Revue Internationale des Sciences de Gestion
ISSN: 2665-7473
Numéro 6 / Volume 3 : numéro 1

La modélisation de la probabilité de défaut par la régression logistique et son impact sur les résultats bancaires.

Modelling the probability of default using logistic regression and its impact on banking results.

EL HADDAD SALIM
Etudiant chercheur en science de gestion
Faculté des sciences juridiques, économiques et sociales-Agdal
Université Mohammad V, Rabat.
Laboratoire des sciences de gestion
Salimelhaddad97@gmail.com

HABACHI MOHAMED
Docteur en sciences de gestion
Faculté des sciences juridiques, économiques et sociales-Agdal
Université Mohammad V, Rabat.
Laboratoire des sciences de gestion
Habachimed1@live.fr

Date de soumission : 30/11/2019.
Date d'acceptation : 07/01/2020.
Pour citer cet article :
EL HADDAD S. & HABACHI M. (2019) « La modélisation de la probabilité de défaut par la régression logistique et son impact sur les résultats bancaires », Revue Internationale des Sciences de Gestion « Numéro 6 / Volume 3 : numéro 1 » pp : 343 - 365

Digital Object Identifier : https://doi.org/10.5281/zenodo.3632233

Résumé

Les techniques de modélisation de la probabilité de défaut sont multiples. De ce fait, dans cet article nous allons utiliser la régression logistique pour déterminer la relation entre le défaut et les caractéristiques des entreprises et la construction des outils de notation interne.

La réglementation bancaire impose aux banques la constitution d'une provision pour risques et charges pour couvrir la perte de crédit attendue, et ce, en conformité avec les dispositions de la norme IFRS 9. De ce fait, les banques doivent utiliser leurs modèles internes que ce soit dans le cadre de l'approche IRB fondation ou avancée.

Notre objectif est de proposer une démarche de construction des outils de notation basés sur la régression logistique et d'évaluer la provision pour risques et charges destinée à couvrir la perte attendue. De ce fait, nous allons modéliser la défaillance par la régression logistique, et nous allons utiliser l'approche IRB fondation pour déterminer la perte attendue.

Mots clés: Probabilité de défaut ; régression logistique ; courbe de ROC ; perte attendue ; approche IRB fondation.

Abstract

The techniques for modelling the probability of default are multiple. Therefore, in this paper we will use logistic regression to determine the relationship between default and firm characteristics and the construction of internal rating tools.

Banking regulations impose on banks the constitution of a provision for risks and charges to cover the expected credit loss, in accordance with the he accounting standard IFRS 9. As a result, banks must use their internal models, whether under the IRB fundation approach or the advanced IRB approach.

Our objective is to propose an approach for the construction of rating tools based on the logistic regression and to evaluate the provision for risks and charges intended to cover the expected loss. Therefore, we will model the failure by logistic regression, and we will use the IRB foundation approach to calculate the expected loss (EL).

Keywords: Probability of default; Logistic regression; ROC curve; Expected loss; Approach IRB foundation.

Introduction

La réglementation bancaire définit plusieurs approches de calcul des fonds propres exigibles pour couvrir le risque de crédit. Ces approches sont composées de l'approche standard basée sur l'application d'un coefficient de pondération sur l'exposition en crédit et de l'approche basée sur la notation interne (Internal Ratings-Basesd- IRB).

Le calcul de la perte attendue est fondé sur la probabilité de défaut de la contrepartie de crédit, la perte au moment de la défaillance et l'exposition en cas de défaut. En effet, diverses recherches ont était menées pour définir une approche de prévision de la probabilité de défaillance. Ces études concernent l'utilisation de l'analyse discriminante linéaire, des techniques d'intelligences, des réseaux bayésiens et des modèles probabilistes.

Pour la régression logistique, diverses études ont était menées pour prédire la défaillance des entreprises notamment ceux menées par (Hosmer & Lemeshow, 1980), (Ohlson, 1980), (Hunter & Isachenkova, 2002), (Hensher, et al., 2007), (Hosmer, et al., 2013), (Benbachir & Habachi, 2018) et (Habachi, 2019).

