de manière subjective. En effet, il parait délicat de considérer que tel profil est « meilleur » du point de vue de la santé financière que tel autre. Mais, pour, pour la clarté de la représentation, il était nécessaire de tenter une telle hiérarchisation. Selon la hiérarchie employée ici, plus la couleur est claire, plus le profil est considéré comme sain. Nous avons indiqué autour des cartes, les

dimensions sous jacentes aux profils en leur attribuant une note combinant des + et des -. Ces notes résument les valeurs moyennes de chaque variable indiquées au sein des tableaux de profils (tableaux 46, 47, 48 et 49)⁴³.

Figure 6 : Cartes des profils
Critère : Lambda de Wilks – Procédure de recherche : Backward

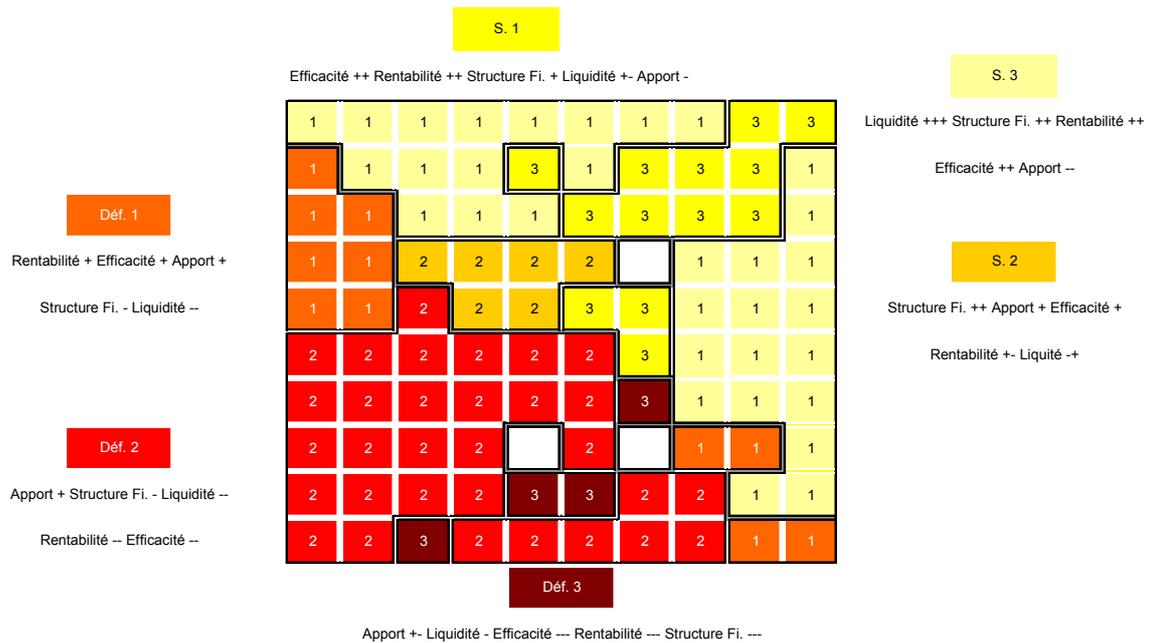


Critère : Vraisemblance – Procédure de recherche : Backward



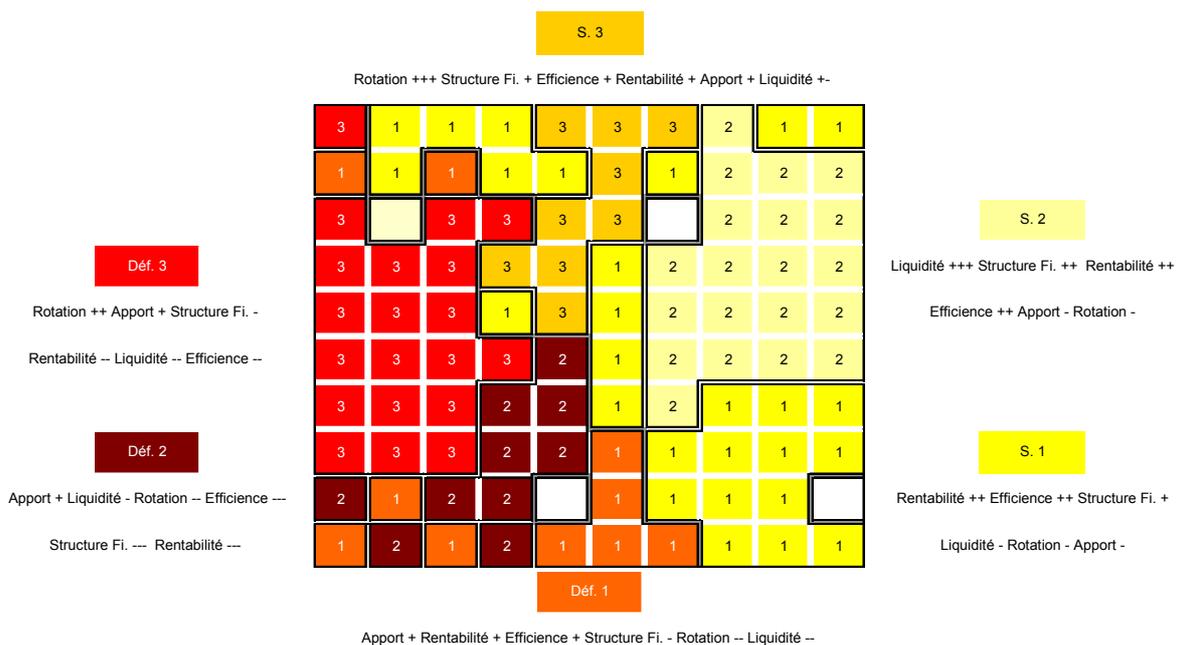
⁴³ Toute dimension recueillant des valeurs supérieures à 1 se voit attribuée +++ ; entre 0,5 et 1, ++ ; entre 0 et 0,5, + ; entre 0 et -0,5, - ; entre -0,5 et -1, --, et au-delà, ---. Les couleurs ont été attribuées selon le nombre de + recueillis par chaque dimension et selon l'importance de la dimension eu égard à la santé financière des firmes appréciée au travers de notre subjectivité.

Critère : Vraisemblance – Procédure de recherche : Forward

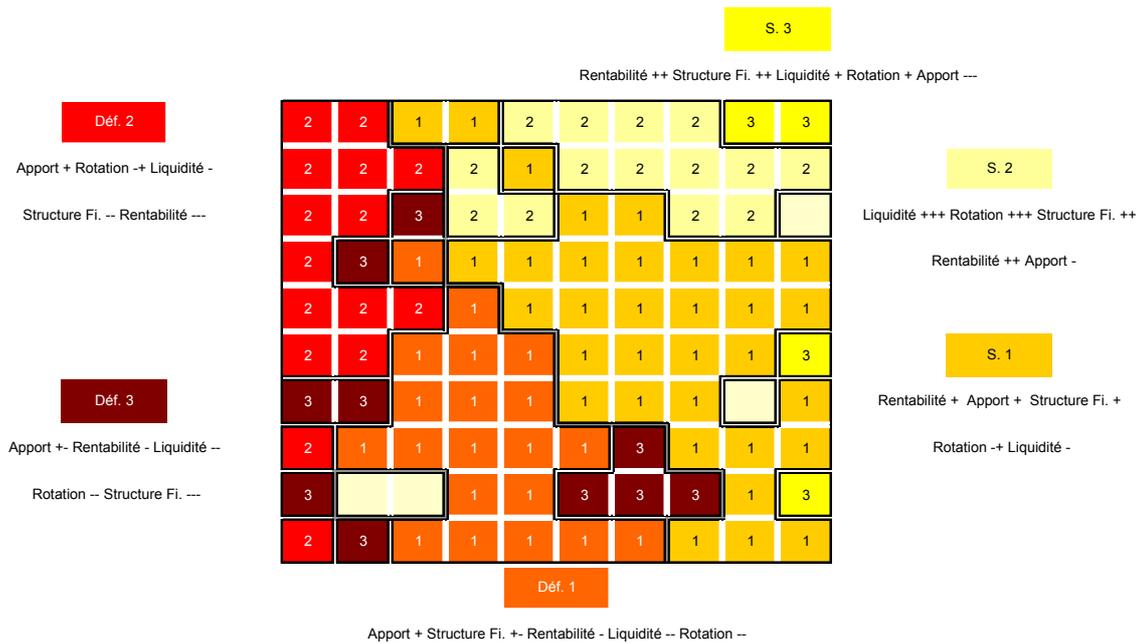


L'orientation des cartes a été « forcée » de manière à ce que l'on retrouve, à chaque fois, les entreprises saines plutôt dans la moitié supérieure droite de chaque schéma et les défailtantes, dans la moitié opposée. Ceci ne change rien aux rapports de proximité et facilite la lecture comparée des représentations. La frontière entre les saines et défailtantes est matérialisée par un double trait avec une couleur grise entre les deux et les frontières entre les profils, par un double trait avec une couleur blanche au milieu.

Critère : Erreur – Procédure de recherche : Backward



Critère : Mesure Ordre 0 – Procédure de recherche : Backward



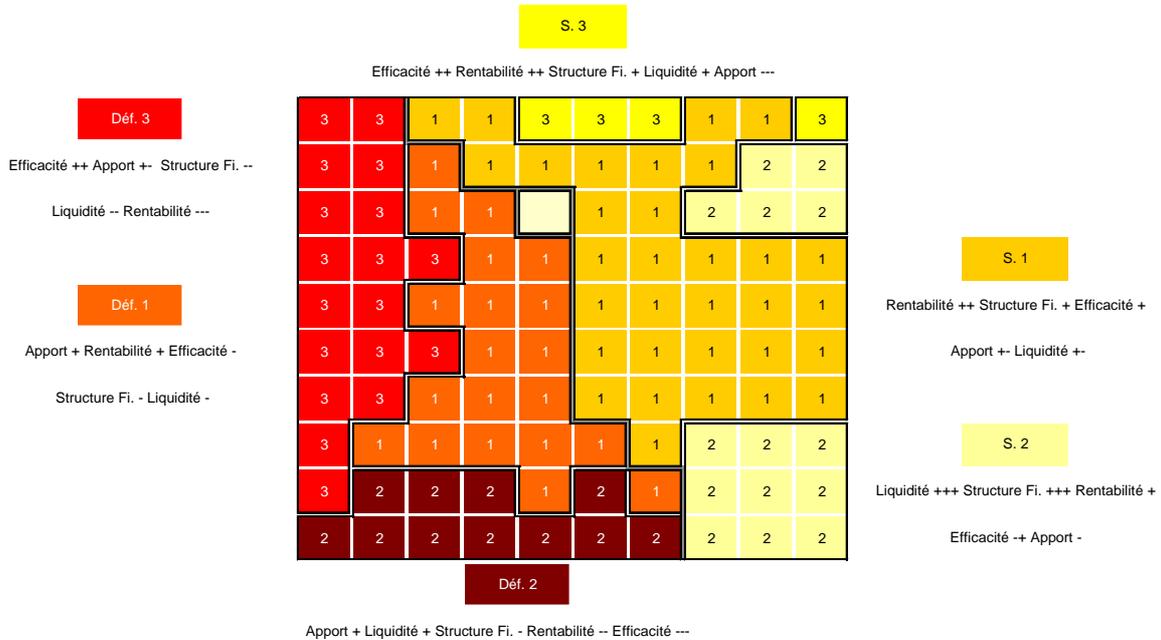
Les cartes montrent tout d'abord que les deux ou trois profils dominants, donc ceux qui sont formés par le plus grand nombre d'entreprises, ont tendance à occuper, chacun, des positions contiguës de l'espace. Sur la carte située ci-dessus, par exemple, on voit clairement que le profil 1 des firmes saines (S. 1) s'étale sur un très grand nombre de neurones contigus, comme le profil 1 des défailtantes (Déf. 1). C'est aussi le cas, mais dans une moindre mesure, des profils 2 de chaque groupe d'entreprises (S. 2 et Déf. 2). En revanche, les autres profils, donc ceux qui s'appuient sur les plus petits ensembles d'observations, ont tendance à venir s'intercaler au sein des autres, et sont donc souvent très fragmentés. L'apparente déstructuration des cartes provient donc de ces « petits » profils qui ne sont pas suffisamment homogènes pour occuper des positions proches. Seule la carte ci-après offre une vision très compacte des données, même si le profil 2 des firmes saines est découpé en deux parties relativement distantes l'une de l'autre.

Tous les groupes très éclatés sur les cartes se retrouvent bien dans le tableau 51, p. 199, comme étant ceux qui présentent les erreurs de quantification les plus élevées et sensiblement différentes de celles des autres. Et lorsque l'erreur due à chaque profil est à peu près similaire, on obtient la topologie décrite sur la carte indiquée page suivante.

Ensuite, on distingue sur toutes les cartes qu'il n'y a pas de distance véritablement marquée entre des prototypes d'entreprises que l'on peut qualifier de financièrement très « sains » et d'autres que l'on peut qualifier de très « dégradés ». En effet, quasiment tous les profils se côtoient au voisinage de la frontière entre saines et défailtantes, certes au travers de peu de

neurones, mais ils partagent néanmoins des points de friction. Il y a donc, en marge de tous les profils, des sous-ensembles qui partagent des caractéristiques avec plusieurs autres, y compris certains de la classe opposée, montrant que la distance à la défaillance s’incarne dans de multiples formes.

Critère : Mesure Ordre 1 – Procédure de recherche : Backward



Dans l’ensemble, l’analyse factorielle a permis de construire des sous-ensembles qui donnent une vision assez compacte des profils mais qui cache néanmoins une certaine disparité au sein des observations participant à la formation de ces sous groupes. Les variables sélectionnées grâce au réseau, de ce point de vue, n’apportent rien de particulier, et la structure plutôt régulière de la carte due au critère d’*ordre 1* relève certainement plus du hasard que d’une quelconque propriété du réseau ou de la méthode de sélection. On ne discerne pas non plus de lien entre une quelconque forme d’organisation des territoires et la performance des modèles sur lesquels reposent les cartes.

Si l’on considère que le chemin de la défaillance s’interprète comme le déplacement d’une entreprise à partir d’une position donnée dans l’espace d’une carte vers une autre qui la rapprocherait de la frontière, on s’aperçoit que de multiples voies peuvent être empruntées, la position de départ ne préjugant pas de celle d’arrivée, puisque l’on ne distingue aucun *contignum* dans l’organisation des états de santé financière. La variabilité constatée est vraisemblablement une propriété forte des données qui confirme bien qu’une entreprise peut glisser vers la défaillance de bien des manières, même si ces dernières peuvent se résumer dans seulement

quelques variables. Ce qui fait toute la diversité des chemins réside dans la multiplicité des combinaisons d'occurrences des ratios qui ne se laisse pas facilement appréhender.

6.3 CONCLUSION DU CHAPITRE

Ce dernier chapitre se donnait pour objectif d'étudier la contribution des modèles créés pour notre étude à la compréhension du phénomène de défaillance, relativement au type d'entreprises des échantillons employés bien entendu. Il s'agissait avant tout, à l'aide de variables isolées par différentes techniques, de tenter de mettre en lumière l'existence de constantes au sein des entreprises vues au travers de profils décrivant des états de santé financière marquant une plus ou moins grande distance à la défaillance.

