Business School

WORKING PAPER SERIES

Working Paper

2014-321

Prévision de la détresse financière des entreprises françaises: Approche par la régression logistique PLS

Sami BEN JABEUR

http://www.ipag.fr/fr/accueil/la-recherche/publications-WP.html

IPAG Business School
184, Boulevard Saint-Germain
75006 Paris
France

IPAG working papers are circulated for discussion and comments only. They have not been peer-reviewed and may not be reproduced without permission of the authors.

Prévision de la détresse financière des entreprises françaises: Approche par la régression logistique PLS

Sami BEN JABEUR

Université de Bretagne-Sud

5 rue kerfranc 56000 Vannes

benjabeursami@yahoo.fr

Résumé :

L'objectif de cet article est d'appliquer la technique de la régression logistique PLS (Partial Least Squares) à la prévision de la détresse financière des firmes françaises. Cette recherche est motivée par les insuffisances des modèles de prévision traditionnels. L'échantillon est composé de 800 PME françaises pour lesquelles des données comptables et financières ont été collectées et une batterie de 33 ratios financiers a été calculée sur la période 2008-2006. L'objet de notre travail s'inscrit dans la perspective d'analyse et d'explication de la détresse financière. Il vise à montrer en quoi l'horizon de prévision offre une explication claire de la détresse à travers d'indicateurs susceptibles de refléter cette situation empirique.

Mots-clés : détresse financière, modèle Logit, régression PLS

Abstract :

The objective of this article is to apply the technique of the PLS logistic regression to the forecast of the financial distress of the French firm. This research is motivated by the shortcomings of traditional models. The sample consisted of 800 French SMEs for which accounting and financial ratios were calculated over the period 2008-2006.

The purpose of our work is within the context of analysis and explanation of financial distress. It aims to show how the forecast provides a clear explanation of distress through indicators that reflect the empirical situation.

Keywords: financial distress, Logit model, PLS regression.

Introduction

L'objet commun des modèles de prévision de la faillite est de tenter, grâce aux ratios comptables sélectionnés, d'affecter une entreprise quelconque à l'un des deux groupes : les entreprises défaillantes et les entreprises non défaillantes. Les études peuvent être répertoriées en deux classes, selon la méthode de classification utilisée. Les plus nombreuses recourent aux méthodes de discrimination statistiques paramétriques : les analyses discriminantes et les techniques économétriques sur données qualitatives. Dans ce papier nous essayons d'appliquer une nouvelle approche qui est celle de la régression logistique PLS qui trouve son succès dans plusieurs domaines notamment la finance.

I - La détresse financière

Wruck (1990) définit la détresse financière comme étant une situation où les cash-flows sont insuffisants pour couvrir les obligations courantes. Ces obligations peuvent inclure les dettes fournisseurs, les dépenses de litige ou les paiements d'intérêts. Selon Baldwin et Scott (1983), quand la situation d'une entreprise se dégrade au point où elle ne peut faire face à ses contraintes financières, la firme entre dans un état de détresse financière. Ces

mêmes auteurs assurent que cette situation est le résultat d'une mauvaise conjoncture économique, d'un déclin de leurs performances et d'une faible qualité de leurs managements. Modigliani et Miller (1959), ont montré que seules les firmes performantes, efficaces et rentables, sont prêtes à supporter un endettement relativement important, car elles sont en mesure de respecter leurs engagements sans problèmes. Jensen (1986) affirme que l'endettement agit en tant que mécanisme disciplinaire sur les dirigeants et se répercute positivement sur les performances de l'entreprise. Selon Finet (2001), l'octroi de la dette résulte des difficultés subies par l'entreprise, il devient un élément de sa fragilisation des performances d'exploitation, de son capital comme de sa réputation. Opter et Titman (1994) estiment de leur côté que la dette est un facteur de détresse financière, susceptible de mettre en danger la survie de l'entreprise. John (1993) considère, lui, que la détresse financière résulterait donc d'une mauvaise synchronisation entre la disponibilité de liquidités des actifs et de ses obligations contractuelles pendant une période donnée. En d'autres termes, c'est l'impossibilité de faire face au passif exigible avec son actif disponible. Le risque d'insolvabilité, ainsi que la violation des clauses contractuelles des

dettes, est un signal avant coureur de la détresse financière.

Une certaine confusion surgit lorsque le terme « insolvabilité » est utilisé comme étant un synonyme de « détresse financière ». L'insolvabilité peut être interprétée en se rapportant aux actifs ou flux, mais les deux sont souvent confondus. Par exemple, le dictionnaire *New World de Webster* définit l'insolvabilité comme étant d'abord, l'incapacité de payer toutes les dettes et ensuite, comme l'incapacité de payer les dettes, qui deviennent exigibles. L'insolvabilité structurelle, ou à base d'actif, apparaît lorsque la valeur des actifs d'une société est inférieure à la valeur de ses dettes ce qui implique des fonds propres négatifs.

Partant de l'hypothèse de Guilhot (2000) : « *Il importe plus de prévoir les faillites que d'en rechercher les causes* ». L'accroissement actuel de la taille des entreprises en faillite, donc des montants de dettes concernés, rappelle vivement la nécessité de prévoir la défaillance. La protection des intérêts des créanciers et le rétablissement de la pérennité de l'entreprise, passent par l'anticipation des graves difficultés économiques et financières qu'une entreprise est susceptible de rencontrer. Actuellement le problème se pose aux banques avec une acuité particulière. Dans ce cadre, les banques

devront proposer une évaluation systématique des risques qu'elles encourent. Cela implique notamment une estimation précise de la probabilité de défaut de leurs clients, donc un éventuel remaniement de leurs méthodes d'évaluation.