L'explication de la défaillance est faite par des variables propres à chaque entreprise. Ces variables peuvent être des variables quantitatives, qualitatives ou macroéconomiques (Habachi, 2019) et (Habachi & Benbachir, 2018). En effet, les études divergent sur le choix des variables du fait que la revue de la littérature fait ressortir que le recours aux variables quantitatives était le premier choix pour les chercheurs comme (Altman, 1968), (Shirata, 1998), (Bardos, 1998) et (Taffler, 1982) alors que les variables qualitatives ont été étudiées par (Courdec & Renault, 2005), (Grunert, et al., 2005), (Yildirak, 2013) et (Habachi & Benbachir, 2019). D'autres recherches ont étudié l'impact des facteurs macroéconomiques sur le défaut des entreprises comme (Figlewski, et al., 2012).

La réglementation bancaire a subit d'énormes changements suite à la crise financière de 2007 notamment ceux concernant l'utilisation des modèles internes de risque de crédit, et ce, en introduisant la comptabilité en temps utile des pertes attendues sur crédit et le déploiement de la norme $IFRS\,9$. En effet, la norme prévoit la détermination de la probabilité de défaut annuelle (PD_a) qui est considérée constante et la probabilité de défaut sur la durée de vie de la créance ($PD_{lifetime}$) dite « probability of default lifetime ».

La perte sur crédit attendue (perte attendue EL) d'un portefeuille de n contreparties s'écrit :

$$EL = \sum_{i=1}^{n} EL_i = \sum_{i=1}^{n} PD_i * LGD_i * EAD_i$$

Avec EL_i est la perte attendue individuelle sur la contrepartie (i).

Pour la probabilité lifetime, les banques doivent construire un modèle d'évolution de la probabilité de défaut sur la durée de crédit du fait que la probabilité de défaut est considérée comme une variable aléatoire. Dans ce cas, le recours à la modélisation par la chaîne de Markov est suggéré par plusieurs recherches telles que (Cohen & Edwards, 2017), (Xu, 2016), (Kristóf & Virag, 2017) et (Vaněk & Hampel, 2017).

La problématique traitée dans cet article consiste à déterminer une démarche de construction d'un modèle de notation en se basant sur la régression logistique et à évaluer la probabilité de défaut annuelle par classe de notation (PD_C) ainsi que la perte attendue qui doit faire l'objet de provision dans le cadre de la norme Ifrs 9, et ce, en utilisant l'approche *IRB* fondation.

Compte tenu de ce qui précède, la première section est consacrée à la quantification du risque de crédit, la deuxième section est réservée à la modélisation de la probabilité de défaut, la troisième section est consacrée à la détermination de la grille de notation, la quatrième section est consacrée au calcul de la perte attendue et la cinquième section à l'étude empirique.

1. La quantification du risque de crédit.

1.1. Définition

Le risque de crédit du portefeuille bancaire est défini comme étant le risque qu'une contrepartie fasse défaut pour un horizon d'un an. La notion de défaut signifiée que la contrepartie est incapable d'honorer ses engagements à l'égard de l'établissement de crédits.

Une relation est considérée en défaut si ses encours sont reclassés en créance en souffrance[1] notamment dans les cas suivants :

(1) Persistance d'impayés pour une durée dépassant 90.
(2) Enregistrement de trois impayés.
(3) Baisse importante d'activité (Chiffre d'affaire).
(4) Reclassement de l'affaire en créances contentieuses.
(5) Dépassement chronique de l'engagement de financement (plafond autorisé).
(6) Engagements échus (autorisations échues) non renouvelés pour une période dépassant 90 jours.
(7) Gel de compte (compte sans mouvement pour une période dépassant 180 jours).

1.1.1. Les caractéristiques du risque de crédit :

Le risque de crédit est caractérisé par quatre composantes qui sont (Habachi, 2019):

(1) La probabilité de défaut (PD) de la contrepartie qui représente la probabilité qu'elle fasse défaut dans un horizon d'un an ;

[1] Les créances en souffrance sont les créances qui génèrent des impayés dépassant 90jours

(2) La perte en cas de défaut (LGD) qui est égale à la part non récupérée de l'encours de crédit octroyé. Il s'agit du montant susceptible d'être perdu en cas de défaut ;

(3) L'exposition en cas de défaut (EAD) qui représente l'encours de crédit auquel la banque sera exposée en cas de défaut ;

(4) La maturité (M) qui représente l'échéance effective du crédit.

1.1.2. *Les pertes attendues (EL):*

La perte attendue est définie comme le résultat de la multiplication de la probabilité de défaut (PD), la perte en cas de défaut (LGD) et l'exposition en cas de défaut (EAD):

$$Montant\ de\ la\ perte\ attendue\ (EL) = PD \times\ LGD \times EAD \quad (1)$$

1.2. L'approche notation interne (IRB-Internal Rating Based)

L'approche IRB est basée sur la détermination des quatre composantes citées en haut, qui sont : la PD , l'EAD, la LGD et la maturité M ainsi que les deux types de pertes qui sont : la perte inattendue (UL) et la perte attendue (EL).

