

L'UTILISATION DU MODÈLE DE COX-PLS DANS LA PRÉVISION DE DÉFAILLANCE DES ENTREPRISES

Sami Ben Jabeur

IPAG Business School, 184 Boulevard Saint-Germain, 75006 Paris
sbenjabeur@gmail.com

Résumé. L'objectif de cet article est d'appliquer le modèle Cox-PLS (Partial Least Squares) à la prévision de la détresse financière d'entreprises françaises afin d'estimer leur risque de défaut et d'élaborer un indicateur du risque de faillite. Cette recherche est motivée par les insuffisances des modèles de prévision traditionnels. L'échantillon est composé de 800 petites et moyennes entreprises françaises pour lesquelles des données comptables et financières ont été collectées et une batterie de 33 ratios financiers a été calculée sur la période 2006-2008. Les prévisions issues de l'analyse Cox-PLS sont comparées à celui de modèle de Cox ; les probabilités de défaut estimées constituent un bon indicateur en termes de prévision du risque de faillite de un à trois ans avant le dépôt de bilan.

Mots-clés. Modèle de Cox, régression PLS, défaillance d'entreprises

Abstract. The objective of this paper is to apply the Cox-PLS model (partial least squares) to the prediction of financial distress of French firms to estimate the risk of corporate default and develop a bankruptcy risk indicator. This research is motivated by the inadequacies of traditional forecasting models. The sample consists of 800 small and medium-sized French companies where accounting and financial data was collected and a battery of 33 financial ratios were calculated over the period 2006-2008. The forecast from the Cox-PLS analysis are compared to that of the Cox model; the estimated default probabilities are a good indicator in terms of forecasting bankruptcy risk of one to three years before filing for bankruptcy.

Keywords. Cox model, PLS regression, business failure

1 Introduction

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

2 Le modèle de Cox-PLS

La régression PLS univariée est un modèle non linéaire reliant une seule variable dépendante Y à un ensemble de variables quantitatives ou qualitatives X_1, \dots, X_k (Wold 1983). Elle peut être obtenue par une suite de régressions simples et multiples. En exploitant les tests statistiques associés au cadre de la régression linéaire classique, 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 , $1 \leq j \leq k$, à l'explication de la variable Y . Il existe plusieurs versions de l'algorithme de régression PLS univariée. Elles diffèrent au niveau des normalisations et des calculs intermédiaires.

Selon Bastien et al. (2005), Fisher (2004), l'algorithme de la régression PLS peut se décomposer comme suit.

On construit tout d'abord une composante PLS T_1 avec : $T_1 = w_{11}X_1 + \dots + w_{1k}X_k$

où :

$$w_{1j} = \frac{\text{cov}(Y, X_j)}{\sqrt{\sum_{j=1}^k \text{cov}(Y, X_j)^2}}, 1 \leq j \leq k$$

Ces coefficients sont ensuite normalisés :

$$w_{1j}^* = \frac{w_{1j}}{\sqrt{\sum_{j=1}^k (w_{1j})^2}}, 1 \leq j \leq k$$

Puis on effectue une régression simple de Y sur T_1 :

$$Y = c_1 T_1 + \varepsilon_1$$

où c_1 est le coefficient de régression et ε_1 le vecteur des résidus.

Toutefois, si l'on estime que le pouvoir explicatif de ce modèle n'est pas suffisant, on cherche une deuxième composante T_2 , combinaison linéaire des X_j , non corrélée à T_1 et expliquant bien le résidu ε_1 .

Cette composante T_2 est combinaison linéaire des résidus χ_{1j} des régressions des variables X_j sur la composante T_1 :

avec
$$\chi_{1j} = X_j - p_{1j} T_1, 1 \leq j \leq k$$

et
$$p_{1j} = \frac{\text{cov}(X_j, T_1)}{\text{var}(T_1)}, 1 \leq j \leq k$$

On obtient alors T_2 :

$$T_2 = w_{21}\chi_{11} + \dots + w_{2k}\chi_{1k}$$

On procède ensuite à une régression multiple de Y sur T_1 et T_2 :

$$Y = c_1 T_1 + c_2 T_2 + \varepsilon_2$$

où c_1 et c_2 sont les coefficients de régression et ε_2 le vecteur des résidus.

On peut itérer ce processus en utilisant de la même manière les résidus $\varepsilon_2, \chi_{21}, \dots, \chi_{2k}$ des régressions de Y, X_1, \dots, X_k sur T_1 et T_2 .

Le nombre de composantes T_h à retenir est habituellement déterminé par validation croisée.