II - La régression logistique PLS

1 - Les limites des modèles probabilistes

Parmi les inconvénients des modèles dichotomiques, qui sont extrêmement sensibles au problème de multicolinéarité, l'inclusion de variables fortement corrélées doit être évitée (Ooghe 1993, 1994 ; Joos, 1998 ;Dumpos et Zopoudinis, 1999) et, comme la régression logistique est basée sur l'analyse des ratios financiers, ces variables sont généralement corrélées parce qu'elles partagent souvent le même numérateur ou le dénominateur.

La régression logistique présente un inconvénient majeur : elle requiert des conditions statistiques strictes. Les variables comptables utilisées doivent en effet suivre une loi multinomiale et leurs matrices de variance-covariance doivent être les mêmes pour l'échantillon des entreprises défaillantes et pour celui des entreprises non-défaillantes. Face à la contrainte d'homoscédasticité, certains auteurs ont recouru aux analyses discriminantes quadratiques,

qui n'exigent que l'hypothèse de multinormalité des ratios (Lachenbruch et al. 1973, Rose et Giroux, 1984 ; Refait, 2004). Mais, malgré le respect de l'hypothèse de multinormalité, l'analyse discriminante quadratique n'est performante que si elle est appliquée à un échantillon très large. Face à ce constat certains auteurs ont préféré recourir à d'autres techniques : les réseaux de neurones (Varetto, 1998 ; Barney et al., 1999, Shin et Lee, 2002 ; Chen et Du, 2009 ; Ravisankar et Ravi, 2010), les techniques de « *support vector machine* » (Vapnik, 1998 ; Hua et al.,2007 ; Ding et al., 2008 ; Li et Sun, 2009 ; Li et al., 2010).

Face au problème de corrélation et pour ne pas supprimer les ratios jugés pertinents pour notre analyse. Plusieurs auteurs (Wold, 1983 ; Tenenhaus, 1998 ; Nguyen et Rocke, 2004 ; Bastien et al., 2005) ont proposé d'autres alternatives plus intéressantes, à savoir la régression PLS que nous allons présenter dans la section suivante.

2 – Algorithme de la régression logistique PLS

La régression PLS univariée est un modèle non linéaire reliant une seule variable dépendante Y à un ensemble de variables indépendantes quantitatives ou qualitatives $X_1, \ldots, X_k$ (Wold 1983). Elle peut

être obtenue par une suite de régressions simples et multiples. En exploitant les tests statistiques associés à la régression linéaire, il est possible de sélectionner les variables explicatives significatives à conserver dans la régression PLS et de choisir le nombre de composantes PLS à retenir.

On cherche de plus à fournir des coefficients de régression interprétables, c'est-à-dire que l'on puisse mesurer la contribution de la variable X_j à l'explication de la variable Y à l'aide du coefficient de régression. Si les signes du coefficient de régression et de corrélation entre la variable explicative et la variable d'intérêt sont opposés, l'interprétation sera délicate pour le praticien.

Cette extension a été proposée par Tenenhaus, (2000). Ces méthodes sont particulièrement utiles, dans des conditions où la régression multiple fonctionne mal. Or, on peut rencontrer des problèmes analogues en régression logistique. Cependant il est tout à fait possible de transposer les principes de la régression PLS à la régression logistique. La régression logistique PLS consiste à adapter l'algorithme de régression PLS1 au cas d'une variable qualitative, binaire ou ordinale (Fischer, 2004 ; Bastien et al., 2005). Il est utile de rappeler que la régression logistique PLS est univariée et ne s'applique donc qu'à la modélisation

d'une seule variable d'intérêt Y que l'on cherche à modéliser à l'aide des variables $X_1,\ldots,X_k$ candidates à l'explication. Nous allons ainsi décrire l'algorithme de régression logistique PLS proposé par Fisher (2004).

Recherche de la première composante $PLSt_1$:

Etape1 : on calcule les coefficients de régression w_{1j} dans la régression logistique simple de Y sur chaque variable X_j.

Etape2 : La composante PLS s'écrit alors :

$$t_1 = \frac{Xw_1}{w_1' w_1}$$

Puis on effectue une régression logistique de Y sur t_1 :

$$\log\frac{\hat{p}}{1-\hat{p}} = \hat{Y} = c_1 t_1 + Y_1$$

où c_1 est le coefficient de régression, Y1 le vecteur des résidus et $\hat{p}$ la probabilité de survenue de l'événement. D'ou une première équation de régression :

$$\hat{Y} = c_1 w_{11}^* x_1 + \ldots + c_1 w_{1k}^* x_k + Y_1$$

avec $$w_1^* = \frac{w_1}{w_1^* w_1}$$

dont les coefficients sont facilement interprétables.

La probabilité de survenue de l'événement pour l'individu « i » est :

$$\hat{p}^i = \frac{\exp(c_1 \sum_{j=1}^{K} w_{1j} X^i{}_j)}{1 + \exp(c_1 \sum_{j=1}^{K} w_{1j} X^i{}_j)}$$

Recherche de la deuxième composante PLS t_2

Etape 1 : si l'on estime que le pouvoir explicatif de ce modèle est faible, on cherche une deuxième composante t_2 , combinaison linéaire de X_j, non corrélée à t_1 et expliquant bien le résidu du Y_1 .Cette composante t_2 est combinaison linéaire des résidus x_{1j} des régressions des variables X_j sur la composante t_1 :

$$X_j = t_1 p'_{1j} + x_{1j}$$

où p_{1j} est le coefficient de régression de chaque X_j sur t_1

Etape2 : pour chaque valeur de j, on calcule le coefficient de régression w_{2j} de x_{1j} dans la régression logistique multiple de Y sur les variables t_1 et x_{1j}.