Si la perte attendue (EL) est définie par la formule (1), la perte inattendue (UL) est égale à l'exposition en cas de défaut (EAD) pondérée par un coefficient α. Le coefficient α est une fonction de la probabilité de défaut (PD), de la perte en cas de défaut LGD et de la maturité (M) dont les formules sont définies par le document publié par le comité de Bâle en 2006, intitulé « International Convergence of Capital Measurement and Capital Standards» au niveau des paragraphes 272-273. De ce fait :

$$UL = \alpha \times EAD \quad (2)$$

Par ailleurs, la réglementation bancaire définit deux approches de notation interne IRB qui sont : l'approche IRB fondation selon laquelle l'établissement doit estimer uniquement la probabilité de défaut (PD), alors que, la perte en cas de défaut (LGD), l'exposition en cas de défaut (EAD) et la maturité (M) sont fournies par le régulateur et l'approche IRB avancée selon laquelle l'établissement doit estimer tous les paramètres (PD, LGD, EAD et M) par ses propres modèles internes.

2. La modélisation de la probabilité de défaut

Dans le cadre de l'approche IRB fondation, le calcul de la probabilité de défaut (PD) est la composante la plus importante. En effet, il s'agit de construire le modèle de notation qui permet de déterminer le profil de risque de chaque contrepartie sur la base de ses données qualitatives et quantitatives. Les entreprises sont réparties en classe de notation. Chaque

classe doit contenir des entreprises ayant des profils similaires de risque. La probabilité de défaut représente le pourcentage des entreprises en défaut dans chaque classe.

De ce fait, nous allons présenter ci-après la démarche de conception de l'outil de notation basé sur la régression logistique:

2.1. Définition des variables

- **Les variables quantitatives**

Soit (X_j), $j = 1,..,9$, les variables quantitatives.

Le choix des variables (X_j), $j = 1,..,9$ est dicté par la disponibilité des données, de ce fait, le tableau 1 résume les variables utilisées :

Table 1: la liste des variables quantitatives

Les variables quantitatives $(X_j, 1 \leq j \leq 9)$
X_1 = Chiffre d'Affaires (CA) X_2 = Nombre d'employés X_3 = Age de l'entreprise $X_4 \frac{Résultat\ net}{Fonds\ propres}$ $X_5 = \frac{Frais\ financiers}{Chiffre\ d'Affaires(CA)}$ $X_6 = \frac{DMLT(Dettes\ Moyen\ et\ Long\ Terme)}{Capitaux\ propres}$ $X_7 = \frac{Fond\ de\ Roulement(FDR)}{Actif\ circulant}$ $X_8 = \frac{resultat\ net}{fonds\ propres}$ $X_9 = \frac{Chiffre\ d'Affaires(CA)}{Total\ actif}$

Source : Auteurs

- **Les variables qualitatives :**

Les variables qualitatives $(q_m, 1 \leq m = 1, \ldots,3)$ sont définies dans le tableau 2 comme suit :

Table 2: la liste des variables qualitatives

Variables qualitatives (q_m)
q_1 : Taux de défaut secteur q_1 : Expérience PDG. q_3 : Forme de société.

Source : Auteurs

2.2. Modélisation de la probabilité.

2.2.1. Analyse univariée et détermination des variables explicatives

2.2.1.1.Analyse univariée

L'analyse univariée permet de déterminer la relation entre le défaut d'une entreprise et les variables quantitatives et qualitatives qui caractérisent cette entreprise (i). Le défaut est modélisé par une variable Y définie dans {0,1} comme suit :

$$Y = \begin{cases} 1 \; si \; l'entreprise \; est \; saine \\ 0 \; si \; l'entreprise \; est \; en \; défaut \end{cases} \quad (3)$$

La relation entre la variable Y qui représente le défaut et les variables explicatives X_j et q_m qui caractérisent l'entreprise est déterminée par le modèle de régression logistique (Benbachir & Habachi, 2018) et (Habachi, 2019).