Nous avons d'abord analysé la représentation générale de chaque groupe de firmes donnée par les six ensembles de variables choisis précédemment. Ceci nous a permis de montrer que les points communs à toutes les entreprises saines et défaillantes, révélés par tous les modèles, sont essentiellement structurés autour du cycle d'exploitation, et sont vraiment révélateurs des mécanismes de gestion fondamentaux des PME. Nous avons aussi montré comment les modèles décrivent plutôt bien l'articulation entre l'exploitation et la capacité de l'entreprise à la mener à bien, le besoin en fond de roulement, la rentabilité, la dette ainsi que la structure capitalistique comme variable d'ajustement en cas de difficulté qui, là aussi, illustre les rouages de base de la logique financière des petites entreprises.

Nous avons ensuite étudié les différents profils définis grâce à une série d'analyses factorielles, à l'aide de cartes de *Kohonen*. Elles ont permis de révéler trois éléments essentiels. D'abord, les entreprises saines présentent une plus grande diversité de profils financiers que les défaillantes, déjà constatée par Pérez (2002), et nos modèles semblent montrer que la capacité d'une règle de classification à prévoir correctement l'issue des entreprises défaillantes est liée à cette faculté d'appréhender toute la variété des firmes saines. Ensuite, les modèles créés parviennent à donner des représentations qui ne sont que partiellement homogènes. En effet, si les cartes obtenues partitionnent bien l'espace en deux, avec d'un côté les saines et de l'autre, les défaillantes, la frontière qui sépare ces deux zones permet quand même de faire cohabiter des profils qui sont, en moyenne, plutôt distants. Enfin, une analyse plus détaillée des profils a aussi permis d'illustrer les rapports de proximité qu'ils entretiennent, sans que ceci permette de dégager de structure apparente dans leur organisation. Rien dans les projections

ne permet de détecter quelque logique dans la façon dont les firmes pourraient cheminer vers la défaillance ou indiquer des points de passage obligés, les profils n'étant pas répartis selon une organisation qui révélerait une gradation dans la sévérité de leur état vis-à-vis de la défaillance. Au contraire, des profils assez distincts peuvent partager des frontières, mêmes si celles-ci sont parfois très étroites. C'est d'ailleurs ces proximités qui sont à l'origine des difficultés à séparer correctement les groupes, et quand bien même dispose-t-on d'un modèle précis, ce dernier ne tire pas ses capacités d'une réorganisation générale de l'espace, mais uniquement d'une meilleure approximation des frontières.

CONCLUSION

Toute la littérature financière suggère que les modèles de défaillance sont de nature réellement conjoncturelle et qu'il est nécessaire, à chaque fois que l'on souhaite en créer un nouveau, de rechercher des variables *ad hoc*, précisément adaptées à l'échantillon employé à un moment donné. À cette caractéristique, s'ajoute leur extrême instabilité, car leur durée de vie et leur champ d'application sont somme toute assez limités. En effet, un modèle ne peut pas conserver des performances même moyennes, si on l'applique à des firmes dont les caractéristiques sont par trop éloignées de celles qui ont été employées pour son paramétrage. C'est le cas, lorsqu'elles sont issues d'un autre secteur d'activité ou qu'elles disposent d'une taille d'actifs différente. Il ne peut pas non plus conserver de bonnes performances au fil du temps, y compris lorsque l'on ne le change pas de contexte d'application.

Les raisons de ces phénomènes sont multiples et difficilement maîtrisables, comme nous l'avons vu, mais on dispose néanmoins d'une fenêtre de temps au cours de laquelle un modèle est opérationnel, et au-delà de laquelle il doit être réévalué, dès lors que l'on a pris la précaution de bien délimiter l'espace de recherche de la fonction de classement qui fera office de règle de prévision. C'est dans ce contexte qu'il faut interpréter les résultats que nous avons obtenus, donc leurs limites.

Nous avons montré qu'un modèle créé avec un réseau de neurones tire mieux partie d'une sélection de variables effectuée avec des méthodes adaptées à ses caractéristiques, qu'il ne peut le faire avec d'autres méthodes et notamment, avec celles couramment mises en œuvre. Ces dernières ont l'avantage d'être simples et rapides d'utilisation et la question des temps de calcul peut légitimer la sous-utilisation des premières. Ceci étant, un certain nombre de travaux de recherche ont exploré d'autres voies, notamment à l'aide d'algorithmes génétiques. Donc la raison qui sous-tend cette sous-utilisation doit être recherchée ailleurs et vraisemblablement dans le manque de passerelles entre les disciplines. On le constate d'ailleurs, lorsque l'on observe la variété des algorithmes neuronaux employés pour construire des modèles de défaillance, qui est en définitive assez faible, puisque depuis l'origine un seul type domine, alors qu'il en existe bien d'autres. Il en est de même pour les méthodes de sélection de variables qui constituent à elle seule un champ disciplinaire bien distinct de la gestion d'entreprise, et qui explique que les techniques qui lui sont propres n'aient pas encore débordé en finance.

Ces résultats demandent donc à être confirmés avec d'autres échantillons, et notamment avec des firmes appartenant à d'autres secteurs, ou au sein du même secteur, avec d'autres structures d'actifs, mais plaident en faveur d'un recours plus marqué à d'autres modes de sélection des variables que ce qui se fait traditionnellement. Au demeurant, les derniers travaux publiés récemment s'appuient, comme d'habitude, soit sur des variables choisies parce que déjà présentes dans d'autres études (Jones et Hensher, 2007), soit selon ce que dit l'analyse financière (Zhu et al. 2007).

Nous avons aussi montré qu'il y a un lien entre la capacité de discrimination d'une variable, exprimée au travers d'un simple test t ou F et sa prédisposition à être choisie par un processus de sélection qui s'appuie sur d'autres propriétés que cette dernière, mais que ce lien est trop détendu pour être utilisé seul dans une stratégie de sélection. En conséquence, il faut être extrêmement prudent quant à l'emploi de tels tests lorsque l'on envisage de créer un modèle neuronal.

L'analyse des différents modèles que nous avons créés a fait ressortir qu'il existe un lien intime entre la problématique de gestion du type d'entreprises considérée ici, et les variables retenues. En effet, le cycle d'exploitation notamment, vu au travers du besoin en fond de roulement, est bien au cœur de la gestion financière des PME, et occupe une place centrale dans tous les modèles. Ceci suggère que recourir à une analyse de la logique de gestion d'un ensemble d'entreprises homogènes, selon les critères que nous avons retenus ici, peut constituer une étape préalable à un processus de sélection. Jusqu'à présent, les seules analyses conduites étaient plutôt de nature conceptuelle, s'appuyant sur des modèles théoriques de la firme, comme nous l'avons évoqué au paragraphe 1.1. Ici, il s'agirait plutôt de s'appuyer sur une modélisation d'une ou de plusieurs mécaniques de gestion pour en déduire des variables.

Nous avons aussi étudié les profils des entreprises qui ressortent de l'analyse des variables de nos modèles. Les résultats, ici, sont moins tangibles que précédemment. D'un côté, la précision d'un modèle semble être liée à sa capacité à appréhender la variété des situations des saines et des défaillantes, les premières présentant vraisemblablement plus de diversité dans leurs profils que les secondes, de l'autre, tous les modèles indiquent que les frontières entre profils sont assez ténues, y compris entre profils appartenant à des groupes d'entreprises différents (saines *versus* défaillantes). Ceci montre qu'il est difficile de trouver des structures stables qui renseigneraient par exemple, sur la façon dont une catégorie d'entreprises saines peut

cheminer vers la défaillance. Manifestement, les parcours sont multiples, comme le laisse augurer la grande variété des occurrences possibles de situations financières.

Nous tirerons deux conclusions majeures de nos travaux. D'abord, une plus grande attention devrait être portée par la communauté scientifique intéressée par la défaillance sur les protocoles de choix de variables, la primauté ayant été donnée jusqu'à présent aux méthodes de création de modèles et ce, quel que soit l'objectif poursuivi ; améliorer la compréhension du phénomène, la précision des modèles... Ensuite, le secteur bancaire notamment, qui par construction a besoin de qualifier le risque de non remboursement pour fixer un loyer de l'argent prêté, et qui pour des raisons prudentielles doit maintenant pouvoir qualifier le risque associé à toute créance, comme le stipule la réforme dite « Bâle II », peut trouver dans les techniques décrites ici des compléments méthodologiques à sa problématique de gestion. Tous les établissements de crédit disposent aujourd'hui d'outils statistiques pour créer des fonctions de *scoring*, mais encore très peu emploient des techniques non linéaires, et encore moins mettent en œuvre des mécanismes de sélection de variables comme ceux que nous avons pu décrire. Ces techniques-ci sont aujourd'hui suffisamment mûres et efficaces pour être implémentées largement, non pas en se substituant à ce qui existe, mais plutôt en venant compléter les méthodes actuelles, pour former des systèmes permettant d'apprécier le risque de diverses manières. Il devient ainsi envisageable de former des indicateurs à plusieurs dimensions, c'est-à-dire issus de plusieurs modèles : discriminant, logistique, neuronal..., l'appréciation du risque final pouvant se faire alors sur la base de plusieurs résultats donnés par différentes règles de prévision et de leur plus ou moins grande convergence.

Pour conclure, plusieurs voies de recherche peuvent être envisagées compte tenu de ce que nous avons étudié. La première, comme nous l'avons évoqué brièvement, renvoie à la répliquabilité des résultats obtenus relatifs à la précision des modèles selon la méthode de choix des variables. La deuxième voie concerne l'approfondissement de l'étude des techniques de sélection. Nous n'avons utilisé que quelques critères, avec des méthodes de recherche simples. Il conviendrait donc de tester l'efficacité de bien d'autres critères d'évaluation (mesures d'information, de dépendance...), propres ou non aux réseaux, et d'autres algorithmes de parcours de l'espace des variables. La troisième voie est liée aux variables mises en exergue par tous les modèles de l'étude. Il s'agirait d'étudier le comportement de modèles bâtis à l'aide de variables issues de plusieurs procédures de sélection, donc à partir de ce que nous

avons appelé, par analogie avec la typologie, des « formes fortes », mais cette fois constituées non pas d'observations, mais de variables. Enfin, une dernière voie peut être envisagée qui concerne le lien entre des mécanismes de gestion de base propres à un type d'entreprise et des choix de prédicteurs. Les modèles théoriques de la firme ont été employés à quelques reprises pour légitimer de tels choix. Ils s'appuient aussi sur une vision simplifiée du mode de fonctionnement des entreprises, comme le modèle de Laitinen (1991). Ici, nous proposons une démarche quelque peu similaire, mais cette fois qui s'appuie non pas sur une mécanique commune à toutes les entreprises, mais à une certaine catégorie, donc qui épouse les spécificités de ses modes de gestion plutôt que les caractéristiques très générales des modes partagés par toute entreprise.

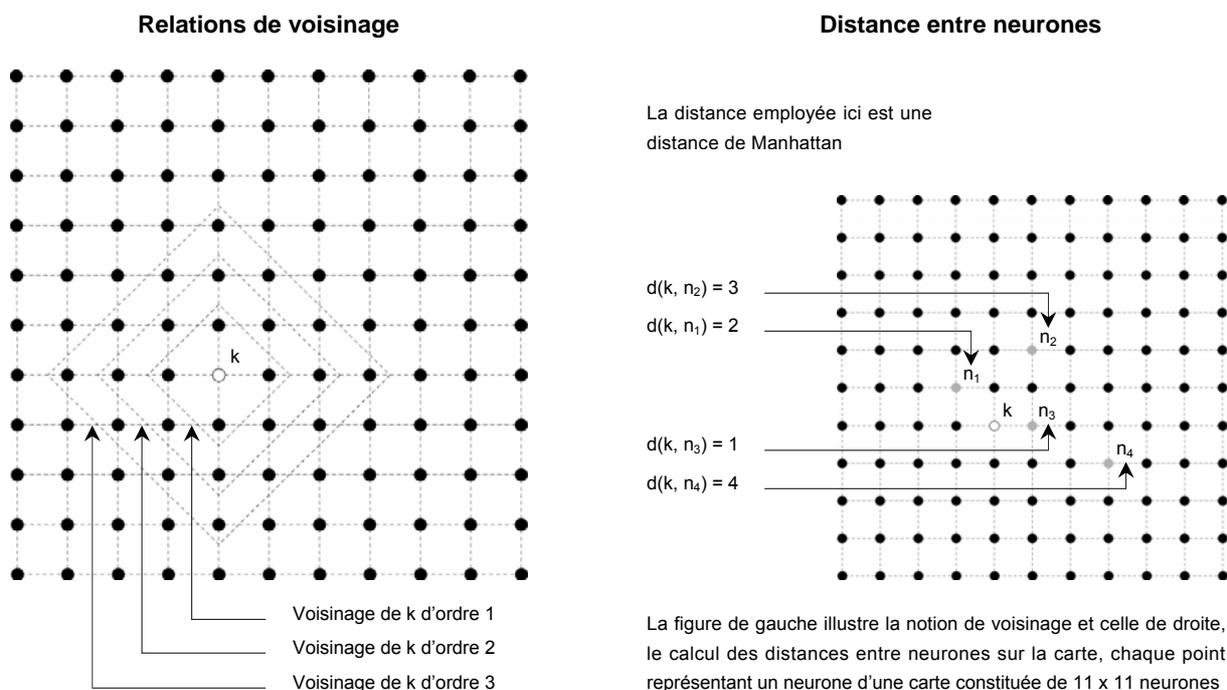
ANNEXES

ANNEXE 1

Principes de fonctionnement des cartes de Kohonen

Les cartes de *Kohonen* sont composées d'un ensemble de neurones répartis sur deux couches. La première, ou couche d'entrée, comporte autant de neurones que de variables à coder et peut être représentée par un vecteur $e = (e_1, \dots, e_n)$, de dimension n , où n correspond au nombre de variables. La seconde, ou carte, est composée de neurones disposés selon une structure le plus souvent matricielle, même si d'autres formes existent. Chaque neurone de la carte est représenté par un vecteur de poids $w = (w_1, \dots, w_n)$, aussi de dimension n . Les neurones de la couche d'entrée sont connectés à tous les neurones de la carte. La topologie de ce réseau est illustrée sur la figure ci-après, tirée de Dreyfus et al. (2004).