Etape 3 : la composante PLS t_2 s'écrit alors :

$$t_2 = \frac{x_1 w_2}{w_2' w_2}$$

où x_1 est le résidu de la régression de X sur la première composante t_1.

On peut aussi réécrire la composante t_2 en fonction des variables initiales X_j :

$$t_2 = \frac{w_2}{w_2' w_2}\left(\frac{1 - w_1 p_1'}{w_1' w_1}\right)X = Xw_2^*$$

On procède ensuite à une régression logistique de Y sur t_1, t_2 :

$$\log\frac{\hat{p}}{1-\hat{p}} = \hat{Y} = c_1 t_1 + c_2 t_2 + Y_2$$

où c_1 et c_2 les coefficients de régression logistique de Y sur t_1, t2 :

Pour la deuxième composante la probabilité de survenue de l'événement est :

$$\hat{p} = \frac{\exp(c_1 t_1 + c_2 t_2)}{1 + \exp(c_1 t_1)}$$

$\hat{p}$ s'écrit en fonction des variables explicatives en remplaçant les composantes t_1,t_2 en fonction de X :

$$\hat{p} = \frac{\exp(c_1 X w_1^* + c_2 X w_2^*)}{1 + \exp(c_1 X w_1^* + c_1 X w_2^*)}$$

Recherche de la h-ième composante PLS t_h :

Etape 1 : on calcule le résidu x_{h-1} de la régression de X sur les premières composantes $t_{1,},...,t_{h-1}$. On notera $x_{h-1,j}$ la j-ème colonne de la matrice x_{h-1} :

$$X = t_1 p_1^{'} + ... + t_{h-1} p_{h-1} + x_{h-1}$$

Etape 2 : pour chaque valeur de j, on calcule le coefficient de régression w_{hj} de $x_{h-1,j}$.

Etape 3 : la composante PLS t_h s'écrit alors :

$$t_h = \frac{X_{h-1} w_h}{w_h^{'} w_h}$$

Pour chaque composante t_h calculée, on procède à la régression logistique de Y sur les composantes $t_{1,},...,t_h$. L'équation de régression logistique PLS est obtenue en exprimant cette équation en fonction des variables d'origine :

$$\log \frac{\hat{p}}{1-\hat{p}} = \hat{Y} = c_1 t_1 + ... + c_h t_h = c_1 x w_1^* + ... + c_h X w_h^* = X\alpha$$

où $$\alpha = c_1 w_1^* + ... + c_h w_h^* = \sum_{\ell=1}^{h} c_\ell w_l^*$$

La probabilité de survenue de l'événement de la composante t_h :

$$\hat{p} = \frac{\exp(c_1 X w_1^* + ... + c_h X w_h^*)}{1 + \exp(c_1 X w_1^* + c_h X w_h^*)} = \frac{\exp(X\alpha)}{1 + \exp(X\alpha)}$$

III - Le choix méthodologique

Nous sommes amenés à construire un dispositif méthodologique fondé sur une approche quantitative que nous tenterons de justifier dans un premier temps en montrant comment nous accédons au terrain avant d'expliquer le choix des ratios comptables.

1 - La construction de l'échantillon

A partir de la base des données DIANE (l'accès instantané aux données des entreprises françaises pour l'analyse économique), nous avons établi des critères de sélection des PME constituant notre échantillon. Une entreprise sera considérée défaillante dès lors qu'elle aura fait l'objet d'une première déclaration d'événement judiciaire auprès du tribunal de commerce durant l'année 2009. Les données étudiées sont organisées de telle sorte que l'exercice comptable est disponible pour l'année 2008, 2007 et 2006. L'échantillon final obtenu au

terme de ce processus de sélection rigoureux se compose de 800 entreprises divisées en deux sous-échantillons : 400 entreprises saines et 400 et défaillantes.

2 - La sélection des indicateurs

Nous avons voulu construire un score opérationnel de ratios permettant une détection précoce des difficultés des entreprises à partir de critères objectifs. La méthode des ratios est un outil largement diffusé en analyse financière. Selon Cohen (1997), le recours à cette méthode traduit le souhait de déterminer des seuils, et des normes qui devraient permettre de juger l'état de santé financière et commerciale d'une entreprise.

Le choix de ratios financiers doit donc s'appuyer sur la nature des informations utilisées pour leur calcul, et sur les caractéristiques économiques et financières qu'ils permettent de mettre en évidence. Se pose alors le problème de la subtilité de la procédure de sélection de tels indicateurs. Les ratios sélectionnés ont subi plusieurs choix logiques et méthodologiques dans le but de constituer une batterie pertinente et crédible susceptible de répondre aux objectifs et attentes des analyses externes. Dans le cas présent, nous justifions le choix de nos ratios par :

- leur récurrence dans la littérature française (Bardos 1995, les travaux de banque de France) et internationale (Altman, 1968, 1984 ; Conan et Holder 1979 ; Rose et Giroux, 1984, Refait, 2004).

-leur pertinence par rapport à l'analyse financière, en incorporant les ratios de base existants dans la plupart des modèles de détection de faillite : ratios de liquidité, de rentabilité de gestion, de productivité et de structure financière.