En effet, l'expression (3) montre que Y est une variable de Bernoulli de paramètre p avec $P(Y = 1) = p$ et $P(Y = 0) = 1 - p$. De ce fait, la loi de probabilité de Y s'écrit :

$$P(Y = y_i) = p^{y_i}(1 - p)^{1-y_i} \quad (4)$$

La régressions logistique entre la variable Y et les caractéristiques de l'entreprise (i) consiste à définir une relation entre Y et une probabilité logistique (p) définie à partir des caractéristiques (X_j) et (q_m). De ce fait, il faut effectuer une transformation logistique (Logit) pour lier la variable dépendante Y à chaque variable parmi les variables $(X_j), \mathrm{j} = 1, \dots, 9$ et $(q_m), m = 1, \dots, 3$.

Pour chaque variable X_j ou q_m, la relation entre le défaut et les variables étudiées est définie par la probabilité :

$$p_j = P(Y = 1/X_j) \text{ ou } p_m = P(Y = 1/q_m)$$

En effet, la modélisation de Y conditionnellement à la variable X_j ou q_m est définie par le modèle :

$$Y = p + \varepsilon \quad (5)$$

$$avec \; p = P(Y = 1/X_j) \; (ou \;\; p = P(Y = 1/q_m))$$

2.2.1.2.La modélisation de la probabilité p:

Nous allons présenter l'analyse univariée des variables X_j et nous procédons de la mêmes façon pour les variables q_m.

De ce fait, soit x_{ij} la valeur observée de la variable X_j pour l'entreprise (i), et p_i la probabilité que l'entreprise (i) soit saine. En effet, p_i est définie par $P(Y_i = 1/X_j = x_{ij})$ avec :

$$Y_i = \begin{cases} 1 \; si \; l'entreprise(i) \; est \; saine \\ 0 \; si \; l'entreprise(i) \; est \; en \; défaut \end{cases}$$

De ce fait, la formule (5) s'écrit pour l'entreprise (i) comme suit :

$$y_i = p_i + \varepsilon_i$$

Sachant que le ratio $\frac{p_i}{1-p_i}$ est appelé l'Oddos de p_i, donc :

$$\text{Oddos}(p_i) = \frac{p_i}{1-p_i} = \frac{P(Y_i = 1/X_j = x_{ij})}{1 - P(Y_i = 1/X_j = x_{ij})} \quad (6)$$

En conséquence, la régression logistique établit une relation de régression linéaire entre $\ln\left(\frac{p_i}{1-p_i}\right)$ et la variable x_{ij} comme suit :

$$Logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 \times x_{ij} \quad \text{d'où}$$

$$p_i = \frac{e^{\beta_0+\beta_1\times x_{ij}}}{1 + e^{\beta_0+\beta_1\times x_{ij}}} \quad (7)$$

En conséquence, la formule (5) et la probabilité p y afférente se présentent ainsi :

$$\begin{cases} Y = p + \varepsilon \\ p = \frac{e^{\beta_0+\beta_1\times X_j}}{1+e^{\beta_0+\beta_1\times X_j}} \end{cases}$$

2.2.1.3.Estimation des paramètres :

La variable Y est définie sur $\{0,1\}$. Soit Y_i la variable qui modélise le défaut de l'entreprise (i). La variable $(Y_i/X_j = x_{ji})$ est une variable Bernoulli de paramètre p_i avec :

$$P(Y_i/X_j = x_{ij}) = P(Y_i = y_i/X_j = x_{ij}) = {p_i}^{y_i} \times (1-p_i)^{(1-y_i)} \quad (8)$$

avec $y_i = 0\ ou\ 1$

❖ ***La fonction de vraisemblance :***

La fonction de vraisemblance de n entreprises est définie par :

$$L = \prod_{i=1}^{n} P(Y_i/X_j = x_{ij}) = \prod_{i=1}^{n} (1-p_{0i})^{(1-y_i)} \times {p_{0i}}^{y_i} \quad (9)$$

$$L = \prod_{i=1}^{n} \left(\frac{e^{\beta_0+\beta_1\times x_{ij}}}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right)^{y_i} \times \left(1 - \frac{e^{\beta_0+\beta_1\times x_{ij}}}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right)^{(1-y_i)}$$

❖ **La fonction de log-vraisemblance :**

La fonction log-vraisemblance est définie par :

$$LL = \sum_{i=1}^{n} y_i \times ln\left(\frac{e^{\beta_0+\beta_1\times x_{ij}}}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right) + \sum_{i=1}^{n} (1-y_i) \times ln\left(1 - \frac{e^{\beta_0+\beta_1\times x_{ij}}}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right) \quad (10)$$

$$LL = \sum_{i=1}^{n} ln\left(\frac{1}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right) + y_i \left(ln\left(\frac{e^{\beta_0+\beta_1\times x_{ij}}}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right) - ln\left(\frac{1}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right)\right)$$

$$LL = \sum_{i=1}^{n} \ln\left(\frac{1}{1 + e^{\beta_0+\beta_1\times x_{ij}}}\right) + y_i \times (\beta_0 + \beta_1 \times x_{ij}) \quad (11)$$

Soit $\beta = (\beta_0, \beta_1)$ donc l'estimation de β consiste à déterminer $\hat{\beta}$ qui maximise la fonction de vraisemblance et en conséquence, maximise la fonction de log-vraisemblance.