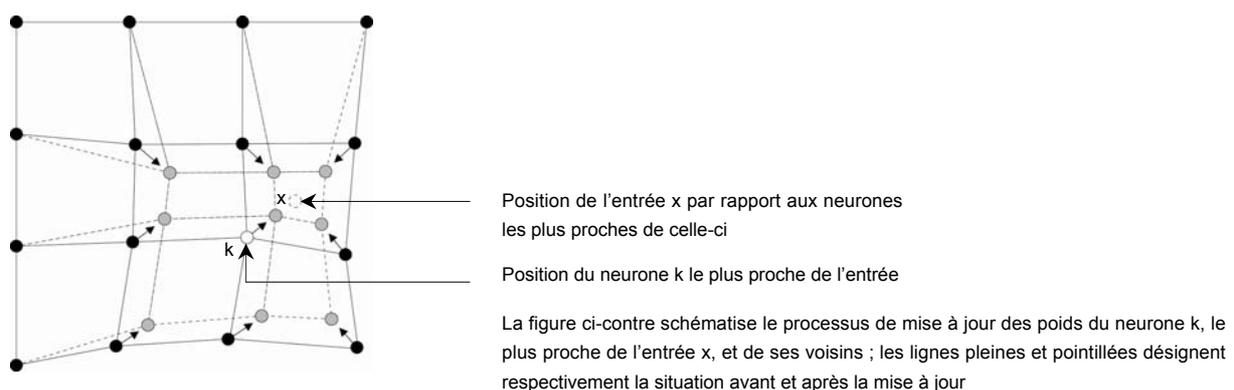
Figure 7 : Topologie d'une carte de Kohonen



Pour qu'une carte apprenne à représenter la topologie d'un ensemble de données, on procède en utilisant une phase d'apprentissage. Au cours de cette phase, les observations de l'échantillon considéré sont comparées à tous les vecteurs de poids par l'intermédiaire d'un calcul de distance. Le neurone de la carte dont le vecteur est le plus proche du vecteur d'entrée voit ses poids modifiés de telle façon que ceux-ci se rapprochent légèrement des valeurs de l'entrée, comme indiqué sur la figure présentée page suivante. Puis, les neurones qui se situent dans le voisinage de ce dernier voient aussi leurs poids modifiés, selon la même logique, mais l'amplitude de la correction est inversement proportionnelle à la distance qui les sépare sur la

carte. Au-delà de cette zone de voisinage, les neurones ne sont affectés par aucune modification. Ce processus se poursuit jusqu'à l'arrêt de l'apprentissage. Au cours de son déroulement, le rayon de voisinage est réduit progressivement pour devenir presque nul à la fin. En fin d'apprentissage, on dispose d'une structure qui résulte de la projection d'un espace de n dimensions sur un espace à 2 dimensions et qui reflète la forme des données, les neurones représentant chacun un prototype de ces dernières. Si l'on souhaite étiqueter les neurones, dès lors que l'on connaît les classes d'appartenance des observations, on peut par exemple affecter aux neurones le numéro de la classe pour laquelle ils constituent un référent. Pour ce faire, on présente une dernière fois l'échantillon au réseau, on repère, pour tous les neurones, les observations dont ils sont les plus proches, au sens de la distance utilisée, et on les étiquette avec le numéro de la classe dominante.

Figure 8 : Processus de mise à jour des poids



L'algorithme d'apprentissage se résume comme indiqué ci-dessous, la couche d'entrée comprenant n neurones, et la carte de *Kohonen* en comportant m :

- fixer la taille m de la carte, avec $m = \text{lignes} * \text{c colonnes}$, puis initialiser aléatoirement les poids ;
- présenter une observation $e = (e_1, \dots, e_n)$ à la couche d'entrée ;
- calculer la distance entre le vecteur $e = (e_1, \dots, e_n)$ et le vecteur de poids $w = (w_{k1}, \dots, w_{kn})$ de chaque neurone w_k et déterminer le neurone w_c pour lequel la distance entre les deux est minimale :

$$\|e - w_c\| = \min \{ \|e - w_k\| \}$$

- modifier les poids, dans le voisinage du neurone déterminé ci-dessus, ainsi :

$$w_k(t+1) = w_k(t) + \alpha(t)h_{ck}(t)(e(t) - w_k(t))$$

avec t représentant le temps, $\alpha(t)$ le pas d'apprentissage, $h_{ck}(t)$ la fonction de voisinage autour du neurone w_c et $e(t)$ l'entrée utilisée. La fonction de voisinage est une fonction

décroissante du temps et de la distance entre un neurone w_k de la carte et le neurone w_c le plus proche de l'entrée à l'instant t . Elle définit la région d'influence qu'une entrée peut avoir sur la carte.

- présenter une nouvelle observation jusqu'à ce que le paramètre t atteigne sa valeur finale.

Différentes fonctions de voisinage ainsi que différentes fonctions de décroissance du pas d'apprentissage sont utilisables (Kohonen et al., 1996, Vesanto et al., 2000). Elles sont décrites ci-après.

Fonctions de voisinage

- $h_{ck}(t) = 1(\sigma(t) - d_{ck})$
- $h_{ck}(t) = e^{-d_{ck}^2 / 2\sigma(t)^2}$
- $h_{ck}(t) = e^{-d_{ck}^2 / 2\sigma(t)^2} 1(\sigma(t) - d_{ck})$
- $h_{ck}(t) = \max\{0, 1 - (\sigma(t) - d_{ck})^2\}$

Paramètres des fonctions

- $\sigma(t)$ correspond au rayon de voisinage à l'instant t , avec $\sigma(t) = \sigma_0(\sigma_n / \sigma_0)^{t/t_{\max}}$, σ_0 et σ_n représentant la valeur initiale et la valeur finale du rayon de voisinage sachant que $\sigma_0 \geq \sigma_n \geq 1$, et t_{\max} la durée de la phase d'apprentissage.
- $d_{ck} = \|w_c - w_k\|$ représente la distance sur la carte entre le neurone w_c et le neurone w_k .
- $1(x)$ représente une fonction seuillée avec $1(x) = 0$ si $x < 0$ et $1(x) = 1$ si $x \geq 0$.

Fonctions de décroissance du pas d'apprentissage

- $\sigma(t) = \alpha_0(\alpha_n / \sigma_0)^{t/t_{\max}}$
- $\sigma(t) = \alpha_0(1 - t/t_{\max})$
- $\sigma(t) = \alpha_0(0,005 / \sigma_0)^{t/t_{\max}}$
- $\sigma(t) = \alpha_0 / (1 + 100t/t_{\max})$

Paramètres des fonctions

- α_0 représente la valeur initiale du pas d'apprentissage.
- α_n représente la valeur finale du pas d'apprentissage avec $\alpha_0 \geq \alpha_n > 1$.
- t_{\max} représente la durée de la phase d'apprentissage généralement exprimée en nombre de cycles de présentation au réseau de l'ensemble des observations.

L'algorithme d'apprentissage qui vient d'être décrit confère à la carte deux propriétés essentielles : résumer « au mieux » les entrées et conserver le « mieux possible » leur topologie. La première dépend du terme $\|e - w_c\|$ qui utilise les entrées pour modéliser les neurones de la carte en les transformant en prototypes de ces dernières. La seconde est liée à la fonction de voisinage h_{ck} qui donne à la carte la capacité de refléter la topologie des données. Ainsi, deux vecteurs proches dans l'espace d'entrée seront-ils aussi proches dans l'espace de sortie.

ANNEXE 2

Algorithme de Sammon

L'algorithme de Sammon cherche à trouver une représentation x_i de chaque objet i dans un espace à 2 dimensions où les distances $d_{ij} = d(x_i, x_j)$ entre paires d'objets mesurées dans l'espace initial correspondent au mieux aux distances $D_{ij} = D(x_i, x_j)$ entre les mêmes paires d'objets, mais mesurées dans l'espace de sortie, avec $i \neq j$. Les distances sont calculées par l'intermédiaire de distances euclidiennes qui s'expriment ainsi :

$$d_{ij} = \|x_i - x_j\|$$

avec $x_i \in \mathbb{R}^2$, $i, j = 1, 2, \dots, n$ et n représentant le nombre d'objets, un objet désignant ici une observation ou un neurone.

La recherche d'une représentation s'effectue via une transformation de l'espace d'entrée en un espace de sortie qui minimise la fonction d'erreur suivante :

$$E(\{x_i\}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} (d_{ij} - D_{ij})^2$$

Le terme w_{ij} permet de pondérer les distances individuellement mais aussi de normaliser la fonction de d'erreur E afin de la rendre indépendante de l'échelle de valeurs qui est associée aux distances calculées dans l'espace de sortie. La pondération peut prendre plusieurs formes selon que l'on souhaite traiter de la même manière toutes les distances ou que l'on préfère magnifier les structures locales en réduisant l'influence des distances les plus grandes. Dans le premier cas, le facteur de pondération s'écrit :

$$w_{ij} = \frac{1}{\sum_{k=1}^{n-1} \sum_{l=k+1}^n D_{kl}^2}$$

Dans le second cas, le facteur de pondération s'écrit :

$$w_{ij} = \frac{2}{n} (n-1) \frac{1}{D_{ij}^2}$$

Sammon (1969) suggère d'employer une normalisation intermédiaire, qui s'exprime comme indiqué ci-dessous. C'est cette dernière expression qui nous avons utilisée :

$$w_{ij} = \frac{1}{\sum_{k=1}^{n-1} \sum_{l=k+1}^n D_{kl}} \frac{1}{D_{ij}}$$

La minimisation de la fonction d'erreur s'effectue de manière itérative par l'intermédiaire d'une méthode newtonienne de descente de gradient à partir d'une initialisation aléatoire de l'espace de représentation $\{x_i\}$: à chaque étape t du processus, un objet i^* est pris en considération et la recherche du minimum s'opère respectivement à x_{i^*} selon l'équation suivante :

$$x_{i^*}^{(t)} = x_{i^*}^{(t+1)} + \eta \Delta_{i^*}$$

avec Δ_{i^*} , pour chaque valeur de $q = 1, 2$:

$$\Delta_{i^*q} = - \frac{\left(\frac{\partial E}{\partial x_{i^*q}}\right)}{\left(\frac{\partial^2 E}{\partial x_{i^*q}^2}\right)}$$

La descente de gradient est régulée par le terme constant η qui porte le nom de *Magic Factor*, et dont la valeur doit être supérieure à 0 et inférieure ou égale à 1.

Malgré ses propriétés intéressantes, la méthode de *Sammon* présente deux limitations qui jouent de manière dramatique sur les temps de calcul :

- l'algorithme converge toujours, après un certain nombre d'itérations, vers un minimum dont rien ne garantit qu'il corresponde à un minimum global. En effet, la solution est tributaire de l'initialisation de l'espace de représentation, comme une carte de *Kohonen* est tributaire de l'initialisation des poids. Il est donc judicieux d'effectuer plusieurs essais afin de pouvoir retenir la solution qui présente, in fine, l'erreur la plus faible possible.
- le nombre d'étapes nécessaires pour passer en revue tous les objets au cours d'une itération est une fonction quadratique du nombre de ces mêmes objets : il faut, en effet, calculer n^2 mises à jour des vecteurs de sortie pour n objets.

BIBLIOGRAPHIE

- ADYA, M., COLLOPY, F. (1998), How Effective are Neural Networks at Forecasting and Prediction ? A Review and Evaluation, *Journal of Forecasting*, vol. 17, pp. 481-497.
- AGARWAL, A. (1999), Abductive Networks for Two-Group Classification : A Comparison with Neural Networks, *Journal of Applied Business Research*, vol. 15, n° 2, pp. 1-12.
- AGARWAL, A., DAVIS, J. T., WARD, T. (2001), Supporting Ordinal Four-State Classification Decisions Using Neural Networks, *Information Technology and Management*, vol. 2, n° 1, pp. 5-26.
- ALAM, P., BOOTH, D., LEE, K., THORDARSON, T. (2000), The Use of Fuzzy Clustering Algorithm and Self-Organizing Neural Networks for Identifying Potentially Failing Banks : An Experimental Study, *Expert Systems with Applications*, vol. 18, n° 3, pp 185-199.
- ALTMAN, E. I. (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance*, vol. 23, n° 4, pp. 589-609.
- ALTMAN, E. I. (1973), Predicting Railroad Bankruptcies in America, *Bell Journal of Economics*, vol. 4, n° 1, pp. 184-211.
- ALTMAN, E. I., EOM, Y. H. , KIM, D. W. (1995), Failure Prediction : Evidence from Korea, *Journal of International Financial Management and Accounting*, vol. 6, n° 3, pp. 230-249.
- ALTMAN, E. I., HALDEMAN, R., NARAYANAN, P. (1977), Zeta Analysis : A New Model to Identify Bankruptcy Risk of Corporations, *Journal of Banking and Finance*, vol. 1, n° 1, pp. 29-51.
- ALTMAN, E. I., MARCO, G., VARETTO, F. (1994), Corporate Distress Diagnosis : Comparisons Using Linear Discriminant Analysis and Neural Network – The Italian Experience, *Journal of Banking and Finance*, vol. 18, n° 3, pp. 505-529.
- ALTMAN, E. I., NARAYANAN, P. (1997), An International Survey of Business Failure Classification Models, *Financial Markets, Institutions and Instruments*, vol. 6, n° 2, pp. 1-57.
- ANANDARAJAN, M., LEE, P., ANANDARAJAN, A. (2001), Bankruptcy Prediction of Financially Stressed firms : An Examination of the Predictive Accuracy of Artificial Neural Networks, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 10, n° 2, pp 69-81.
- ATIYA, A. F. (2001), Bankruptcy Prediction for Credit Risk Using Neural Networks : A Survey and New Results, *IEEE Transactions on Neural Networks*, vol. 12, n° 4, pp. 929-935.
- AZIZ, A., EMANUEL, D. C., LAWSON, G. C. (1988), Bankruptcy Prediction : An Investigation of Cash Flow Based Models, *Journal of Management Studies*, vol. 25, n° 5, pp. 419-437.
- AZIZ, A., DAR, H. A. (2004), Predicting Corporate Bankruptcy : Whither do We Stand ?, Loughborough University, Department of Economics, Research Paper; n° ERP04-01, 51 p.
- BACK, B., LAITINEN, T., HEKANAHU, J., SERE, K. (1997), The Effect of Sample Size on Different Failure Prediction Methods, *Turku Centre for Computer Science, Technical Report*, n° 155, 23 p.
- BACK, B., LAITINEN, T., SERE, K. (1996 B), Neural Networks and Genetic Algorithms for Bankruptcy Predictions, *Expert Systems with Applications*, vol. 11, n° 4, pp. 407-413.