Les variables explicatives retenues, majoritairement des ratios comptables, sont diverses. Rose et Giroux (1984) en ont recensé plus de 130 différentes. Cependant, les caractéristiques des entreprises servant de base à l'étude sont généralement similaires. Conforment à l'enseignement de l'analyse financière, la rentabilité économique ou financière de l'entreprise, la structure de son bilan et sa capacité de remboursement sont les trois éléments les plus corrélés à la défaillance. Selon Bardos (2008), la préparation des ratios économiques et financiers est délicate et longue. En effet, lorsqu'on ne dispose pas de l'ensemble des bilans d'une entreprise, le manque de données ne permet pas de calculer les variations pour toutes les entreprises ni de faire de comparaisons, puisque les résultats seraient aberrants. Ainsi, une série de 33 ratios (R01 à R33) a été retenue (voir annexe I) parmi ceux

couramment utilisés par la littérature et qui présentent un contenu informationnel significatif dans l'analyse de la situation financière de l'entreprise.

IV – Les résultats de l'étude

1 - Les résultats de la régression logistique PLS un an avant la défaillance (Logit-PLS1)

La régression logistique PLS permet de relier la variable Y à l'ensemble des variables explicatives, cette méthode est en quelque sorte une réponse au regret qu'éprouve l'utilisateur d'une régression pas à pas, et de devoir choisir une seule et unique variable explicative par bloc de variables explicatives. Nous présenterons les résultats de la méthode pour les trois horizons de la défaillance.

Les coefficients de la fonction discriminante obtenus à l'aide du logiciel SAS sont les suivants.

$Z_k = 0{,}518 - 0{,}002R02_i - 0{,}002R04_i + 0{,}011R20_i + 0{,}156R23_i - 0{,}005R24_i + 0{,}014R25_i – 0{,}008R26_i$

1.1- Pouvoir discriminant des ratios un an avant la défaillance

La VIP permet de hiérarchiser les ratios selon leur pouvoir explicatif sur les variables cibles. Plus la valeur est grande, plus elle est intéressante (Wold,

1983) ; de manière très simplifiée, nous dirons qu'une variable est à considérer avec attention dès lors que (VIP ≥ 0,8). Dans le tableau qui suit nous résumons ce constat.

Table 1: Poids des ratios dans la fonction discriminante un an avant la défaillance

Ratios	VIP
R20	1,974
R04	1,181
R23	1,077
R26	0,998
R25	0,954
R24	0,858
R02	0,806

Les résultats fournis par le tableau ci-dessus permettent de constater que seuls les ratios R20 (ratio d'exportation), R04 (autonomie financière), R23 (productivité du capital investi), R26 (rentabilité nette), R25 (performance), R24 (rentabilité économique), R02 (indépendance financière) sont les plus discriminants. Les entreprises saines sont caractérisées par un taux d'exportation plus élevé par rapport aux entreprises défaillantes, un niveau de liquidités certes, mais

aussi de solvabilité plus important. Les entreprises saines et défaillantes se différencient également d'une manière significative sur le plan de l'activité commerciale. Les entreprises défaillantes semblent avoir des carnets de commandes, un dynamisme commercial, des niveaux de productivité notamment moins favorables que les entreprises saines. Ces dernières sont en outre plus performantes sur le plan de l'autonomie financière et de la rentabilité économique.

Les entreprises défaillantes se caractérisent par un taux d'endettement moyennement élevé par rapport aux entreprises saines. Plus ce taux est élevé, plus l'entreprise est endettée ; le remboursement des échéances étant un prélèvement inévitable, il peut conduire l'entreprise à la cessation des paiements. Le ratio R02 indique que les entreprises françaises saines, sont relativement peu endettées en comparaison à celles qui sont en difficulté, ce qui confirme l'importance de la contribution des banques dans le financement de ces entreprises. En effet, pour de multiples raisons, notamment pour financer leur croissance, les entreprises ont recours à l'endettement, et plus particulièrement aux crédits bancaires. Le recours au marché financier reste toujours limité même pour les entreprises cotées. L'analyse détaillée de la structure du capital des

entreprises, confirme l'idée précédente, puisque d'après le ratio R04, les entreprises saines disposent d'une autonomie financière plus élevée que les autres entreprises, et d'une capacité de remboursement meilleure. La détermination de l'autonomie financière (R04), a pour but fondamental de mesurer l'indépendance de l'entreprise vis-à-vis de ses créanciers, et d'évaluer le rapport de force entre l'entreprise, les banques et les obligataires.

Les ratios de liquidité générale et réduite indiquent que les entreprises saines présentent un niveau de liquidité plus important que celui des entreprises défaillantes. Cela peut paraître évident, dans la mesure où les entreprises en difficulté, ont du mal à honorer leurs engagements. En effet, lorsque la solvabilité de l'entreprise se dégrade, les banques voient leur rôle s'accroître ; aussi, pour prévenir le risque, elles sont amenées à prendre des mesures de resserrement des conditions de crédit. A l'issue de cette analyse, nous constatons que la liquidité place l'entreprise, au fil du temps, dans l'incapacité de régler sa situation vis-à-vis de ses créanciers, ce qui conduit à une baisse notable de sa rentabilité qui se maintient de manière relativement faible par rapport à celle des entreprises non défaillantes.

La croissance de l'activité des entreprises peut fournir une appréciation sur l'utilisation des moyens de production ainsi que sur les opportunités de croissance potentielles. Toutefois, les entreprises défaillantes se trouvent dans l'impossibilité d'honorer leurs engagements par manque de liquidités à la suite d'une baisse de l'activité ou de la profitabilité. Cette situation, qui peut résulter d'un choix stratégique de recherche de parts de marché ou d'une évolution défavorable des coûts, s'oppose nettement au comportement des entreprises non défaillantes qui, elles, ont augmenté leurs marges. Par conséquent, le niveau bas des ratios indiquant l'activité, la productivité ainsi que de la croissance, valide le principe prédéfini.