Pour maximiser la fonction LL il faut résoudre les équations déterminées par les conditions suivantes :

$$\frac{\partial LL}{\partial \beta} = 0 \; et \; \frac{\partial^2 LL}{\partial^2 \beta} < 0 \tag{12}$$

La résolution du problème est faite par l'algorithme de Newton-Raphson.

2.2.1.4.Test de validité de modèle

Le test de Wald pour la validation du modèle se présente comme suit :

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \; \beta_1 \neq 0 \end{cases} \tag{13}$$

La statistique S_1 de test est définie comme suit :

$$S_1 = \frac{\hat{\beta}_1^2}{\hat{\sigma}_{\hat{\beta}_1}^2} \tag{14}$$

Sous l'hypothèse H_0, S_1 est un χ^2 de degré 1. De ce fait, H_0 est rejetée si $S_1 > \chi_1^2$ (0,05)

2.2.1.5.Pouvoir discriminant :

La capacité du modèle à prédire des situations futures détermine son pouvoir discriminant. De ce fait, nous allons utiliser la courbe de ROC pour déterminer le pouvoir discriminant de chaque variable $X_j \; et \; q_m$.

La courbe de ROC est déterminée à partir du tableau 3 appelé tableau de classification qui se présente comme suit :

Table 3: Tableau de classification

	Saine ($Y = 1$)	Défaut ($Y = 0$)
Saine ($Y = 1$)	Vrais Saine (VS)	Faux Saine (FS)
Défaut ($Y = 0$)	Faux Défaut (FD)	Vrais Défaut (VD)

Source : Auteurs

(1) La sensibilité (sensivity (SV)) désigne la proportion des entreprises saines correctement classées :

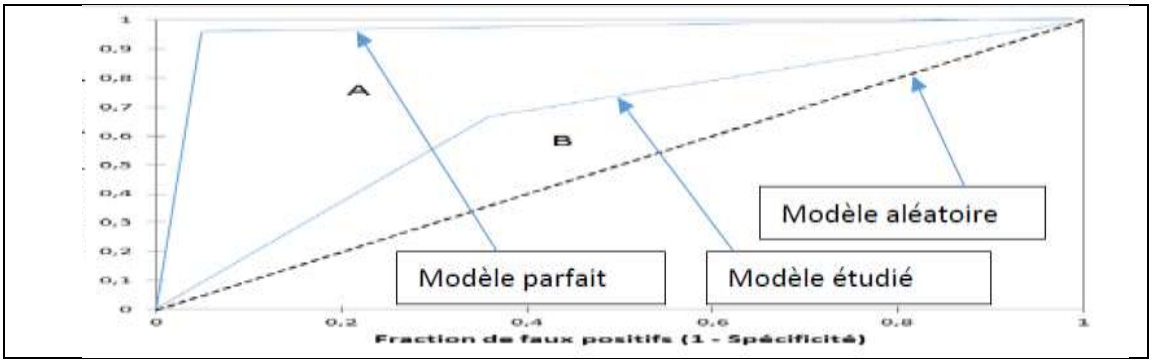
$$SV = \frac{VS}{VS+FS}$$

(2) la spécificité (specificity (SP)) désigne la proportion des entreprises en défaut correctement classées :

$$SP = \frac{VD}{VD + FD}$$

Si on fait varier la 'probabilité seuil' à partir de laquelle une entreprise est considérée saine, la sensibilité et la spécificité varient respectivement. Les points définis par le couple (1-SP, SV) forment la courbe ROC.

Figure 1: courbe de ROC



Source : auteurs

L'aire sous la courbe de ROC est la mesure de la surface en dessous de la courbe de ROC notée AUC. En effet, ses valeurs appartiennent à l'intervalle [0,1].

2.2.1.6.Détermination des variables explicatives

Le choix des variables explicatives à retenir pour le modèle de classification est basé sur l'analyse de régression logistique univariée entre le défaut et chacune des variables listées dans les tableaux 1 et 2. En effet, elles sont déterminées par le test de signification de Wald et le pouvoir discriminant représenté par l'aire sous la courbe de ROC qui doit être supérieure à 0,60.