- BACK, B., LAITINEN, T., SERE, K., VAN WEZEL, M. (1996 A), Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms, Turku Centre for Computer Science, Technical Report, n° 40, 18 p.
- BACK, B., OOSTEROM, G., SERE, K., VAN WEZEL, M. (1994), A Comparative Study of Neural Networks in Bankruptcy Prediction, Proceedings of the 10th Conference on Artificial Intelligence Research in Finland, Finnish Artificial Intelligence Society, pp. 140-148.
- BAEK, J., CHO. S. (2003), Bankruptcy Prediction for Credit Risk Using an Auto-Associative Neural Network in Korean Firms, Proceedings of the IEEE International Conference on Computational Intelligence for Financial Engineering, Hong Kong, March 20-23, pp. 25-29.
- BALCAEN, S., OOGHE H. (2004), Alternative Methodologies in Studies on Business Failure : Do They Produce Better Results than the Classical Statistical Methods ?, Ghent University, Working Paper, n° 2004-249, 40 p.
- BALCAEN, S., OOGHE H. (2006), 35 Years of Studies on Business Failure : An Overview of the Classical Statistical Methodologies and their Related Problems, British Accounting Review, vol. 38, n° 1, pp. 63-93.
- BARANOFF, E.G., SAGER, T. W., SHIVELY, T. S. (2000), A Semiparametric Stochastic Spline Model as a Managerial Tool for Potential Insolvency, Journal of Risk and Insurance, vol. 67, n° 3, pp. 369-396.
- BARDOS, M. (1995), Détection précoce des défaillances d'entreprises à partir des documents comptables, Bulletin de la Banque de France, Supplément « Études », 3^{ème} trimestre, pp. 57-71.
- BARDOS, M. (2001), Développements récents de la méthode des scores de la Banque de France, Bulletin de la Banque de France, n° 90, pp. 73-92.
- BARDOS, M. (2005), Les scores de la Banque de France : leur développement, leurs applications, leur maintenance, Bulletin de la Banque de France, n° 144, pp. 63-73.
- BARNIV, R., AGARWAL, A., LEACH, R. (1997), Predicting the Outcome Following Bankruptcy Filing : A Three-State Classification Using Neural Networks, International Journal of Intelligent Systems in Accounting, Finance and Management, vol. 6, n° 3, pp. 177-194.
- BARNIV, R., AGARWAL, A., LEACH, R. (2002), Predicting Bankruptcy Resolution, Journal of Business Finance and Accounting, vol. 29, n° 3, pp. 497-520.
- BARNIV, R., HERSHBARGER, A. (1990), Classifying Financial Distress in the Life Insurance Industry, Journal of Risk and Insurance, vol. 57, n° 1, pp. 110-136.
- BARNIV, R., MCDONALD, J. B. (1992), Identifying Financial Distress in the Insurance Industry : A Synthesis of Methodological and Empirical Issues, Journal of Risk and Insurance, vol. 59, n° 4, pp. 543-574.
- BARNIV, R., RAVEH, A. (1989), Identifying Financial Distress : A New Non-Parametric Approach, Journal of Business Finance and Accounting, vol. 16, n° 3, pp. 361-383.
- BATTITI, R. (1989), Accelerated Back-propagation Learning : Two Optimization Methods, Complex Systems, vol. 3, n° 4, pp. 331-342.
- BATTITI, R. (1992), First and Second-Order Methods for Learning : Between Steepest Descent and Newton's Method, Neural Computation, vol. 4, n° 2, pp. 141-166.
- BEAVER, W. H. (1966), Financial Ratios as Predictors of Failure, Empirical Research in Accounting, Selected Studies, Journal of Accounting Research, Supplement, vol. 4, pp. 71-111.

- BEAVER, W. H. (1968), Market Price, Financial Ratios and the Prediction of Failure, *Journal of Accounting Research*, vol. 6, n° 2, pp. 179-192.
- BELL, T. B., RIBAR, G. S., VERCHIO, J. R. (1990), Neural Nets Vs. Logistic Regression : A Comparison of Each Model's Ability to Predict Commercial Bank Failures, *Proceedings of the 1990 Deloitte and Touche-University of Kansas Symposium on Auditing Problems*, pp. 29-58.
- BERG, D. (2005), Bankruptcy Prediction by Generalized Additive Models, University of Oslo, Department of Mathematics, *Statistical Report*, n° 1, January, 36 p.
- BEYNON, M. J., PEEL, M. J. (2001), Variable Precision Rough Set Theory and Data Discretisation : An Application to Corporate Failure Prediction, *Omega – International Journal of Management Science*, vol. 29, n° 6, pp. 561-576.
- BILDERBEEK, J. (1979), De continuïteitsfactor als beoordelingsinstrument van ondernemingen, *Accountancy en Bedrijfskunde Kwartaalschrift*, vol. 4, no. 3, pp 58-61.
- BLAZY, R., COMBIER, J. (1997), La défaillance d'entreprise : causes économiques, traitement judiciaire et impact financier », *Insee Méthodes, Economica*, 180 p.
- BLUM, M. (1974), Failing Company Discriminant Analysis, *Journal of Accounting Research*, vol. 12, n° 1, pp. 1-25.
- BOOTH, P. J. (1983), Decomposition Measure and the Prediction of Financial Failure, *Journal of Business Finance and Accounting*, vol. 10, n° 1, pp. 67-82.
- BORITZ, J. E., KENNEDY, D. B. (1995), Effectiveness of Neural Network Types for Prediction of Business Failure, *Expert Systems with Applications*, vol. 9, n° 4, pp. 503-512.
- BOSE, I., PAL, R. (2006), Predicting the Survival or Failure of Click-and-Mortar Corporations : A Knowledge Discovery Approach, *European Journal of Operational Research*, vol. 174, n° 2, pp. 959-982.
- BRABAZON, A., KEENAN, P. B. (2004), A Hybrid Genetic Model for the Prediction of Corporate Failure, *Computational Management Science, Special Issue*, vol. 1, n° 3-4, pp. 293-310.
- BRABAZON, A., O'NEILL, M. (2004), Diagnosing Corporate Stability Using Grammatical Evolution, *International Journal of Applied Mathematics and Computer Science*, vol. 14, n° 3, pp. 363-374.
- BRADLEY, D. B. (2004), Small Business : Causes of Bankruptcy, Small Business Advancement National Center, University of Central Arkansas, College of Business Administration, Research Paper, 2004, 15 p.
- BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., STONE, C. J. (1984), Classification and Regression Trees, Wadsworth International Group, Belmont, California, pp. 43-49.
- BREIMAN, L. (1996), Bagging Predictors, *Machine Learning*, vol. 24, n° 2, pp. 123-140.
- BROCKETT, P. L., COOPER, W. W., GOLDEN, L. L., PITAKTONG, U. (1994), A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency, *Journal of Risk and Insurance*, vol. 61, n° 3, pp. 402-424.
- CHAPADOS, N., BENGIO, Y. (2003), Comment améliorer la capacité de généralisation des algorithmes d'apprentissage pour la prise de décisions financières, Centre interuniversitaire de recherche en analyse des organisations, Montréal, Série Scientifique, n° 2003s-20, 63 p.
- CHARALAMBOUS, C., CHARITOU, A., KAOUROU, F. (2000), Application of Feature Extractive Algorithm to Bankruptcy Prediction, *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, July 24-27, Como, Italy, vol. 5, pp. 303-308.

- CHARITOU, A., NEOPHYTOU, E., CHARALAMBOUS, C. (2004), Predicting Corporate Failure : Empirical Evidence for the UK, *European Accounting Review*, vol. 13, n° 3, pp. 465-497.
- CHARREAUX, G. (1996), *Gestion financière*, Collection Expertise Comptable, 5^{ème} édition, Litec, 810 p.
- CHAVA, S., JARROW, R. A. (2004), Bankruptcy Prediction with Industry Effects, *Review of Finance*, vol. 8, n° 4, pp. 537-569.
- CIBAS, T., FOGELMAN SOULIÉ, F.F., GALLINARI, P., RAUDYS, S. (1994 A), Variable Selection with Neural Networks, *Proceedings of the International Conference on Artificial Neural Networks*, Sorrento, Italy, vol. 2, pp. 1464-1469.
- CIBAS, T., FOGELMAN SOULIÉ, F.F., GALLINARI, P., RAUDYS, S. (1994 B), Variable Selection with Optimal Cell Damage, *Proceedings of the International Conference on Artificial Neural Networks*, Sorrento, Italy, May, pp. 727-730.
- CIELEN, A., PEETERS, L., VANHOOF, K. (2004), Bankruptcy Prediction Using a Data Envelopment Analysis, *European Journal of Operational Research*, vol. 154, n° 2, pp. 526-532.
- COAKLEY, J. R., BROWN, C. E. (2000), Artificial Neural Networks in Accounting and Finance : Modeling Issues, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 9, n° 2, pp. 119-144.
- COATS, P. K., FANT, L. F. (1993), Recognizing Financial Distress Patterns Using a Neural Network Tool, *Financial Management*, vol. 22, n° 3, pp. 142-154.
- COHEN, E. (1997), *Analyse financière*, Economica, 4^{ème} édition, 635 p.
- CYBINSKI, P. J. (1995), A Discrete-Valued Risk Function for Modelling Financial Distress in Private Australian Companies, *Accounting and Finance*, vol. 35, n° 2, pp. 17-32.
- DAMBOLENA, I. G., KHOURY, S. J. (1980), Ratios Stability and Corporate Failure, *Journal of Finance*, vol. 35, n° 4, pp. 1017-1026.
- DASH, M., LIU, H. (1997), Feature Selection for Classification, *Intelligent Data Analysis*, vol. 1, n° 3, pp. 131-156.
- DEAKIN, E. B. (1972), A Discriminant Analysis of Predictors of Business Failures, *Journal of Accounting Research*, vol. 10, n° 1, pp. 167-179.
- DEVENA, L. (1994), Automatic Selection of the Most Relevant Features to Recognize Objects, *Proceedings of the International Conference on Artificial Neural Networks*, vol. 2, pp. 1113-1116.
- DIMITRAS, A., ZANAKIS, S., ZOPOUDINIS, C. (1996), A Survey of Business Failures with an Emphasis on Prediction Methods and Industrial Applications, *European Journal of Operational Research*, vol. 90, n° 3, pp. 487-513.
- DIMITRAS, A. I., SLOWINSKI, R., SUSMAGA, R., ZAPOUNIDIS, C. (1999), Business Failure Prediction Using Rough Sets, *European Journal of Operational Research*, vol. 114, n° 2, pp. 263-280.
- DODIER, R. H., HENZE, G. P. (2004), Statistical Analysis of Neural Networks as Applied to Building Energy Prediction, *Journal of Solar Energy Engineering*, vol. 126, n° 1, pp. 592-600.
- DORSEY, R. E., EDMISTER, R. O., JOHNSON, J. D. (1995), Bankruptcy Prediction Using Artificial Neural Systems, University of Mississippi, School of Business, Working Paper, 72 p.
- DOUMPOS, M., ZOPOUNIDIS, C. (1999), A Multicriteria Discrimination Method for the Prediction of Financial Distress : The Case of Greece, *Multinational Finance Journal*, vol. 3, n° 2, p. 71-101.

- DREYFUS, G., SAMUELIDES, M., MARTINEZ, J. M., GORDON, M. B., BADRAN, F., THIRIA, S., HERAULT, L. (2004), Réseaux de neurones : Méthodologies et applications, Eyrolles, 2^{ème} édition, Paris, 408 p.
- DREZNER, Z., MARCOULIDES, G. A., STOHS, M. H (2001), Financial Application of a Tabu Search Variable Selection Model, *Journal of Applied Mathematics and Decision Sciences*, vol. 5, n° 4, pp. 215-234.
- EATON, H. A., OLIVIER, T. L. (1992), Learning Coefficient Dependence on Training Set Size, *Neural Networks*, vol. 5, pp. 283-288.
- EDMISTER, R. O. (1972), An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction, *Journal of Financial and Quantitative Analysis*, vol. 7, n° 2, pp. 1477-1493.
- EFRON, B., TIBSHIRANI, R. (1995), Cross-validation and the Bootstrap : Estimating the Error Rate of a Prediction Rule, Stanford University, Department of Statistics, Technical Report, n° TR-477, 28 p.
- EISENBEIS, R. A. (1977), Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics, *Journal of Finance*, vol. 32, n° 3, pp. 875-900.
- EL DEREDY, W., BRANSTON, N.M. (1995), Identification of Relevant Features in HMR Tumor Spectra Using Neural Networks, *Proceedings of the 4th International Conference on Artificial Neural Networks*, pp. 454-458.
- EL HENNAWY, R. H., MORRIS, R. C. (1983), The Significance of Base Year in Developing Failure Prediction Models, *Journal of Business Finance and Accounting*, vol. 10, n° 2, pp. 209-223.
- ELMER, P. J., BOROWSKI, D. M. (1988), An Expert System Approach to Financial Analysis : The Case of Savings and Loan Bankruptcy, *Financial Management*, vol. 17, n° 3, pp. 66-76.
- ETHERIDGE, H. L., SRIRAM, R. S. (1997), A Comparison of the Relative Costs of Financial Distress Models : Artificial Neural Networks, Logit and Multivariate Discriminant Analysis, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 6, n° 3, pp. 235-248.
- FAHLMAN, S. E. (1988), An Empirical Study of Learning Speed in Back-Propagation Networks, CMU-CS-88-162
- FAN, A., PALANISWAMI, M. (2000), Selecting Bankruptcy Predictors Using a Support Vector Machine Approach, *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, Como, Italy, July 24-27, vol. 6, pp. 354-359.
- FANNING, K. M., COGGER, K. O. (1994), A Comparative Analysis of Artificial Neural Networks Using Financial Distress Prediction, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 3, n° 3, pp. 241-252.
- FERGUSON, J. A., PATTEN, C. A., SCHROEDER, D. R., OFFORD, K. P., EBERMAN, K. M., HURT R. D. (2003), Predictors of 6-Month Tobacco Abstinence Among 1224 Cigarette Smokers Treated for Nicotine Dependence, *Addictive Behaviors*, vol. 28, n° 7, pp. 1203-1218.
- FITZPATRICK, P. J. (1932), A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms, *Certified Public Accountant*, vol. 12, October, November, December, pp. 598-605, 656-662, 727-731.
- FLAGG, J. C., GIROUX, G. A., WIGGINS, C. E. (1991), Predicting Corporate Bankruptcy Using Failing Firms, *Review of Financial Economics*, vol. 1, n° 1, pp. 67-79.
- FOREMAN, R. D. (2003), A Logistic Analysis of Bankruptcy within the US Local Telecommunications Industry, *Journal of Economics and Business*, vol. 55, n° 2, pp. 135-166.