Pour chaque valeur h , on calcule des prédictions de la variable d'intérêt Y . Plus précisément, pour un échantillon de n individus, on calcule les prédictions \hat{Y}_{hi} de Y_i , $1 \leq i \leq n$, en utilisant tous les individus, puis $\hat{Y}_{h(-i)}$ sans utiliser l'individu i . Puis on calcule les critères RSS_h (Residual Sum of Squares) et $PRESS_h$ (Prediction Error Sum of Squares) définis par :

$$RSS_h = \sum_{i=1}^n (Y_i - \hat{Y}_{hi})^2$$

et

$$PRESS_h = \sum_{i=1}^n (Y_i - \hat{Y}_{h(-i)})^2$$

Selon Umetri (1996), la composante T_h sera retenue si :

$$\sqrt{PRESS_h} \leq 0.95 \leq \sqrt{RSS_{h-1}}$$

Une fois que l'on a retenu le nombre des composantes, la prédiction de la fonction de risque de défaillance s'effectue ensuite en ajustant le modèle de Cox sur les composantes, non pas sur les variables d'origine.

Le principe de l'approche Cox-PLS étant maintenant introduit, nous allons présenter les caractéristiques de l'échantillon utilisé puis les résultats de l'étude.

3 Données

Une batterie de trente trois ratios financiers ($R01, \dots, R33$) a été calculée (voir annexe 1). Ces ratios ont été choisis parmi ceux qui sont couramment utilisés dans la littérature financière (Altman 1968, Deakin 1972, Ohlson 1980, Hamer 1983, Liou et Smith 2007) ou qui présentent un contenu informationnel significatif dans l'analyse de la situation financière des entreprises. La construction des modèles de prévision est faite sur la base d'un échantillon de 800 PME françaises appartenant à différents secteurs d'activité issu de la base des données DIANE (l'accès instantané aux données des entreprises françaises pour l'analyse économique), divisé en deux groupes de 400 entreprises chacun, le groupe des entreprises saines et le groupe des entreprises défaillantes. Une entreprise sera considérée défaillante dès lors qu'elle aura fait l'objet d'une liquidation 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 les années 2008, 2007 et 2006. Pour la construction du modèle de prévision, la variable explicative Y est la durée de vie de l'entreprise calculée à partir de la base des données DIANE.

4 Résultats

Les corrélations entre les variables potentiellement explicatives sont données dans le tableau suivant.

Table 1: Tableau de corrélation

Variables explicatives	Coefficient de Pearson
(R01,R21)	0.806
(R06, R22)	-0.954
(R07,R08)	0.983
(R07, R17)	0.950
(R07,R18)	0.884
(R07,R19)	0.611
(R12,R13)	0.508
(R14, R18)	0.760
(R17,R18)	0.922
(R18, R08)	0.855
(R18,R19)	0.760

Afin d'estimer les paramètres des modèles, nous avons utilisé l'ensemble de l'échantillon. Ces données posent problème en régression de Cox, à cause des fortes corrélations entre les variables explicatives, ce qui peut conduire à exclure à tort du modèle plusieurs ratios pourtant jugés pertinents dans l'explication de la défaillance. En régression PLS, on remplace ce bloc de variables par la(es) composante(s) PLS et toutes les variables importantes sont conservées. Les résultats de notre étude montrent que les modèles de régression Cox-PLS sont un outil intéressant pour la prédiction du risque de défaillance des entreprises. Ce constat nous a permis d'affirmer qu'il est plus réaliste et plus intéressant de prévoir la défaillance par un modèle de régression Cox-PLS, plutôt que par le modèle de Cox classique. 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 Cox-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 des données.

Bibliographie

- [1] Bardos, M. (1998), Detecting the risk of company failure at the Banque of France, *Journal of Banking and Finance*, 22, 1405–1419.
- [2] Bastien, P., Esposito, V., et Tenenhaus, M. (2005), PLS generalised linear regression, *Comput. Stat. Data Anal*, 48–1, 17–46.
- [3] Nguyen, D., Rocke, D. (2004), On partial least squares dimension reduction for microarraybased classification: a simulation study, *Comput. Stat. Data Anal*, 46, 407–425.
- [4] Fischer, N. (2004), *Fusion statistique de fichiers de données*, thèse de doctorat, CNAM.

Annexe 1 : 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 ajoutée	Taux de valeur ajoutée	R29
	Part des salariés	R30
	Part de l'état	R31
	Part des prêteurs	R32
	Part de l'autofinancement	R33