2.2.2. Analyse multivariée et la fonction score

2.2.2.1.Analyse multivariée

L'analyse discriminante multivariée permet d'étudier le lien entre la variable à expliquer Y et les variables explicatives $V_m, m = 1, \dots, l$ retenues ; parmi, les variables $X_j, j = 1, \dots, 9$ et $q_k, k = 1, \dots, 3$. Le modèle peut être exprimé par la relation suivante :

$$Y = f_1(V_1, V_2, \dots, V_l) + \epsilon \qquad (15)$$

Avec f_1 est la probabilité logistique.

Soit, n la taille de l'échantillon des entreprises utilisées pour la modélisation, et soit $V_{im}, 1 \leq i \leq$ n les vecteurs définies par les variables explicatives retenues notées $V_m, 1 \leq m \leq l$. Notons par $V = (v_{ij}), 1 \leq i \leq n, 1 \leq m \leq l$, la matrice dont les colonnes sont $V_m, 1 \leq m \leq l$.

L'objectif de cette étude est la modélisation de la variable Y par le modèle suivant :

$$Y = p_0 + \varepsilon$$

Avec $p_0 = P(Y = 1/V_1, V_2, \dots, V_l)$

Soit $(v_{i1}, \cdots, v_{il})$ les données de l'entreprise (i) qui représentent les réalisations des variables $(V_1, V_2, \dots, V_l)$ par l'entreprise (i) et y_i la réalisation de Y_i pour cette entreprise alors y_i s'écrit

$$y_i = p_{0i} + \varepsilon_i \quad (16)$$

$$avec\ p_{0i} = P(Y_i = 1/v_{i1}, \cdots, v_{il})$$

Pour déterminer l'expression de p_{0i} par le modèle de la régression logistique, on procède comme suit:

Soit,

$$\text{Oddos}(p_{0i}) = \frac{p_{0i}}{1 - p_{0i}} = \frac{P(Y_i = 1/v_{i1}, \cdots, v_{il})}{1 - P(Y_i = 1/v_{i1}, \cdots, v_{il})}$$

Donc, nous pouvons écrire le modèle sous la forme logarithmique comme suit:

$$Logit(p_{0i}) = \ln\left(\frac{p_{0i}}{1-p_{0i}}\right) = \beta_0 + \beta_1 \times v_{i1} + \beta_2 \times v_{i2} + \cdots + \beta_l \times x_{il} = \beta {v_i}'$$

Avec $\beta = (\beta_0, .., \beta_{12})$ et ${v_i}' = transposé(1, v_{i1}, \cdots, v_{il})$ donc

$$p_{0i} = \frac{e^{\beta {v_i}'}}{1 + e^{\beta {v_i}'}} \quad (17)$$

Soit $V' = transposé(1, v_1, \cdots, v_l)$ alors p_0 et Y peuvent s'écrire :

$$\begin{cases} p_0 = \frac{e^{\beta V'}}{1+e^{\beta V'}} \\ Y = \frac{e^{\beta V'}}{1+e^{\beta V'}} + \varepsilon \end{cases}$$

2.2.2.2.Estimation des paramètres :

La variable aléatoire Y est définie dans $\{0,1\}$. Soit y_i la valeur Y_i observée pour l'entreprise (i) on modélise la variable conditionnelle $(Y_i/v_{i1}, \cdots, v_{il})$ comme suit :

$$P(Y_i = y_i/v_{i1}, \cdots, v_{il}) = {p_{0i}}^{y_i} \times (1 - p_{0i})^{(1-y_i)} \quad (18)$$

❖ **La fonction de vraisemblance :**

La fonction de vraisemblance de n entreprises est définie par :

$$L = \prod_{i=1}^{n} P(Y_i = y_i/v_{i1}, \cdots, v_{il}) = \prod_{i=1}^{n} (1 - p_{0i})^{(1-y_i)} \times {p_{0i}}^{y_i} \quad (19)$$

$$L = \prod_{i=1}^{n} \left(\frac{e^{\beta {v_i}'}}{1 + e^{\beta {v_i}'}}\right)^{y_i} \times \left(1 - \frac{e^{\beta {v_i}'}}{1 + e^{\beta {v_i}'}}\right)^{(1-y_i)} \quad (20)$$