- FRYDMAN, H., ALTMAN, E. I., KAO, D. (1985), Introducing Recursive Partitioning for Financial Classification : The Case of Financial Distress, *Journal of Finance*, vol. 40, n° 1, pp. 269-291.
- GALLINARI, P., CIBAS, T. (1999), Practical Complexity Control in Multilayer Perceptrons, *Signal Processing*, vol. 74, n° 1, pp. 29-46.
- GENTRY, J. A., NEWBOLD, P., WHITFORD, D. T. (1985), Classifying Bankrupt Firms with Funds Flow Components, *Journal of Accounting Research*, vol. 23, n° 1, pp. 146-160.
- GENTRY, J. A., NEWBOLD, P., WHITFORD, D. T. (1987), Funds Flow Components, Financial Ratios and Bankruptcy, *Journal of Business Finance and Accountancy*, vol. 14, pp. 595-606.
- GENTRY, J. A., SHAW, M. J., TESSMER, A. C., WHITFORD, D. T. (2002), Using Inductive Learning to Predict Bankruptcy, *Journal of Organizational Computing and Electronic Commerce*, vol. 12, n° 1, pp. 39-57.
- GILBERT, L. R., KRISHNAGOPAL, M., SCHWARTZ, K. B. (1990), Predicting Bankruptcy for Firms in Financial Distress, *Journal of Business Finance and Accounting*, vol. 17, n° 1, pp. 161-171.
- GLOUBOS, G., GRAMMATICOS, J. (1988), Success of Bankruptcy Prediction Models in Greece, *Studies in Banking and Finance, International Business Failure Prediction Models*, vol. 7, pp. 37-46.
- GODLEWSKI, C. J. (2003), Modélisation de la prévision de défaillance bancaire : une application aux banques des pays émergents, Université Robert Schuman, Institut d'Etudes Politiques, Working Paper, 35 p.
- GOMBOLA, M. J., HASKINS, M. E., KETZ, J. E., WILLIAMS, D. D. (1987), Cash Flow in Bankruptcy Prediction, *Financial Management*, vol. 16, n° 4, pp. 55-65.
- GOSS, E. P., RAMCHANDANI, H. (1995), Comparing Classification Accuracy of Neural Networks, Binary Logit Regression and Discriminant Analysis for Insolvency Prediction of Life Insurer, *Journal of Economics and Finance*, vol. 19, n° 3, pp. 1-18.
- GRANDVALET, Y. (2001), Bagging Can Stabilize Without Reducing Variance, *Proceedings of the International Conference on Artificial Neural Networks*, Vienna, Austria, August 21- 25, pp. 49-56.
- GRANDVALET, Y. (2004), Bagging Equalizes Influence, *Machine Learning*, vol. 55, n° 3, pp. 251-270.
- GRICE, J. S., INGRAM, R. W. (2001), Tests of the Generalizability of Altman's Bankruptcy Prediction Model, *Journal of Business Research*, vol. 54, n° 1, pp. 53-61.
- GRICE, J. S., DUGAN, M. T. (2003), Reestimations of the Zmijewski and Ohlson Bankruptcy Prediction Models, *Advances in Accounting*, vol. 20, pp. 77-93.
- GUPTA, M. C. (1969), The Effect of Size, Growth, and Industry on The Financial Structure of Manufacturing Companies, *Journal of Finance*, vol. 24 , n° 3, pp. 517-529.
- GUPTA, Y. P., RAO, R. P., BAGCHI, P. K. (1990), Linear Goal Programming as an Alternative to Multivariate Discriminant Analysis : A Note, *Journal of Business Finance and Accounting*, vol. 17, n° 4, pp 593-598.
- GUYON, I., ELISSEEFF, A. (2003), An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research*, vol. 3, n° 7-8, pp. 1157-1182.
- HÄRDLE, W., MORO, R. A., SCHÄFER, D. (2005), Predicting Bankruptcy with Support Vector Machines, Humboldt-Universität, Collaborative Research Center SFB 649 – Economic Risk, Discussion Paper 2005-009, 25 p.

- HASSIBI, B., STORK, D.G. (1993), Second Order Derivatives for Network Pruning : Optimal Brain Surgeon, in SEJNOWSKI, T. J., HINTON, G. E., TOURETZKY, D. S., *Advances in Neural Information Processing Systems*, Morgan Kaufmann, San Mateo, California, vol. 5, pp. 164-171.
- HER, Y. W., CHOE, C. (1999), A Comparative Study of Australian and Korean Accounting Data in Business Failure Prediction Models, La Trobe University, Australia, School of Business, Department of Economics, Working Paper, Series A, n° 99.07, 36 p.
- HORNIK, K., STINCHOMBE, M., WHITE, H. (1989), Multilayer Feedforward Networks are Universal Approximators, *Neural Networks*, vol. 2, n° 5, pp. 359-366.
- HORNIK, K., STINCHOMBE, M., WHITE, H. (1990), Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks, *Neural Networks*, vol. 3, n° 5, pp. 551-560.
- HORNIK, K. (1991), Approximation Capabilities of Multilayer Feedforward Networks, *Neural Networks*, vol. 4, n° 2, pp. 251-257.
- HORNIK, K., STINCHCOMBE, M., WHITE, H., AUER, P. (1994), Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and their Derivatives, *Neural Computation*, vol. 6, n° 6, pp. 1262-1275.
- HORRIGAN, J. O. (1983), Methodological Implications of Non-Normally Distributed Financial Ratios, *Journal of Business Finance and Accounting*, vol. 10, n° 4, pp. 683-689.
- HOWES, P., CROOK, N. (1999), Using Input Parameter Influences to Support the Decisions of Feedforward Neural Networks, *Neurocomputing*, vol. 24, n° 1-3, pp. 191-206.
- HSIEH, S. J. (1993), A Note on the Optimal Cutoff Point in Bankruptcy Prediction Models, *Journal of Business Finance and Accounting*, vol. 20, n° 3, pp. 457-464.
- HUA, Z., WANG, Y., XU, X., ZHANG, B., LIANG, L. (2007), Predicting Corporate Financial Distress Based on Integration of Support Vector Machine and Logistic Regression, *Expert Systems with Applications*, vol. 33, n° 2, pp. 434-440.
- HUANG, C. S., DORSEY, R. E., BOOSE, M. A. (1994), Life Insurer Financial Distress Prediction : A Neural Network Model, *Journal of Insurance Regulation*, vol. 13, n° 2, pp. 131-167.
- HUNTER, J., ISACHENKOVA, N. (2001), Failure Risk : A Comparative Study of UK and Russian Firms, *Journal of Policy Modeling*, vol. 23, n° 5, pp 511-521.
- HWANG, D. Y., LEE, C. F., LIAW, K. T. (1997), Forecasting Bank Failures and Deposit Insurance Premium, *International Review of Economics and Finance*, vol. 6, n° 3, pp. 317-334.
- JAIN, A. K., ZONGKER, D. (1997), Feature Selection: Evaluation, Application and Small Sample Performance, *IEEE Transactions on Pattern Analysis and Machine Intelligence archive*, vol. 19, n° 2, pp. 153-158
- JO, H., HAN, I., LEE, H. (1997), Bankruptcy Prediction Using Case-Based Reasoning, *Neural Networks and Discriminant Analysis*, *Expert Systems with Applications*, vol. 13, n° 2, pp. 97-108.
- JOHN, G. H., KOHAVI, R, PFLEGER, K. (1994), Irrelevant Features and the Subset Selection Problem, in *Machine Learning : Proceedings of the Eleventh International Conference*, New Brunswick, New Jersey, July 10-15, pp. 121-129.
- JONES, S., HENSHER, D. A. (2004), Modelling Corporate Failure : A Multinomial Nested Logit Analysis for Unordered Outcomes, *British Accounting Review* , vol. 39, n° 1, pp. 89-107.

- JONES, S., HENSHER, D. A. (2007), Predicting Firm Financial Distress : A Mixed Logit Model, *Accounting Review*, vol. 79, n° 4, pp. 1011-1038.
- KANE, G. D., RICHARDSON, F. M., MEADE, N. L. (1998), Rank Transformations and the Prediction of Corporate Failure, *Contemporary Accounting Research*, vol. 15, n° 2, pp 145-166.
- KARELS, G. V., PRAKASH, A. J. (1987), Multivariate Normality and Forecasting of Business Bankruptcy, *Journal of Business Finance and Accounting*, vol. 14, n° 4, pp. 573-593.
- KEASEY, K., MCGUINNESS, P. (1990), The Failure of UK Industrial Firms for the Period 1976-1984, *Logistic Analysis and Entropy Measures*, *Journal of Business Finance and Accounting*, vol. 17, n° 1, pp. 119-135.
- KEASEY, K., WATSON, R. (1987), Non-Financial Symptoms and the Prediction of Small Company Failure : A Test of Argenti's Hypothesis, *Journal of Business Finance and Accounting*, vol. 14, n° 3, pp. 335-354.
- KEASEY, K., WATSON, R. (1991), Financial Distress Models : A Review of Their Usefulness, *British Journal of Management*, vol. 2, n° 2, pp. 89-102.
- KIM, M. J., HAN, I. (2003), The Discovery of Experts' Decision Rules from Qualitative Bankruptcy Data Using Genetic Algorithms, *Expert Systems with Applications*, vol. 25, n° 4, pp. 637-646.
- KIM, C. N., JUN, S. G., TAM, K. (2003), A Neural Network Approach to Compare Predictive Value of Accounting Versus Market Data, *International Journal of Digital Management*, Hanyang University, vol. 3, pp. 1-25.
- KIM, C. N., MCLEOD, R. (1999), Expert, Linear Models and Nonlinear Models of Expert Decision Making in Bankruptcy Prediction : A Lens Model Analysis, *Journal of Management Information Systems*, vol. 16, n° 1, pp. 189-206.
- KIVILUOTO, K. (1998), Predicting Bankruptcies with the Self-Organizing Map, *Neurocomputing*, vol. 21, n° 1-3, pp. 191-220.
- KOENIG, G. (1985), Entreprises en difficulté : des symptômes aux remèdes, *Revue Française de Gestion*, Janvier-Février, pp. 84-92.
- KOHAVI, R. (1995), A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection, *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montréal, vol. 2, pp. 1137-1143.
- KOHAVI, R., JOHN, G. H. (1997), Wrappers for Feature Subset Selection, *Artificial Intelligence*, Special Issue on Relevance, vol. 97, n° 1-2, pp. 273-324.
- KOHAVI, R., JOHN, G. H. (1998), The Wrapper Approach, in LIU, H., MOTODA, H., *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers, pp 33-50.
- KOHONEN, T. (1989), *Self-Organization and Associative Memory*, Heidelberg: Springer-Verlag, Berlin, 3d Edition, 255 p.
- KOHONEN, T., HYNINEN, J., KANGAS, J., LAAKSONEN, J. (1996), SOM Pak : The Self-Organizing Map Program Package, Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, 27 p.
- KOLARI, J., GLENNON, D., SHIN, H., CAPUTO, M. (2002), Predicting Large US Commercial Bank Failures, *Journal of Economics and Business*, vol. 54, n° 4, pp. 361-387.

- KOTSIANTIS, S., TZELEPIS, D., KOUMANAKOS, E., TAMPAKAS, V. (2005), Efficiency of Machine Learning Techniques in Bankruptcy Prediction, 2d International Conference on Enterprise Systems and Accounting, Thessaloniki, Greece, July 11-12, pp. 39-49.
- KOUAM, A., BADRAN, F., THIRIA, S. (1992), Approche méthodologique pour l'étude de la prévision à l'aide des réseaux de neurones, Actes du congrès Neuro-Nîmes 92 : Neural Networks and their Applications, Nîmes, 2-6 Novembre, p. 117-127.
- KRUSCHKE, J. K., MOVELLAN, J. R. (1991), Benefits of Gain : Speeded Learning and Minimal Layers in Back-Propagation Networks, IEEE Transactions on Systems, Man and Cybernetics, vol. 21, pp. 273-280.
- KUMAR, N., KROVI, R., RAJAGOPALAN, B. (1997), Financial Decision Support with Hybrid Genetic and Neural Based Modeling Tools, European Journal of Operational Research, vol. 103, n° 2, pp. 339-349.
- KURUPPU, N., LASWAD, F., OYELERE, P. (2003), The Efficacy of Liquidation and Bankruptcy Prediction Models for Assessing Going Concern, Managerial Auditing Journal, vol. 18, n° 6-7, pp. 577-590.
- LA ROCCA, M., PERNA, C. (2003), Modelling complex structures by artificial neural networks, Proceedings of the IASC-INTERFACE-IFCS Workshop on Knowledge Extraction by Modeling, Capri, Italy, September 4-6, pp. 1-12.
- LA ROCCA, M., PERNA, C. (2005), Variable Selection in Neural Network Regression Models with Dependent Data : a Subsampling Approach, Computational Statistics and Data Analysis, vol. 48, n° 2, pp. 415-429.
- LACHER, R.C., COATS, P. K., SHARMA, S. C., FANT, L. F. (1995), A Neural Network for Classifying the Financial Health of a Firm, European Journal of Operational Research, vol. 85, n° 1, pp. 53-65.
- LAITINEN, E. K. (1991), Financial Ratios and Different Failure Processes, Journal of Business Finance and Accounting, vol. 18, n° 5, pp. 649-673.
- LAITINEN, T. M., KANKAANPAA, M. (1999), Comparative Analysis of Failure Prediction Methods : The Finnish Case, European Accounting Review, vol. 8, n° 1, pp. 67-92.
- LAITINEN, E. K., LAITINEN, T. (1995), The Duality of Bankruptcy Process in Finland, European Accounting Review, vol. 4, n° 3, pp. 433-454.
- LAITINEN, E. K., LAITINEN, T. (1998), Cash Management Behaviour and Failure Prediction, Journal of Business Finance and Accounting, vol. 25, n° 7-8, pp. 893-919.
- LAITINEN, E. K., LAITINEN, T. (2000), Bankruptcy Prediction : Application of the Taylor's Expansion in Logistic Regression, International Review of Financial Analysis, vol. 9, n° 4, pp. 327-349.
- LANGLEY, P. (1994), Selection of Relevant Features in Machine Learning, Proceedings of the AAAI Fall Symposium on Relevances, New Orleans, pp. 140-144.
- LANINE, G., VANDER VENNET, R. (2005), Failure Prediction in the Russian Bank Sector with Logit and Trait Recognition Models, Ghent University, Working Paper, n° 2005-329, 31 p.
- LAU, A. H. (1987), A Five-State Financial Distress Prediction Model, Journal of Accounting Research, vol. 25, n° 1, pp. 127-138.
- LE CUN, Y. (1989), Generalization and Network Design Strategies, Technical Report CRG-TR-89-4, University of Toronto, Department of Computer Science, 20 p.
- LE CUN, Y., BOTTOU, L., ORR, G. B. , MULLER, K. R. (1998), Efficient BackProp, in ORR, G. B. , MULLER, K. - R., Neural Networks : Tricks of the Trade, Lecture Notes in Computer Sciences, n° 1524, pp. 5-50.

- LE CUN, Y., DENKER, J. S., SOLLA, S. A. (1990), Optimal Brain Damage, in TOURETZKY, D. S., *Advances in Neural Information Processing Systems*, Morgan Kaufmann, San Mateo, California, vol. 2, pp. 598-605.
- LEE, H., MEHROTRA, K., MOHAN, C., RANKA, S. (1993), Selection Procedures for Redundant Inputs in Neural Networks, *Proceeding of the World Congress on Neural Networks*, vol. 1, pp. 300-303.
- LEE, T. S. (2004), Incorporating Financial Ratios and Intellectual Capital in Bankruptcy Predictions, *Proceedings of the National Taiwan University International Conference in Finance*, Taiwan, December 20-21, 33 p.
- LEE, K., BOOTH, D., ALAM, P. (2005), A Comparison of Supervised and Unsupervised Neural Networks in Predicting Bankruptcy of Korean Firms, *Expert Systems with Applications*, vol. 29, n° 1, pp. 1-16.
- LEE, K. C., HAN, I., KWON, Y. (1996), Hybrid Neural Network Models for Bankruptcy Predictions, *Decision Support Systems*, vol. 18, n° 1, pp. 63-72.
- LEE, S. H., URRUTIA, J. L. (1996), Analysis and Prediction of Insolvency in the Property-Liability Insurance Industry : A Comparison of Logit and Hazardous Models, *Journal of Risk and Insurance*, vol. 63, n° 1, pp. 121-130.
- LELOGEAIS, L. (2003), Un score sur variables qualitatives pour la détection précoce des défaillances d'entreprises, *Bulletin de la Banque de France*, n° 114, pp. 29-46.
- LENNOX, C. (1999), Identifying Failing Companies : A Re-evaluation of the Logit, Probit and DA Approaches, *Journal of Economics and Business*, vol. 51, n° 4, pp. 347-364.
- LENSBERG, T., EILIFSEN, A., MCKEE, T. E. (2006), Bankruptcy Theory Development and Classification Via Genetic Programming, *European Journal of Operational Research*, vol. 169, pp. 677-697.
- LERAY, P. (1998), Apprentissage et diagnostic de systèmes complexes : réseaux de neurones et réseaux bayésiens – Application à la gestion en temps réel du trafic téléphonique français, Thèse de Doctorat, Université Pierre et Marie Curie – Paris VI, 193 p.
- LESHNO, M., SPECTOR, Y. (1996), Neural Network Prediction Analysis : The Bankruptcy Case, *Neurocomputing*, vol. 10, n° 2, pp. 125-147.
- LEV, B., SUNDER, S. (1979), Methodological Issues in the Use of Financial Ratios, *Journal of Accounting and Economics*, vol. 1, n° 3, pp. 187-210.
- LI, X., GUPTA, J. N. D. (2002), Neural Networks in Financial Failure Prediction : A Revisit, *Proceedings of the Annual Decision Sciences Institute Conference*, San Diego, California, pp 233-238.
- LIN, L., PIESSE, J (2004), Identification of Corporate Distress in UK Industrials : A Conditional Probability Analysis Approach, *Applied Financial Economics*, vol. 14, n° 2, pp. 73-82.
- LINDSAY, D. H., CAMPBELL, A. (1996), A Chaos Approach to Bankruptcy Prediction, *Journal of Applied Business Research*, vol. 12, n° 4, pp. 1-9.
- LIU, H., LI, J., WONG, L. (2002), A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Pattern, *Genome Informatics*, vol. 13, n° 51, pp. 51-60.
- LIU, H., YU, L. (2005), Toward Integrating Feature Selection Algorithms for Classification and Clustering, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, n° 3, pp. 1-12.
- LUSSIER, R. N. (1995), A Nonfinancial Business Success Versus Failure Prediction Model for Young Firms, *Journal of Small Business Management*, vol. 33, n° 1, pp. 8-20.