Etudier la rentabilité d'une entreprise, c'est apprécier son aptitude à dégager des profits. Au-delà des objectifs de pérennité et de croissance, le but final poursuivi par les propriétaires est d'assurer, par le biais du succès de l'entreprise qu'ils contrôlent, leur propre enrichissement. La rentabilité de l'entreprise peut être appréciée sous une optique d'exploitation, économique ou financière, ce qui permet d'avoir une idée de la rentabilité générale de toutes ses opérations.

L’appréciation de la disponibilité des liquidités renvoie au taux de liquidité immédiate ou future. L’augmentation des créances clients reflète le dynamisme des ventes et souligne le niveau de sa liquidité future. Cependant, comme l’économie dans son ensemble constitue un vaste réseau d’entreprises, il est probable que la faillite d’une entreprise ait, par effet domino, des conséquences désastreuses sur d’autres. De plus, le gonflement des créances traduit une augmentation du besoin en fonds de roulement et des difficultés du paiement des clients notamment des créances douteuses qui forment des risques financiers énormes pour les entreprises.

En effet, plus le niveau de solvabilité, de rentabilité, d’activité et de productivité augmente, plus la fonction de discrimination est élevée et la probabilité de tomber en faillite faible. Les résultats obtenus rejoignent ceux de Keasy et Mc Guinnes (1990), Pompe et Belderbeek (2004), Smith et Liou (2007), Hamza et Bagdadi (2008). Ces derniers estiment qu’il y a une relation causale entre la probabilité de défaillance et la rentabilité des entreprises

1.2 - La classification des entreprises

La régression logistique PLS a permis de retenir quinze composantes, les résultats de classification sont présentés dans le tableau suivant :

Table 2: Validation du modèle (Logit-PLS1)

Un an avant la défaillance		Nature d'entreprise	Groupe d'affectation prévu par le modèle		Total
			Saine	Défaillante	
Groupe d'affectation réelle	Nombre	Saine Défaillante	382 14	18 387	400 400
	%	Saine Défaillante	95,5 3,25	4,5 96,75	100 100

Nous constatons que le modèle a pu distinguer 382 entreprises saines parmi 400, soit un taux de succès de 95,5 %. De surcroît il a permis de classer 387 entreprises défaillantes qui ont effectivement fait faillite, soit donc un taux de 96,75 %. En contrepartie, le taux d'erreur de première catégorie est de l'ordre de 3,25 %, alors que celui de la seconde est de 4,5 %. Nous concluons donc, que la fonction de discrimination, appliquée un an avant la défaillance, a permis de classer 769 entreprises parmi 800 dans le bon groupe soit un taux de bon classement de l'ordre de 96,125 %. La méthode PLS permet de relier l'ensemble des variables

explicatives tandis qu'en régression logistique on écarte les ratios fortement corrélés de l'analyse.

2 - Les résultats de l'analyse discriminante PLS deux ans avant la défaillance (PLS-DA2)

Nous allons procéder de la même manière que précédemment, avec les mêmes entreprises mais, pour l'exercice clos deux ans, avant la défaillance. L'estimation des coefficients du modèle seront formulés comme suit :

$Z_k = 0{,}893 - 0{,}001\ R02_i - 0{,}001R04_i + 0{,}001R10_i + 0{,}009R17_i - 0{,}015R25_i - 0{,}007R26_i$

D'après cette équation, ce ne sont pas les mêmes ratios qui présentent les pouvoirs discriminants les plus importants un et deux avant la défaillance. Ce manque de conformité s'explique par le fait que les modèles de prévision sont construits en utilisant des informations à des horizons différents : un, deux et trois ans avant la défaillance. L'analyse de la capacité de discrimination de ces ratios pour prévoir la défaillance des PME françaises deux ans avant la défaillance s'avère nécessaire.

2.1- Pouvoir discriminant des ratios deux ans avant la défaillance

Pour déterminer l'ordre d'importance des ratios comptables dans la discrimination entre des deux catégories d'entreprises, nous présentons le tableau suivant :

Table 2: Poids des ratios dans la fonction discriminante deux ans avant la défaillance

Ratios	VIP
R26	1,380
R25	1,330
R04	1,176
R17	1,162
R10	0,911
R02	0,878

L'analyse PLS a permis de retenir cinq ratios financiers, le ratio de rentabilité (R26) ainsi que le ratio de rentabilité économique apparaissent comme les variables les plus discriminant**e**s entre les deux catégories d'entreprises, ce qui permet de constater qu'il existe une relation négative entre la rentabilité et le risque d'insolvabilité. D'autres ratios ont été

sélectionnés avec différents critères de pondération tels que les ratios d'endettement (R04, R17 et R02). Le ratio R17 (couverture du chiffre d'affaire par le fonds de roulement) est quant à lui déterminant de la défaillance avec un coefficient de pondération à l'ordre de 1,162. La prise en compte de l'indicateur VIP dans l'évaluation des contributions des ratios dans la discrimination et la prévision de la défaillance des entreprises permet de dégager quelques régularités afin de comprendre le comportement des entreprises françaises. La liquidité, l'endettement, la croissance de l'activité, la solvabilité et l'équilibre financier, représentent les indicateurs les plus significatifs de la situation financière des entreprises. Souvent, ces indicateurs sont utilisés pour évaluer les différentes dimensions de la firme et, de ne fait, ils n'ont pas la même capacité prédictive avant la défaillance. Donc, l'investigation dans les causes de la défaillance permet d'évaluer un ordre d'importance des ratios financiers afin de sélectionner ceux qui peuvent éventuellement être utilisés dans les modèles de prévision.