❖ **La fonction de log-vraisemblance :**

La fonction de log-vraisemblance de n entreprises est définie par :

$$LL = \ln(L) = \sum_{i=1}^{n} y_i \times ln\left(\frac{e^{\beta v_i'}}{1+e^{\beta v_i'}}\right) + \sum_{i=1}^{n}(1-y_i) \times ln\,(1-\frac{e^{\beta v_i'}}{1+e^{\beta v_i'}}) \quad (21)$$

$$LL = \sum_{i=1}^{n} \ln\left(\frac{1}{1+e^{\beta v_i'}}\right) + y_i \beta v_i' \quad (22)$$

L'estimation de β consiste à déterminer $\hat{\beta}$ qui maximise la fonction de vraisemblance et en conséquence maximise la fonction log-vraisemblance.

Pour maximiser la fonction LL, il suffit de résoudre le systéme des équations déterminées par les conditions suivantes :

$$\frac{\partial LL}{\partial \beta} = 0 \; et \; \frac{\partial^2 LL}{\partial^2 \beta} < 0$$

2.2.2.3.Test de validité du modèle multivarié :

Pour tester la validité du modèle multivarié, nous allons procéder comme suit :

❖ **Test de signification**

Soit le modèle logistique définie par :

$$y_i = p_i + \varepsilon_i$$

Avec $p_i = \frac{e^{\beta V'}}{1+e^{\beta V'}}$, $\beta = (\beta_0, .., \beta_l)$ et $V' = tr(1, v_1, .., v_l)$

Les paramètres de la régression doivent être significatifs. De ce fait, il faut tester la signification globale et individuelle. L'hypothèse H_0 pour les deux tests se présente comme suit :

(1) Pour le test de signification individuelle l'hypothèse à tester est :

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases}$$

(2) Pour le test de signification globale l'hypothèse à tester est :

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \cdots = \beta_l = 0 \\ H_1 : \exists k \in \{1, \ldots, l\} : \beta_k \neq 0 (au\ mois\ un\ \beta_k \neq 0\ ) \end{cases}$$

Pour tester la signification des variables explicatives $V_m, 1 \leq m \leq l$, trois testes sont utilisés :

✓ **Test de Wald :**

Le test de Wald pour la signification individuelle est basé sur la statistique suivante :

$$W_i = \frac{\hat{\beta}_i^2}{\hat{\sigma}(\hat{\beta}_i)^2}$$

La statistique W_i sous l'hypothèse H_0, est un χ^2 de degré 1. De ce fait, H_0 est rejetée si $W_i > \chi_1^2(0{,}05)$

✓ **Test de rapport de vraisemblance :**

Le test de signification globale est basé sur le ratio de vraisemblance qui représente le rapport entre le modèle qui ne contient pas les variables explicatives et le modèle qui contient les variables explicatives. En effet, le test peut être formulé comme suit :

$$\begin{cases} H_0 : logit\left(P(Y=1)\right) = \beta_0 \\ H_1 : logit\left(P(Y=1)\right) = \beta Z' \end{cases} \quad (23)$$

La statistique de test RV est définie par :

$$RV = -2\left[\frac{log - vraisemblance\ au\ maximum\ du\ modéle\ M_1}{log - vraisemblance\ au\ maximum\ du\ modéle\ M_2}\right]$$

Avec :

- M_1 : le modèle définit par $logit\left(P(Y=1)\right) = \beta_0$
- La fonction log-vraisemblance au maximum du modèle M_1 :

$$M_1 : LL = \sum_{i=1}^{n} \ln\left(\frac{1}{1+e^{\beta_0}}\right) + y_i\beta_0$$

- La fonction log-vraisemblance au maximum du modèle M_2 est définie par :

$$M_2 : LL = \sum_{i=1}^{n} \ln\left(\frac{1}{1+e^{\beta Z}}\right) + y_i\beta Z'$$

Sous l'hypothèse H_0 la statistique RV est un χ^2 de degré l. De ce fait, on rejette H_0 si $\text{RV} > \chi_l^2(0{,}05)$. En conséquence, on déduit que le modèle M_2 est meilleur que le modèle M_1.

❖ **Test d'adéquation, *Test de Hosmer-Lemeshow***

Le test d'adéquation consiste à évaluer l'ajustement entre les valeurs prédites et celles observées. En effet, les données doivent être classées par ordre croissant des probabilités déterminées par le modèle ; ensuite, elles doivent être réparties au plus en 10 groupes.