- MALECOT, J. F. (1991), Analyses théoriques des défaillances d'entreprise : une revue de la littérature, *Revue d'économie financière*, n° 19, pp. 205-227.
- MARCO, L. (1989), La montée des faillites en France XIX^{ème}-XX^{ème} siècles, L'Harmattan, Collection Logiques Economiques, Paris, 191 p.
- MCGURR, P. T., DE VANEY, S. A. (1998), Predicting Business Failure of Retail Firms : An Analysis Using Mixed Industry Models, *Journal of Business Research*, vol. 43, n° 3, pp. 169-176.
- MCKAY, D. J. C. (1992), A practical Bayesian framework for Backpropagation Networks, *Neural Computation*, vol. 4, n° 3, pp. 448-472.
- MCKEE, T. E., GREENSTEIN, M. (2000), Predicting Bankruptcy Using Recursive Partitioning and a Realistically Proportioned Data Set, *Journal of Forecasting*, vol. 19, n° 3, pp. 219-230.
- MCKEE, T. E., LENSBERG, T. (2002), Genetic Programming and Rough Sets : A Hybrid Approach to Bankruptcy Classification, *European Journal of Operational Research*, vol. 138, n° 2, pp. 436-451.
- MENSAH, Y. M. (1984), An Examination of the Stationarity of Multivariate Bankruptcy Prediction Models : A Methodological Study, *Journal of Accounting Research*, vol. 22, n° 1, pp. 380-395.
- MERWIN, C. L. (1942), Financing Small Corporations in Five Manufacturing Industries, 1926-1936, National Bureau of Economic Research, Financial Research Program III, Studies in Business Financing, 172 p.
- MESSER, K., KITTLER, J. (1998), Choosing an Optimal Neural Network Size to Aid a Search through a Large Image Database, *Proceedings of the 9th British Machine Vision Conference*, Southampton, 10 p.
- MEYER, P. A., PIFER, H. W. (1970), Prediction of Bank Failures, *Journal of Finance*, vol. 25, n° 4, pp. 853-868.
- MIN, J. H., LEE, Y. C. (2004), Business Failure Prediction with Support Vector Machines And Neural Networks : A Comparative Study, *Proceedings of the 9th Asia-Pacific Decision Sciences Institute Conference*, Seoul, Korea, 14 p.
- MIN, J. H., LEE, Y. C. (2005), Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters, *Expert Systems with Applications*, vol. 28, n° 4, pp. 603-614.
- MIN, S. H., LEE, J., HAN, I. (2006), Hybrid Genetic Algorithms and Support Vector Machines for Bankruptcy Prediction, *Expert Systems with Applications*, vol. 31, n° 3, pp 652-660.
- MONARI, G. (1999), Sélection de modèles non linéaires par leave-one-out : étude théorique et application des réseaux de neurones au procédé de soudage par points, Thèse de Doctorat, Université Pierre et Marie Curie – Paris VI, 154 p.
- MOODY, J. E., UTANS, J. (1992), Principled Architecture Selection for Neural Networks : Application to Corporate Bond Rating Prediction, in MOODY, J. E., HANSON, S. J., LIPPMANN, R. P., *Advances in Neural Information Processing Systems 4*, Morgan Kaufmann Publishers, San Mateo, California, pp. 683-690.
- MOODY, J. E. (1994), Prediction Risk and Architecture Selection for Neural Networks, in CHERKASSKY, V., FRIEDMAN, J.H., WECHSLER, H., *From Statistics to Neural Networks : Theory and Pattern Recognition Applications*, NATO ASI Series F, Springer-Verlag, vol. 136, pp. 147-165.
- MOSSMAN, C. E., BELL, G. G., SWARTZ, L. M.; TURTLE, H. (1998), An Empirical Comparison of Bankruptcy Models, *Financial Review*, vol. 33, n° 2, pp. 35-53.
- MOYER, R. C. (1977), Forecasting Financial Failure : A Re-Examination, *Financial Management*, vol. 6, n° 1, pp. 11-17.

- NAM, J. H., JINN, T. (2000), Bankruptcy Prediction : Evidence from Korean Listed Companies During the IMF Crisis, *Journal of International Financial Management and Accounting*, vol. 11, n° 3, pp. 178-197.
- NANDA, S., PENDHARKAR, P. (2001), Linear Models for Minimizing Misclassification Costs in Bankruptcy Prediction, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 10, n° 3, pp. 155-168.
- NARENDRA, P. M., FUKUNAGA., K. (1977), A Branch and Bound Algorithm for Feature Subset Selection, *IEEE Transactions on Computer*, vol. c-26, n° 9, pp. 917-922.
- NEOPHYTOU, E., MAR-MOLINERO, C. (2004), Predicting Corporate Failure in the UK : A Multidimensional Scaling Approach, *Journal of Business Finance and Accounting*, vol. 31, n° 5-6, pp. 677-710.
- NORTON, C. L., SMITH, R. E. (1979), A Comparison of General Price Level and Historical Cost Financial Statements in the Prediction of Bankruptcy; *Accounting Review*, vol. 54, n° 1, pp. 72-87.
- ODOM, M. C., SHARDA, R. (1990), A Neural Network Model for Bankruptcy Prediction, *Proceedings of the IEEE International Joint Conference on Neural Networks*, San Diego, California, vol. 2, pp. 163-168.
- OOGHE, H., CLAUS, H., SIERENS, N., CAMERLYNCK, J. (1999), International Comparison of Failure Prediction Models from Different Countries : An Empirical Analysis, Ghent University, Working Paper, n° 99-79, 33 p.
- OOGHE, H., JOOS, P., DE VOS, D. (1991), Failure Prediction Models, Gent University, Department of Business Finance, Working Paper, 213 p.
- OOGHE, H., VERBAERE, E. (1982), Determinanten van faling: verklaring en predictie, *Accountancy, Bedrijfsfinanciering en Beleidsinformatie*, pp. 1-166.
- OHLSON, J. A. (1980), Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, vol. 18, n° 1, pp. 109-131.
- O'LEARY, D. E. (1998), Using Neural Networks to Predict Corporate Failure, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 7, n° 3, pp. 187-197.
- PACEY, J. W., PHAM, T. M. (1990), The Predictiveness of Bankruptcy Models : Methodological Problems and Evidence, *Australian Journal of Management*, vol. 15, n° 2, pp. 315-337.
- PANTALONE, C. C., PLATT, M. B. (1987), Predicting Failure of Savings and Loan Associations, *American Real Estate and Urban Economics Association Journal*, vol. 15, n° 2, pp. 46-64.
- PEDERSEN, P. E. (1997), Validating a Connectionist Model of Financial Diagnosis, in WEIGEND, A. S., ABU-MOSTAFA, Y., REFENES, A. P., *Decision Technologies for Financial Engineering*, World Scientific Publishing Company, pp. 138-151.
- PEEL, M., J., PEEL, D., A. (1987), Some Further Empirical Evidence on Predicting Private Company Failure, *Accounting and Business Research*, vol. 18, n° 69, pp. 57-66.
- PENDHARKAR, P. C. (2005), A Threshold-Varying Artificial Neural Network Approach for Classification and its Application to Bankruptcy Prediction Problem, *Computers and Operations Research*, vol. 32, n° 10, pp. 2561-2582.
- PEREZ, M. (2002), De l'analyse de la performance à la prévision de défaillance : les apports de la classification neuronale, Thèse de doctorat, Université Jean-Moulin, Lyon III, 410 p.

- PHILOSOPHOV, L. V., PHILOSOPHOV, V. L. (2002), Corporate Bankruptcy Prognosis : An Attempt at a Combined Prediction of the Bankruptcy Event and Time Interval of its Occurrence, *International Review of Financial-Analysis*, vol. 11, n° 3, pp. 375-406.
- PINCHES, G. E., TRIESCHMANN, J. S. (1974), The Efficiency of Alternative Models For Solvency Surveillance in the Insurance Industry, *Journal of Risk and Insurance*, vol. 41, n° 4, pp. 563-578.
- PIRAMUTHU, S., RAVAGAN, H., SHOW, M. J. (1998), Using Feature Construction to Improve the Performance of Neural Networks, *Management Science*, vol. 44, n° 3, pp. 416-430.
- PLATT, H. D., PLATT, M. B. (1990), Development of a Class of Stable Predictive Variables : The Case of Bankruptcy Prediction, *Journal of Business Finance and Accounting*, vol. 17, n° 1, pp. 31-51.
- PLATT, H. D., PLATT, M. B. (2002), Predicting Corporate Financial Distress : Reflections on Choice-Based Sample Bias, *Journal of Economics and Finance*, vol. 26, n° 2, pp. 184-199.
- PLATT, H. D., PLATT, M. B., PEDERSEN, J. G. (1994), Bankruptcy Discrimination with Real Variables, *Journal of Business Finance and Accounting*, vol. 21, n° 4, pp. 491-510.
- POCHET, C. (2001), Traitement légal de la défaillance et gouvernance : une comparaison internationale, *Revue internationale de droit économique*, vol. 15, n° 4, pp. 465-488.
- POMPE, P. P. M., BILDERBEEK, J. (2005 A), The Prediction of Bankruptcy of Small- and Medium-Sized Industrial Firms, *Journal of Business Venturing*, vol. 20, n° 6, pp. 847-868.
- POMPE, P. P. M., BILDERBEEK, J. (2005 B), Bankruptcy Prediction : The Influence of the Year Prior to Failure Selected for Model Building and the Effects in a Period of Economic Decline, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 13, n° 6, pp. 95-112.
- POMPE, P. P. M., FEELDERS, A. J. (1997), Using Machine Learning, Neural Networks and Statistics to Predict Corporate Bankruptcy, *Microcomputers in Civil Engineering*, vol. 12, n° 4, pp. 267-276.
- PUDIL, P., NOVOVICOVA, J., KITTLER, J. (1994), Floating Search Methods in Feature Selection, *Pattern Recognition Letters*, vol. 15, n° 11, pp. 1119-1125.
- RAHIMIAN, E., SINGH, S., THAMMACHOTE, T., VIRMANI, R. (1993), Bankruptcy Prediction by Neural Networks, in TRIPPI, R. R., TURBAN, E., *Neural Networks in Finance and Investing : Using Artificial Intelligence to Improve Real-World Performance*, Probus Publishing Company, Chicago, pp. 159-171.
- RAHMAN, S., TAN, L. H., HEW, O. L., TAN, Y. S. (2004), Identifying Financial Distress Indicators of Selected Banks in Asia, *Asian Economic Journal*, vol. 18, n° 1, pp. 45-57.
- RAMSER, J., FOSTER, L. (1931), A Demonstration of Ratio Analysis, *Bulletin n° 40*, University of Illinois, Bureau of Business Research, 52 p.
- RAPINEL, L. (1994), Caractéristiques financières des PMI ayant bénéficié d'un plan de redressement judiciaire, *Bulletin de la Banque de France*, 3^{ème} trimestre, Supplément « Études », pp. 55-69.
- RAVI KUMAR, P., RAVI, V. (2007), Bankruptcy Prediction in Banks and Firms Via Statistical and Intelligent Techniques, *European Journal of Operational Research*, vol. 180, n° 1, pp. 1-28.
- REFENES, A.-P. N., ZAPRANIS, A. D. (1999), Neural Model Identification, Variable Selection and Model Adequacy, *Journal of Forecasting*, vol. 18, n° 5, pp. 299-232.

- RICHARDSON, F. M., KANE, G. D., LOBINGIER, P. (1998), The Impact of Recession on the Prediction of Corporate Failure, *Journal of Business Finance and Accounting*, vol. 25, n° 1-2, pp. 167-186.
- RIEDMILLER, M., BRAUN, H. (1993), A Direct Adaptive Method for Faster Backpropagation Learning : the Rprop Algorithm, *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, pp. 586-591.
- RIVALS, I., PERSONNAZ, L. (2003), MLPs (Mono-Layer Polynomials and Multi-Layer Perceptrons) for Nonlinear Modeling, *Journal of Machine Learning Research*, vol. 3, n° 7-8, pp. 1383-1398.
- RUMELHART, D. E., HINTON, G. E., WILLIAMS, R. J. (1986), Learning Internal Representations by Error Propagation, in RUMELHART, D. E. and McCLELLAND, J. L., *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, vol. 1, chap. 8, 318-362.
- SALCEDO-SANZ, S., DEPRADO-CUMPLIDO, M., SEGOVIA-VARGAS, M. J., PÉREZ-CRUZ, F., BOUSOÑO-CALZÓN, C. (2004), Feature Selection Methods Involving Support Vector Machines for Prediction of Insolvency in Non-Life Insurance Companies, *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 12, n° 4, pp. 261-281.
- SALCEDO-SANZ, S., FERNÁNDEZ-VILLACAÑAS, J. L., SEGOVIA-VARGAS, M. J., BOUSOÑO-CALZÓN, C. (2005), Genetic Programming for the Prediction of Insolvency in Non-Life Insurance Companies, *Computers and Operations Research*, vol. 32, n° 4, pp. 749-765.
- SALCHENBERGER, L. M., CINAR, E. M., LASH, N. A. (1992), Neural Networks : A New Tool for Predicting Thrift Failures, *Decision Sciences*, vol. 23, n° 4, pp. 899-916.
- SALMI, T., MARTIKAINEN, T. (1994), A Review of the Theoretical and Empirical Basis of Financial Ratio Analysis, *Finnish Journal of Business Economics*, vol. 4, n° 94, pp. 426-448.
- SAMMON, J. W. (1969), A Nonlinear Mapping for Data Structure Analysis, *IEEE Transactions on Computers*, vol. c-18, n° 5, pp. 401-409.
- SAPORTA, G. (1990), *Probabilités, analyse des données et statistiques*, Technip, Paris, 493 p.
- SARKAR, S., SRIRAM, R. S. (2001), Bayesian Models for Early Warning of Bank Failures, *Management Science*, vol. 47, n° 11, pp. 1457-1475.
- SAUERBREI, W., SCHUMACHER, M. (1992), A Bootstrap Resampling Procedure for Model Building: Application to the Cox Regression Model, *Statistics in Medicine*, vol. 11, n° 16, pp. 2093-2109.
- SEMANI, D., FRÉLICOT, C., COURTELLEMONT, P. (2004), Combinaison d'étiquettes floues/possibilistes pour la sélection de variables, 14^{ème} Congrès Francophone AFRIF – AFIA – Reconnaissance des formes et Intelligence artificielle, Toulouse, 28-30 Janvier, vol. 2, pp. 479-488.
- SEN, T. K., GHANDFOROUSH, P., STIVASON, C. T. (2004), Improving Prediction of Neural Networks : A Study of Two Financial Prediction Tasks, *Journal of Applied Mathematics and Decision Sciences*, vol. 8, n° 4, pp. 219-233.
- SERRANO-CINCA, C. (1996), Self-Organizing Neural Networks for Financial Diagnosis, *Decision Support Systems*, vol. 17, n° 3, pp. 227-238.
- SERRANO-CINCA, C. (1997), Feedforward Neural Networks in the Classification of Financial Information, *European Journal of Finance*, vol. 3, n° 3, pp. 183-202.
- SERRANO-CINCA, C. (2005), Country and Size Effects in Financial Ratios : A European Perspective, *Global Finance Journal*, vol. 16, n° 1, pp. 26-47.