Le ratio R04, joue un rôle prépondérant dans la discrimination entre les deux groupes d'entreprises. En effet, le montant des créances détenues dans une entreprise informe sur la relation avec la clientèle et

indique le niveau de sa liquidité future. A cet égard, et pour les deux catégories d'entreprises, le niveau de liquidité future constitue un facteur de discrimination entre les entreprises saines et défaillantes, confirmé par le pouvoir prédictif élevé du ratio R04 un et deux ans avant la défaillance. Toutefois, il est important de vérifier s'il y a une rotation lente des créances ou l'existence d'une part non négligeable des créances douteuses qui constituent des risques financiers énormes pour les entreprises. En outre, nous avons remarqué à un et deux ans avant la faillite, que la baisse du niveau d'activité reste un facteur durable et dominant pour l'explication des causes de faillites de nos entreprises. En effet, le déclin des performances correspond à une dégradation durable de l'activité de l'entreprise, ce qui engendre par la suite un besoin en fonds de roulement, une rentabilité insuffisante et enfin une baisse de la liquidité ce qui aboutit à sa liquidation.

Le ratio R10 (durée de crédit clients) est déclaré significatif entre les deux types d'entreprises, le recours à ce type de financement est trop risqué dans la mesure où l'entreprise a peu de contrôle sur cette variable. Au début de son cycle de vie, l'entreprise rencontre d'importants problèmes de reconnaissance, de développement de marché et de

gestion. Cela a été confirmé par Dietsch (1997), en accordant un délai de paiement à un client, toute entreprise s'expose au risque de subir les effets d'un comportement de « mauvais payeur » de celui-ci, ce qui engendre un allongement de la durée du crédit et une augmentation des coûts de surveillance.

2.2- La classification des entreprises deux ans avant la défaillance

Les résultats du pouvoir prédictif de ce modèle sont représentés dans le tableau suivant :

Table 3: Validation du modèle (Logit-PLS2)

Deux ans avant la défaillance		Nature d'entreprise	Groupe d'affectation prévu par le modèle		Total
			Saine	Défaillante	
Groupe d'affectation réelle	Nombre	Saine Défaillante	250 119	150 281	400 400
	%	Saine Défaillante	62,5 29,75	37,5 70,25	100 100

Sur quatre cents entreprises saines, la régression PLS permis d'en distinguer 250, soit un taux de l'ordre de 62,5 %. En revanche, sur quatre cents entreprises défaillantes la méthode a pu reconnaître 281 entreprises, soit un taux de 70,25 %.

Testons maintenant le pouvoir du modèle dans le temps en passant de un à deux ans avant la

défaillance : nous avons constaté une chute du taux de classement des entreprises saines de 95,5 % à 62,5 % et des entreprises défaillantes de 96,75 % à 70.25 %. De ce fait, plus l'horizon de prévision est lointain, plus l'exactitude du modèle diminue.

3 - Les résultats de la régression logistique PLS Trois ans avant la défaillance (Logit-PLS3)

L'estimation des coefficients du modèle seront formulés comme suit :

$Z_k = 0{,}653 - 0{,}008R02_i - 0{,}0021R04_i + 0{,}0003R10_i + 0{,}0040R17_i - 0{,}013R25_i - 0{,}0001R26_i + 0{,}003R27_i$

3.1- Pouvoir discriminant des ratios trois ans avant la défaillance

Pour déterminer l'ordre d'importance des ratios comptables dans la discrimination entre des deux types d'entreprises, nous présentons le tableau suivant :

Table 5: Poids des ratios dans la fonction discriminante trois ans avant la défaillance

Ratios	VIP
R04	1,575
R10	1,090
R02	1,023
R26	0,962

R25	0,927
R17	0,893
R27	0,890
R33	0,873

Le tableau ci-dessus, indique que l'analyse discriminante parvient à prévoir la situation financière future des entreprises en accordant une importance inégale aux ratios financiers utilisés. En effet, nous signalons depuis le début, que le ratio d'activité R04 a été sélectionné dans les premiers pas, avec un VIP de l'ordre de 1,575, ce qui nous renseigne sur le niveau d'endettement au sein de l'entreprise. Ce résultat est conforme à ceux d'Altman et al. (1994), Mensah (1984), Platt et Platt (1991), Chralambous et al. (2000), Liou et Smith (2007), plus le niveau d'endettement est élevé, plus la situation la situation financière de l'entreprise se dégrade. La variable « durée de crédit clients » est sélectionnée en deuxième position pour les trois horizons de prévision de la défaillance. Une PME avant sa faillite accorde des délais plus longs, ce qui tient au laxisme de sa gestion commerciale, au manque de suivi des créances, au souci de garder ou d'élargir la clientèle. L'entreprise prend des clients sans oser leur imposer des conditions des paiements strictes. Ce résultat corrobore à ceux de Dietsch

(1997), Pacitto (1998). Le ratio de performance R02 figure en troisième position. Ce constat est conforme à ceux de Taffler (1982), Weiss (1996), Mossman et al. (1998), Atiya (2001), Pompe et Belderbak (2004), il existe une relation négative entre la rentabilité et la défaillance d'entreprises.