- SETIONO, R., LIU, H. (1997), Neural Network Feature Selector, IEEE Transactions on Neural Networks, vol. 8, n° 3, pp. 654-661, May 1997.
- SEXTON, R. S., SRIRAM, R. S., ETHERIDGE, H. (2003), Improving Decision Effectiveness of Artificial Neural Networks : A Modified Genetic Algorithm Approach, Decision Sciences, vol. 34, n° 3, pp. 421-442.
- SHARMA, D. S., ISELIN, E. R. (2003), The Relative Relevance of Cash Flow and Accrual Information for Solvency Assessments : A Multi-Method Approach, Journal of Business Finance and Accounting, vol. 30, n° 7-8, pp. 1115-1140.
- SHARMA, S., MAHAJAN, V. (1980), Early Warning Indicators of Business Failure, Journal of Marketing, vol. 44, n° 4, pp. 80-89.
- SHIN, K. S., LEE, Y. J. (2003), A Genetic Algorithm Application in Bankruptcy Prediction Modeling, Expert Systems with Applications, vol. 23, n° 3, pp. 321-328.
- SHIN, K. S., LEE, T. S., KIM, H. J. (2005), An Application of Support Vector Machines in Bankruptcy Prediction Model, Expert Systems with Applications, vol. 28, n° 1, pp. 127-135.
- SHIRATA, C. Y. (1998), Financial Ratios as Predictors of Bankruptcy in Japan : An Empirical Research, Proceedings of the Second Asian Pacific Interdisciplinary Research in Accounting Conference, Osaka, August 4-6, pp. 437-445.
- SHUMWAY, T. (2001), Forecasting Bankruptcy More Accurately : A Simple Hazard Model, Journal of Business, vol. 74, n° 1, pp. 101-124.
- SIEDLECKI, W., SKLANSKY, J. (1988), On Automatic Feature Selection, International Journal of Pattern Recognition and Artificial Intelligence, vol. 2, n° 2, pp. 197-220.
- SINKEY, J. F. (1975), A Multivariate Statistical Analysis of the Characteristics of Problem Banks, Journal of Finance, vol. 30, n° 1, pp 21-36.
- SKOGSVIK, K. (1990), Current Cost Accounting Ratios as Predictors of Business Failure : The Swedish Case, Journal of Business Finance and Accounting, vol. 17, n° 1, pp. 137-160.
- SOMOL, P., PUDIL, P., NOVOVICOVA, J, PACLIK, P. (1999), Adaptive Floating Search Methods in Feature Selection, Pattern Recognition Letters, vol. 20, n° 11-13, pp. 1157-1163.
- STAHLBERGER, A., RIEDMILLER, M. (1997), Fast Network Pruning and Feature Extraction Using the Unit-OBS Algorithm, in MOZER, M. C., JORDAN, M. I., PETSCHKE, T., Advances in Neural Information Processing Systems, vol. 9, Cambridge, Massachusetts, The MIT Press, pp. 655-661.
- STOPPIGLIA, H. (1997), Méthodes statistiques de sélection de modèles neuronaux : applications financières et bancaires, Thèse de Doctorat, Université Pierre et Marie Curie – Paris VI, 158 p.
- STRICKER, H. (2000), Réseaux de neurones pour le traitement automatique du langage : conception et réalisation de filtres d'informations, Thèse de Doctorat, Université Pierre et Marie Curie – Paris VI, 287 p.
- SULLIVAN, T. A., WARREN, E., WESTBROOK, J. (1998), Financial Difficulties of Small Businesses and Reasons for their Failure, U.S. Small Business Administration, Working Paper, n° SBA-95-0403, 44 p.
- SUN, L., ETTREDGE, M., SRIVASTAVA, R. P. (2005), A Further Investigation on the Bankruptcy Probability of Firms with Unhealthy Z-score, Proceedings of the 13th Annual Conference on Pacific Basin Finance, Economics and Accounting, June 10-11, New Jersey, 15 p.

- SUNG, T. K., CHANG, N., LEE, G. (1999), Dynamic of Modeling in Data Mining : Interpretive Approach to Bankruptcy Prediction, *Journal of Management Information Systems*, vol. 16, n° 1, pp. 63-85.
- TAFFLER, R. J. (1983), The Assessment of Company Solvency and Performance Using a Statistical Model, *Accounting and Business Research*, vol. 13, n° 52, pp. 295-307.
- TAM, K. Y., KIANG, M. Y. (1990), Predicting Bank Failures : A Neural Network Approach, *Applied Artificial Intelligence*, vol. 4, n° 4, pp. 265-282.
- TAM, K. Y., KIANG, M. Y. (1992), Managerial Applications of Neural Networks : The Case of Bank Failure Predictions, *Management Science*, vol. 38, n° 7, pp. 926-947.
- TAN, C. N. W., DIHARDJO, H. (2001), A Study on Using Artificial Neural Networks to Develop an Early Warning Predictor for Credit Union Financial Distress with Comparison to the Probit Model, *Managerial Finance*, vol. 27, n° 4, pp. 56-77.
- TANG, T. C., CHI, L. C. (2005), Neural Networks Analysis in Business Failure Prediction of Chinese Importers : A Between-Countries Approach, *Expert Systems with Applications*, vol. 29, n° 2, pp. 244-255.
- THEODOSSIOU, P. T. (1991), Alternative Models for Assessing the Financial Condition of Business in Greece, *Journal of Business Finance and Accounting*, vol. 18, n° 5, pp. 697-720.
- THEODOSSIOU, P. T. (1993), Predicting Shifts in the Mean of a Multivariate Time Series Process : An Application in Predicting Business Failures, *Journal of the American Statistical Association*, vol. 88, n° 422, pp. 441-449.
- THORNHILL, S., AMIT, R. (2003), Comprendre l'échec : mortalité organisationnelle et approche fondée sur les ressources, *Statistique Canada, Division de l'analyse microéconomique, Document de recherche, n° 11F0019MIF – 202*, 29 p.
- TIRAPAT, S., NITTAYAGASETWAT, A. (1999), An Investigation of Thai Listed Firms' Financial Distress Using Macro and Micro Variables, *Multinational Finance Journal*, vol. 3, n° 2, pp. 103-125.
- TRESP, V., NEUNEIER, R., ZIMMERMANN, H. G. (1997), Early Brain Damage, in MOZER, M., JORDAN, M. I., PETSCHKE, T., *Advances in Neural Information Processing Systems*, Cambridge, Massachusetts, MIT Press, vol. 9, pp. 669-675.
- TUNG, W. L., QUEK, C., CHENG, P. (2004), GenSo-EWS : A Novel Neural-Fuzzy Based Early Warning System for Predicting Bank Failures, *Neural Networks*, vol. 17, n° 4, pp. 567-587.
- TYREE, E. W., LONG, J. A. (1996), Bankruptcy Prediction Models : Probabilistic Neural Networks versus Discriminant Analysis and Backpropagation Neural Networks, *City University, School of Informatics, Department of Business Computing, Working Paper*, 26 p.
- UTANS, J., MOODY, J. E. (1991), Selecting Neural Network Architectures via the Prediction Risk : Application to Corporate Bond Rating Prediction, *Proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street*, October 9-11, IEEE Computer Society Press, Los Alamitos, California, pp.35-41.
- VAN CAILLIE, D. (2000), La détection des signaux financiers annonciateurs de faillite en contexte PME : une approche méthodologique spécifique, *Actes du 5^{ème} Congrès International Francophone sur la PME*, Lille, 25-27 octobre, 19 p.
- VAN DE LAAR, P., GIELEN, S., HESKES, T. (1997), Input Selection with Partial Retraining, *Artificial Neural Networks*, volume 1327, pp. 469-474.

- VAN GESTEL, T. V., BAESENS, B., SUYKENS, J. A. K., VAN DEN POEL, D., BAESTAENS, D. E., WILLEKENS, M. (2006), Bayesian Kernel-Based Classification for Financial Distress Detection, *European Journal of Operational Research*, vol. 172, n° 3, pp. 979-1003.
- VAN WYMEERSCH, C., WOLFS, A. (1996), La trajectoire de faillite des entreprises : une analyse chronologique sur base des comptes annuels, Facultés Universitaires Notre-Dame de la Paix, Département de Gestion de l'Entreprise, Namur, Working Paper, n° 218, 31 p.
- VARETTO, F. (1998), Genetic Algorithms Applications in the Analysis of Insolvency Risk, *Journal of Banking and Finance*, vol. 22, n° 10-11, pp. 1421-1439.
- VELLIDO, A., LISBOA, P. J., VAUGHAN, J. (1999), Neural Networks in Business : A Survey of Applications (1992-1998), *Expert Systems with Applications*, vol. 17, pp. 51-70.
- VESANTO, J., HIMBERG, J., ALHONIEMI, E., PARHANKANGAS, J. (2000), SOM Toolbox for MatLab 5, Report A57, Laboratory of Computer and Information Science, Helsinki University of Technology, 60 p.
- VIEIRA, A., CASTILLO, P. A., MERELO, J. J. (2003), Application of HLVQ and G-Prop Neural Networks to the Problem of Bankruptcy Prediction, *Proceedings of the 7th International Work-Conference on Artificial and Natural Neural Networks*, Menorca, Spain, June 3-6, vol. 2, pp. 655-652.
- WALLRAFEN, J., PROTZEL, P., POPP, H. (1996), Genetically Optimized Neural Network Classifiers for Bankruptcy Prediction – An Empirical Study, *Proceedings of the 29th Hawaii International Conference on System Sciences*, January 3-6, Maui, Hawaii, vol. 2, pp 419-426.
- WANG, Z. (2004), Financial Ratio Selection for Default-Rating Modeling : A Model-Free Approach and Its Empirical Performance, *Journal of Applied Finance*, vol. 14, n° 1, pp. 20-35.
- WEISS, L. A., CAPKUN, V. (2005), The Impact of Incorporating the Cost of Errors into Bankruptcy Prediction Models, École des HEC Lausanne, Unité d'Enseignement et de Recherche en Comptabilité et Contrôle, Working Paper, 34 p.
- WEST, D., DELLANA, S., QIAN, J. (2005), Neural Network Ensemble Strategies for Financial Decision Applications, *Computers and Operations Research*, vol. 32, n° 10, pp. 2543-2559.
- WESTGAARD, S., VAN DER WIJST, N. (2001), Default Probabilities in a Corporate Bank Portfolio : A Logistic Model Approach, *European Journal of Operational Research*, vol. 135, n° 2, pp. 338-349.
- WHITE, H. (1989), Learning in Artificial Neural Networks : A Statistical Perspective, *Neural Computation*, vol. 1, n° 4, pp. 425-464.
- WILCOX, J. W. (1973), A Prediction of Business Failure Using Accounting Data, *Journal of Accounting Research*, vol. 11, n° 3, pp. 163-179.
- WILSON, N., CHONG, K. S., PEEL, M. J. (1995), Neural Network Simulation and the Prediction of Corporate Outcomes : Some Empirical Findings, *International Journal of the Economics of Business*, vol. 2, n° 1, pp. 31-50.
- WILSON, R. L., SHARDA, R. (1994), Bankruptcy Prediction Using Neural Networks, *Decision Support System*, vol. 11, n° 5, pp. 545-557.
- WINAKOR, A. H., SMITH, R. F. (1935), Changes in the Financial Structure of Unsuccessful Industrial Corporations, University of Illinois, Bureau of Business Research, Bulletin n° 51, 44 p.

- WONG, B. K., BODNOVICH, T. A., SELVI, Y. (1997), Neural Network Applications in Business Neural Network Applications in Business : A Review and Analysis of the Literature (1988–1995), *Decision Support Systems*, vol. 19, n° 4, pp. 301-320.
- WU, C. H., TZENG, G. H., GOO, Y. J. FANG, W. C. (2007), A Real-Valued Genetic Algorithm to Optimize the Parameters of Support Vector Machine for Predicting Bankruptcy, *Expert Systems with Applications*, vol. 32, n° 2, pp. 397-408.
- YACOUB, M., BENNANI, Y. (1997), HVS : A Heuristic for Variable Selection in Multilayer Artificial Neural Network Classifier, *Proceedings of the International Conference on Artificial Neural Networks and Intelligent Engineering*, Missouri, USA, pp. 527-532.
- YANG, Z. R., PLATT, M. B., PLATT, H. D. (1999), Probabilistic Neural Networks in Bankruptcy Prediction, *Journal of Business Research*, vol. 44, n° 2, pp. 67-74.
- YANG, Z. R., HARRISON, R. G. (2002), Analysing Company Performance Using Templates, *Intelligent Data Analysis*, vol. 6, n° 1, pp. 3-15.
- YIM, J., MITCHELL, H. (2002), A Comparison of Corporate Failure Models in Australia : Hybrid Neural Networks, Logit Models and Discriminant Analysis, RMIT Business, School of Economics and Finance, Working Paper, n° 10-2002, 43 p.
- YIM, J., MITCHELL, H. (2005), A Comparison of Corporate Distress Prediction Models in Brazil : Hybrid Neural Networks, Logit Models and Discriminant Analysis, Universidade Federal de Minas Gerais, Economics Department, Nova Economia, 43 p.
- ZAVGREN, C. V. (1985), Assessing the Vulnerability to Failure of American Industrial Firms : A Logistic Analysis, *Journal of Business Finance and Accounting*, vol. 12, n° 1, pp. 19-45.
- ZELLNER, D., KELLER, F., ZELLNER, G. E. (2004), Variable Selection in Logistic Regression Models, *Communications in Statistics, Simulation and Computation*, vol. 33, n° 3, pp. 787-805.
- ZHANG, G. P., PATUWO, B. E., HU, M. Y. (1998), Forecasting with Artificial Neural Networks : The State of the Art, *International Journal of Forecasting*, vol. 14, n° 1, pp. 35-62.
- ZHANG, G., YU, M. Y., PATUWO, B. E., INDRO, D. C. (1999), Artificial Neural Networks in Bankruptcy Prediction : General Framework and Cross-Validation Analysis, *European Journal of Operational Research*, vol. 116, n° 1, pp. 16-32.
- ZHU, Z., HE, H., STARZYK, J. A., TSENG, C. (2007), Self-Organizing Learning Array and its Application to Economic and Financial Problems, *Information Sciences*, vol. 177, n° 5, pp. 1180-1192.
- ZMIJEWSKI, M. E. (1984), Methodological Issues Related to the Estimation of Financial Distress Prediction Models, *Journal of Accounting Research*, vol. 22, Supplement, pp. 59-82.
- ZOPOUNIDIS, C., DOUMPOS, M. (2002), Multi-Group Discrimination Using Multi-Criteria Analysis : Illustrations from the Field of Finance, *European Journal of Operational Research*, vol. 139, n° 2, pp. 371-389.
- ZURADA, J.M., FOSTER B. P., WARD T.J., BARKER R.M. (1998), Neural Networks Versus Logit Regression Models for Predicting Financial Distress Response Variables, *Journal of Applied Business Research*, vol. 15, n° 1, pp. 21-29.