3.2- La classification des entreprises trois ans avant la défaillance

La régression logistique PLS a pu retenir quinze composantes, les résultats de cette classification sont présentés dans le tableau suivant :

Table 6: Validation du modèle (Logit-PLS3)

Trois ans avant la défaillance		Nature d'entreprise	Groupe d'affectation prévu par le modèle		Total
			Saine	Défaillante	
Groupe d'affectation réelle	Nombre	Saine Défaillante	234 161	166 239	400 400
	%	Saine Défaillante	58,5 40,25	41,5 59,75	100 100

Ce tableau nous permet d'observer que le taux de bons classements des entreprises est de l'ordre de 59,125 %, soit un taux d'erreurs de seconde espèce de l'ordre de 41,5 % et un taux d'erreur de première espèce de l'ordre de 40,25 %.

En ce qui concerne le pouvoir prédictif du modèle dans le temps, en passant de deux à trois ans avant la défaillance, nous avons constaté une chute du taux de classement des entreprises saines de 62,5 % à 58,5 %, et des entreprises défaillantes de 70,25 % à 59,75 %. De ce fait, plus l'horizon de prévision est lointain, plus la pertinence et l'efficacité du modèle diminuent. Les résultats de notre étude rejoignent ceux de Bardos et Zhu (1997), Bardos (1998), Pompe et Belderbeek (2004), Platt et Platt (1999), Hamza et Baghdadi (2008). Ils sont toutefois moins probants que ceux d'Altman (1968), Marco et Varetto (1994) et Varetto (1998). Ces dernières études obtiennent en effet des résultats extrêmement fiables qui dépassent les 90 % le taux de bons classements en moyenne sur les différents horizons de prévision. La plupart des travaux proposent, comme notre étude, un horizon de prévision à trois ans (Altman, 1968, Marco et Varetto, 1994 ; Bardos et Zhu, 1997 ; Varetto, 1998, Jones et Hensher, 2007, Ding et al., 2008), alors que d'autres sont réalisés cinq ans avant la défaillance (Yang et al., 1999 ; Pompe et Bilderbeek, 2004), ou encore sont réalisées sur les échantillons parfois supérieurs à 4 000 entreprises et sur une période de 4 à 13 ans. Par rapport aux travaux proposés par la littérature, notre étude présente un modèle qui porte sur 33

ratios comptables, alors que dans la plupart des cas, l'étude est faite à partir d'une moyenne de 7 ratios. Certains auteurs, notamment Pompe et Bilderbeek (2004), utilisent une batterie riche en ratios comptables et financiers composée de 43 ratios, et obtiennent des résultats qui se situent dans la moyenne de l'ensemble des travaux. Enfin, les études qui proposent les horizons de prévision les plus éloignés sont celles qui offrent les taux de bon classement les moins élevés.

La régression logistique PLS peut être considérée comme une méthode de sélection de variables, alternative à la régression pas à pas. La régression logistique PLS est, en quelque sorte, une réponse au regret qu'éprouve l'utilisation d'une régression pas à pas, d'être contraint de choisir une seule variable explicative dans chaque bloc de variables explicatives. En régression PLS on remplace ce bloc de variables par un résumé : la composante PLS. En régression pas à pas il y a exclusion de variables éventuellement importantes alors qu'en régression PLS toutes les variables importantes sont conservées et les variables sans importance sont soit exclues, soit retenues avec un poids faible.

Conclusion

Comme l'ont prouvé la majorité la majorité des études antérieures ainsi que d'autres plus contemporaines dans ce domaine de recherche, nous constatons que le pouvoir prédictif s'affaiblit pour les modèles qui utilisent des informations de moins en moins récentes, ou pour les modèles qui sont destinés à être utilisés pour prévoir la défaillance des entreprises sur un horizon de plus en plus lointain. Ceux-ci sont appréciés par la supériorité du modèle PLS1 par rapport à PLS2 et PLS3.

La méthode PLS donne des résultats satisfaisants par rapport aux autres méthodes statistiques : lorsque les ratios ne suivent pas des lois multinomiales, et qu'une loi logistique ou une loi normale ne semble pas adaptée pour décrire les erreurs, le recours aux méthodes non-paramétriques est une solution, puisqu'elles ne nécessitent aucune hypothèse relative à la distribution des variables.

L'ensemble des travaux antérieurs sont hétérogènes sur des dimensions telles que : la taille de l'échantillon, la longueur de la période de l'étude, l'horizon de prévision, la démarche de validation externe du modèle estimé. De ce fait, il devient difficile de comparer le résultat final en termes de prédiction en raison de ces disparités de nature

méthodologique. Enfin, la plupart des travaux empiriques sont réalisés à partir d'échantillons de grande taille qui ne sont pas équilibrés puisqu'ils ne font pas la distinction entre entreprises saines et défaillantes, et portent sur des prédictions à moyen terme effectuées avec un nombre réduit d'indicateurs comptables. La régression PLS permet de pallier ce dernier problème en intégrant un grand nombre de ratios dans le modèle ; en outre, elle permet de résoudre le problème de corrélation, tel que la prise en compte des données manquantes dans la matrice.

Bibliographie

Abid F., Zouari A. (2000), *Financial distress prediction using neural networks: the Tunisian experience*, in the proceedings of the MS' 2000 international conference on modelling and simulation, SPAIN, pp.399-406.

Abid F., Zouari A. (2001), *Predicting corporate financial distress: A neural networks approach, in proceedings of the international finance conference*, Tunisia, pp.183-195.

Altman E.I. (1968), *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, Journal of Finance, 23(4), September, 589-609.

Altman E.I. (1984), *A Further Empirical Investigation of the Bankruptcy Cost Question*, The Journal of Finance, 39(4), september, 1067-1089.

Amemiya T. (1981), *Qualitative Response Models: A Survey*, Journal of Economic Litterature, 19(4), 481-536.

Anderson J.A. (1972), *Separate sample logistic discrimination*, Biometrika, 59, 19-35.