TABLE DES MATIÈRES

REMERCIEMENTS	1
INTRODUCTION	4
PARTIE 1 : LE RÔLE DES RÉSEAUX DE NEURONES DANS LA PRÉVISION DE LA DÉFAILLANCE	11
CHAPITRE 1 : PRÉVISION DE LA DÉFAILLANCE – UN ÉTAT DE L'ART	12
1.1 DÉFINITION ET HORIZON DE LA DÉFAILLANCE	12
1.1.1 MODÈLES DE CESSATION DE PAIEMENT	15
1.1.2 MODÈLES ADOSSÉS À UNE DATE DE DÉFAILLANCE	21
1.1.3 MODÈLES DE PRÉVISION D'UN ÉTAT CRITIQUE	23
1.2 FACTEURS DE DÉFAILLANCE	24
1.3 VARIABLES INCARNANT LES FACTEURS DE DÉFAILLANCE	32
1.4 MODALITÉS DE CONSTITUTION D'UN ÉCHANTILLON	53
1.4.1 ÉCHANTILLONNAGE	54
1.4.2 STATIONNARITÉ ET STABILITÉ DES DONNÉES	66
1.4.3 HISTORIQUE ET HORIZON PRÉVISIONNEL	76
1.4.4 CRITÈRE DE PERFORMANCE ET CRITÈRE DE CLASSIFICATION	79
1.5 CONCLUSION DU CHAPITRE	82
CHAPITRE 2 : CONSTRUCTION DE MODÈLES DE PRÉVISION À L'AIDE DE RÉSEAUX DE NEURONES	84
2.1 TECHNIQUES DOMINANTES	84
2.2 CARACTÉRISTIQUES DES RÉSEAUX DE NEURONES	87
2.2.1 ARCHITECTURE	88
2.2.2 PROPRIÉTÉS	91
2.2.2.1 APPROXIMATION UNIVERSELLE	91
2.2.2.2 PARCIMONIE	92
2.3 APPRENTISSAGE DES RÉSEAUX	93
2.3.1 FONCTION DE COÛT	94
2.3.2 TECHNIQUES D'OPTIMISATION	94
2.3.2.1 DESCENTE DE GRADIENT	95
2.3.2.2 MÉTHODES NEWTONIENNES	97
2.3.2.3 MÉTHODES QUASI-NEWTONIENNES	98
2.3.2.4 METHODE DE LEVENBERG-MARQUARDT	98
2.3.2.5 CHOIX D'UNE TECHNIQUE	99
2.3.3 STRATÉGIES D'APPRENTISSAGE	100
2.3.3.1 CHOIX DES DONNÉES	100
2.3.3.2 SURAJUSTEMENT	101

2.3.3.3 CONTRÔLE DU SURAJUSTEMENT	102
2.3.3.3.1 CONTRÔLE UNE FOIS L'APPRENTISSAGE TERMINÉ	103
2.3.3.3.1.1 ENSEMBLE DE TEST	103
2.3.3.3.1.2 MESURES D'INFORMATION	105
2.3.3.3.2 CONTRÔLE EN COURS D'APPRENTISSAGE	105
2.3.3.3.2.1 ARRÊT PRÉMATURÉ	106
2.3.3.3.2.2 MODÉRATION DES POIDS	106
2.4 PARAMÉTRAGE DES MODÈLES DE DÉFAILLANCE	107
2.5 CONCLUSION DU CHAPITRE	115
CHAPITRE 3 : MÉTHODES DE SÉLECTION DE VARIABLES	117
3.1 PROBLÉMATIQUE DE LA SÉLECTION	118
3.2 CARACTÉRISTIQUES DES MÉTHODES	119
3.2.1 PROCÉDURE DE RECHERCHE	119
3.2.1.1 MÉTHODES COMPLÈTES	119
3.2.1.2 MÉTHODES SÉQUENTIELLES	120
3.2.1.3 MÉTHODES ALÉATOIRES	121
3.2.2 CRITÈRE D'ÉVALUATION	121
3.2.2.1 CRITÈRE INDÉPENDANT	121
3.2.2.2 CRITÈRE DÉPENDANT	123
3.2.3 CRITÈRE D'ARRÊT	124
3.3 SÉLECTION AVEC LES RÉSEAUX DE NEURONES	124
3.3.1 CARACTÉRISTIQUES GÉNÉRALES DES MÉTHODES	125
3.3.1.1 PROCÉDURE DE RECHERCHE	125
3.3.1.2 CRITÈRE D'ÉVALUATION	126
3.3.1.3 CRITÈRE D'ARRÊT	126
3.3.2 MESURES D'ORDRE « ZÉRO »	127
3.3.3 MESURES D'ORDRE « UN »	128
3.3.4 MESURES D'ORDRE « DEUX »	130
3.3.5 ÉVALUATION DES MÉTHODES ET STABILITÉ DES RÉSULTATS	132
3.4 CONCLUSION DU CHAPITRE	133
PARTIE 2 : LES MÉTHODES DE SÉLECTION DE VARIABLES ET LES MODÈLES NEURONAUX DE DÉFAILLANCE	135
CHAPITRE 4 : MÉTHODOLOGIE	136
4.1 MÉTHODES DE CONSTRUCTION DES MODÈLES	136
4.2 MÉTHODES DE SÉLECTION DES VARIABLES	141
4.3 PROTOCOLE DE CHOIX DES VARIABLES	143
4.4 PROTOCOLE DE CRÉATION ET DE VALIDATION DES MODÈLES	144

4.5 ÉCHANTILLONS	146
4.6 VARIABLES.....	150
4.7 RETRAITEMENT DES DONNÉES.....	153
4.8 LOGICIELS EMPLOYÉS ET PROCÉDURES DE CALCUL.....	156
4.9 CONCLUSION DU CHAPITRE	157
 CHAPITRE 5 : STRUCTURES ET PERFORMANCES DES MODÈLES	 158
5.1 VARIABLES SÉLECTIONNÉES.....	159
5.2 STRUCTURE DES MODÈLES.....	163
5.3 PERFORMANCE DES MODÈLES.....	168
5.3.1 PERFORMANCE MESURÉE À L'AIDE DE VARIABLES TIRÉES AU HASARD	169
5.3.2 PERFORMANCE MESURÉE À L'AIDE DE TOUTES LES VARIABLES INITIALES	169
5.3.3 PERFORMANCE MESURÉE SELON LES COUPLES « MÉTHODE DE CLASSIFICATION – TECHNIQUE DE SÉLECTION »	170
5.4 CONCLUSION DU CHAPITRE	180
 CHAPITRE 6 : APPORT DES MODÈLES À LA COMPRÉHENSION DE LA DÉFAILLANCE	 182
6.1 CLASSEMENT DES ENTREPRISES	182
6.2 TOPOLOGIE DES DONNÉES	192
6.3 CONCLUSION DU CHAPITRE	206
 CONCLUSION.....	 208
 ANNEXES	 212
ANNEXE 1 : PRINCIPES DE FONCTIONNEMENT DES CARTES DE KOHONEN	213
ANNEXE 2 : ALGORITHME DE SAMMON.....	217
 BIBLIOGRAPHIE.....	 219

TABLEAUX

Tableau 1 :	Catégories de défaillance déterminées par le droit selon les pays	14
Tableau 2 :	Études dont l'objectif consiste à tenter de discriminer d'autres catégories d'entreprises que les catégories habituelles : « saine » versus « défaillante » ou « en faillite »	16
Tableau 3 :	Résultats des prévisions obtenus grâce à un réseau de neurones sur des échantillons d'entreprises constitués à partir d'une définition juridique de la défaillance	19
Tableau 4 :	Études dont la durée minimale prise en compte entre la date de « défaillance » et la date de mesure des données est inférieure ou supérieure à 1 an	22
Tableau 5 :	Principaux facteurs de défaillance selon la Caisse nationale de marchés de l'État (1977) et Conan et Holder en (1979).....	25
Tableau 6 :	Principaux facteurs de défaillance selon Blazy et Combiér (1997)	26
Tableau 7 :	Facteurs de défaillance et modalités de traduction comptable	30
Tableau 8 :	Principaux facteurs de défaillance des petites entreprises selon Sullivan (1998).....	31
Tableau 9 :	Principaux facteurs de défaillance des petites entreprises selon Bradley (2004)	31
Tableau 10 :	Typologie des variables explicatives utilisées par les modèles de défaillance	33
Tableau 11 :	Types de variables employés pour construire les modèles de prévision	33
Tableau 12 :	Modalités de sélection des variables employées par les études faisant intervenir un réseau de neurones dans l'élaboration des modèles	38
Tableau 13 :	Modalités de sélection des variables employées par les études ne faisant pas intervenir un réseau de neurones dans l'élaboration des modèles	41
Tableau 14 :	Problématiques des études faisant intervenir un réseau de neurones dans l'élaboration des modèles	47
Tableau 15 :	Problématiques des études ne faisant pas intervenir de réseau de neurones dans l'élaboration des modèles	49
Tableau 16 :	Influence de la composition des échantillons d'apprentissage sur les résultats des modèles	56
Tableau 17 :	Critères de sélection des entreprises utilisés pour constituer les échantillons des études sur la défaillance	58
Tableau 18 :	Périodes de prélèvement des entreprises défaillantes utilisées pour constituer les échantillons	68
Tableau 19 :	Taille des échantillons employés par les études opposant des entreprises non défaillantes à des entreprises défaillantes.....	71
Tableau 20 :	Taille des échantillons employés dans les études opposant plus de deux catégories d'entreprises	74
Tableau 21 :	Résultats des prévisions effectuées un et deux ans avant la défaillance	78
Tableau 22 :	Algorithmes neuronaux employés par les principales études sur la défaillance d'entreprise	85
Tableau 23 :	Taille des échantillons et architectures des réseaux de type MLP employés pour la construction de modèles de prévision de la défaillance	108
Tableau 24 :	Modalités de paramétrage des réseaux de type MLP employés pour la construction de modèles de prévision de la défaillance	110
Tableau 25 :	Modalités de validation des modèles de prévision de la défaillance élaborés à partir de réseaux de type MLP	113
Tableau 26 :	Comparaison des critères d'évaluation	123
Tableau 27 :	Synthèse des principaux travaux relatifs à la prévision de défaillance à l'aide de techniques de classification et de régression	137
Tableau 28 :	Techniques de sélection de variables	143
Tableau 29 :	Structures des échantillons	149
Tableau 30 :	Postes comptables inclus dans les ratios des modèles de défaillance	150
Tableau 31 :	Ratios et variables retenus	151

Tableau 32 : Caractéristiques des variables des échantillons d'apprentissage et de test Quartiles – Tests de normalité et de différences entre les groupes	154
Tableau 33 : Variables sélectionnées.....	159
Tableau 34 : Fréquences de sélection des variables	161
Tableau 35 : Fréquences de sélection des variables à l'aide de techniques propres au réseau de neurones	161
Tableau 36 : Classement des variables selon un F de Fisher	162
Tableau 37 : Classement des variables selon le critère d'évaluation.....	165
Tableau 38 : Résultats des prévisions obtenus à partir de ratios tirés aléatoirement et calculés à l'aide des échantillons de test.....	169
Tableau 39 : Résultats des prévisions obtenus à partir de tous les ratios et calculés à l'aide des échantillons de test.....	170
Tableau 40 : Résultats des prévisions obtenus pour différents couples « méthodes de classification – technique de sélection » et calculés à l'aide des échantillons de test	170
Tableau 41 : Résultats des prévisions obtenus pour différents couples « méthodes de classification – technique de sélection » et calculés à l'aide des échantillons de validation	173
Tableau 42 : Résultats des prévisions obtenus par chaque méthode de classification et calculés à l'aide des échantillons de test avec les variables sélectionnées par : Critère : Lambda de Wilks – Procédure de recherche : Stepwise	175
Tableau 43 : Résultats des prévisions obtenus par chaque méthode de classification et calculés à l'aide des échantillons de test avec les variables sélectionnées par : Critère : Vraisemblance – Procédure de recherche : Backward Stepwise	176
Tableau 44 : Résultats des prévisions obtenus par chaque méthode de classification et calculés à l'aide des échantillons de test avec les variables sélectionnées par : Critère : Erreur – Procédure de recherche : Backward.....	177
Tableau 45 : Corrélations des variables aux axes factoriels	183
Tableau 46 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales Critère : Lambda de Wilks – Procédure de recherche : Stepwise.....	185
Tableau 47 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales	186
Tableau 48 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales – Critère : Erreur – Procédure de recherche : Backward	188
Tableau 49 : Moyennes des variables obtenues pour chaque groupe isolé par une analyse en composantes principales – Critère : Mesure Ordre 0 – Procédure de recherche : Backward	190
Tableau 50 : Caractéristiques des groupes d'entreprises selon leur représentation sur les cartes	198
Tableau 51 : Caractéristiques des profils d'entreprises selon leur représentation sur les cartes	199
Tableau 52 : Structure des cartes et performances des modèles.....	200
Tableau 53 : Structure des cartes et performances des modèles par groupe d'entreprises.....	200

FIGURES

Figure 1 : Architecture d'un neurone	88
Figure 2 : Exemple d'architecture d'un réseau de neurones comprenant n entrées, nc neurones cachés et 1 neurone de sortie.....	89
Figure 3 : Processus de sélection de variables	119
Figure 4 : Deux approches de la sélection de variables.....	122
Figure 5 : Cartes des classes et projection de Sammon	195
Figure 6 : Cartes des profils.....	202
Figure 7 : Topologie d'une carte de Kohonen	213
Figure 8 : Processus de mise à jour des poids.....	214