Bardos M. (2008), *Scoring sur données d'entreprises : instrument de diagnostic individuel et outil d'analyse de portefeuille d'une clientèle*, Revue MODULA, 38, 159 – 177.

Bardos M. (2001) : Analyse discriminante: application au risque et scoring financier, Dunod.

Bardos M. (1998), *Detecting the risk of company failure at the Banque of France*, Journal of Banking and Finance, 22, 1405-1419.

Bardos M., Zhu W.H. (1997), *Comparaison de l'analyse discriminante linéaire et des réseaux neuronaux : application à la détection de défaillance d'entreprises*, Revue Statistique Appliquée, XLV (4), 65-92.

Baldawin C. Y, Scott P. M. (1983), *The resolution of Claims in financial distress: The case of Massey Ferguson*, Journal of Finance, 38, 505-516.

Bastien P, Vinzi V. E, Tenenhaus M. (2005), *PLS generalised linear regression*, Computational *Statistics and Data Analysis, 48-1.17-46.*

Ben Jabeur.S. (2009), *Prévision de la détresse financière des entreprises : approche par l'analyse discriminante PLS*, AFFI 2009 colloque international de l'association française de la finance, 13-14-15 Mai, Brest.

Chen W-S., Yin-Kuan Du Y-K.(2009), Using neural networks and data mining techniques for the financial distress prediction model, Expert Systems with Applications 36, 4075-4086.

Charalambous C, Charitou A et Kaourou F. (2000), *Comparative analysis of artificial neural network*

models: application in bankruptcy prediction, Annals of operations research, Octobre, university of Cyprus.

Fisher N. (2004), *Fusion statistique de fichiers de données*, Thèse de doctorat, CNAM, Paris.

Fort G., Lambert-Lacroix S. (2005), *Classification using Partial Least Squares with Penalized Logistic Regression*, Bioinformatics, 21(7), 1104-1111.

Gourieroux C. (1984), *Econométrie des variables qualitatives*, Economica, Paris.

Guilhot B. (2000), *Défaillances d'entreprise : soixante-dix ans d'analyse théoriques et empiriques*, Revue Française de Gestion, 130, 52-67.

Hamza T., Bagdadi K. (2008), *Profil et déterminants financiers de la défaillance des PME tunisiennes (1993-2003)*, Bnque & Marchés, 93, 45-62.

Li H., Jie Sun J. (2009), *Predicting business failure using multiple case-based reasoning combined with support vector machine*, Expert Systems with Applications 36, 10085–10096.

Li H, Sun J, Wu J. (2010), *Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods*, Expert Systems with Applications 37, 5895–5904

Nguyen D., Rocke D. (2004), *On partial least squares dimension reduction for microarraybased classification: a simulation study*, Comput. Stat. Data Anal, 46, 407-425.

Ohlson J.A. (1980), *Financial ratios and the probabilistic prediction of bankruptcy*, Journal of Accounting Research, spring, 109-131.

Pages J., Tenenhaus M. (2001), *Multiple factor analysis combined with PLS path modelling*, Revue de Statistique Appliquée, XLIV (2), 35-60.

Refait C. (2004), *La prévision de la faillite fondée sur l'analyse financière de l'entreprise: un état des lieux*, Économie et Prévision, 162, 129-147.

Tenenhaus M. (1998), *La régression PLS : théorie et pratique*, Technip, Paris.

Tenenhaus M. (1995), *Nouvelles méthodes de régression PLS*, Les cahiers de recherche, CR 540.

Tenenhaus M. (2000), *La Régression Logistique PLS*, Journées d'Etudes en Statistique, Modèles Statistiques pour données Qualitatives, 261-273.

Tenenhaus M., Gauchi J.P., Menardo C. (1995), *Régression PLS et Applications*, Revue de Statistiques Appliquée, 43(1), 7-63.

Wei-Sen C., Yin-Kuan D. (2009), *Using neural networks and data mining techniques for the financial distress prediction model*, Expert Systems with Applications 36, 4075–4086.

Wruck K.H. (1990), *Financial distress, reorganization, and organizational efficiency*, Journal of Financial Economies, 27, 419-444.

Annexe I : Liste de 33 ratios financiers de l'étude

Type	Nom	Codage
Structure et Liquidité	Equilibre financier	R01
	Indépendance financière	R02
	Endettement	R03
	Autonomie financière	R04
	Degré d'amortissement des immobilisations corporelles	R05
	Financement de l'actif circulant net	R06
	Liquidité générale	R07
	Liquidité réduite	R08
Gestion	Rotation des stocks	R09
	Durée du crédit clients	R10
	Durée du crédit fournisseurs	R11
	Taux d'intérêt financier	R12
	Intérêt / Chiffre d'affaires	R13
	Endettement global	R14
	Taux d'endettement	R15

	Capacité d’autofinancement	R16
	Couverture du chiffre d’affaires par le fonds de roulement	R17
	Couverture du chiffre d’affaire par le besoin en fonds de roulements	R18
	Poids du besoin en fonds de roulement d'exploitation	R19
	Exportation	R20
	Productivité de potentiel de production	R21
Productivité et Rentabilité	Productivité du capital financier	R22
	Productivité du capital investi	R23
	Rentabilité économique	R24
	Performance	R25
	Rentabilité nette	R26
	Rendements des capitaux propres nets	R27
	Rendement des ressources durables nettes	R28
Marge et valeur	Taux de valeur ajoutée	R29
	Part des salariés	R30
	Part de l’état	R31

ajoutée	Part des prêteurs	R32
	Part de l’autofinancement	R33