Les hypothèses du test sont :

$$\begin{cases} H_0 : le\ modèle\ et\ correcte \\ H_1 : le\ modèle\ n'est\ pas\ correcte \end{cases}$$

La statistique de test SHL est définie par :

$$SHL = \sum_{g=1}^{10} \frac{(O_g - E_g)^2}{E_g} \quad (24)$$

Avec $O_g = \sum_{i\epsilon g} y_i$ désigne le nombre d'évènement observé $et\ E_g = \sum_{i\epsilon g} \hat{p}_i$ le nombre d'évènement théorique. La statistique SHL est un χ^2 de degré 8. De ce fait, l'hypothèse H_0 est rejetée si $\text{SHL} > \chi_8^2(0{,}05)$.

2.2.2.4.Performance de modèle

Pour déterminer la fiabilité des estimations fournies par le modèle et comparer les modèles qui s'ajustent avec les données observées, on doit évaluer la performance du modèle. En effet, la mesurer de la performance du modèle est basée sur le calcul de l'aire sous la courbe de ROC (AUC) présentée en haut.

Une plus grande AUC signifiée une meilleure capacité de prédiction du modèle. En conséquence, le modèle sera retenu si l'AUC est supérieure ou égale à 0,7.

L'utilisation de l'aire sous la courbe de ROC (AUC) et l'Accuracy ratio (AR) permet de déterminer le pouvoir discriminant du modèle.

- **Aire sous la courbe de ROC (AUC).**

L'aire sous la courbe de ROC (AUC) a été étudiée par (Bewick, et al., 2004) qui a montré que cette mesure fournit une évaluation globale de l'ajustement du modèle. En effet, L'AUC varie de 0,5 (absence de capacité prédictive) à 1 (capacité prédictive parfaite).

- **Accuracy ratio (AR).**

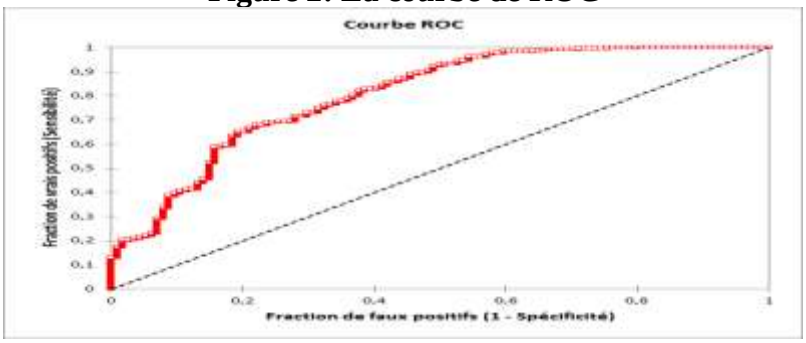
L'accuracy ratio est défini par la relation :

$$AR = 2AUC - 1 \qquad (25)$$

L'AR prend des valeurs entre 0 et 1.

La pouvoir discriminant du modèle est satisfaisant si l'AUC dépasse 70% et l'AR est supérieur à 40%. La courbe de ROC est présentée par la figure 2 :

Figure 2: La courbe de ROC



Source : Auteurs

2.2.2.5. La fonction score :

Pour déterminer les classes de notation, nous allons utiliser la fonction score définie comme suit :

$$S = \beta V' = \beta_0 + \beta_1 V_1 + \beta_2 \times V_2 + \cdots + \beta_l \times V_l \qquad (26)$$

3. La grille de notation.

Une entreprise (i) est considérée saine si $p_i \geq \frac{1}{2}$, en effet :

$$p_i = \frac{e^S}{1+e^S} \geq \frac{1}{2} \leftrightarrow e^S \geq \frac{1}{2}(1+e^S)$$
$$\leftrightarrow e^S \geq 1 \leftrightarrow S \geq 0 \quad (27)$$

De ce fait, l'entreprise (i) est considérée saine si son score est positif.

3.1. La démarche de construction de la grille de notation

(Habachi & Benbachir, 2018, 2019) ont défini la démarche de construction d'un outil de notation. En effet, cette démarche de classification des entreprises saines est basée sur la fonction score S. La grille de notation est structurée en 8 classes qui varient entre A et H.

Chaque entreprise (i) est logée dans une seule classe de notation en fonction de son score. La grille de notation retenue est définie par le tableau 4 suivant :

Tableau 4 : La grille de notation (Habachi & Benbachir, 2018, 2019)

Classe	A	B	C	D	E	F	G	H
Score notation	[90-100]	[80-90[	[70-80[	[60-70[	[50-60[	[40-50[	[30-40[	<30

Source : Auteurs

Le score de notation SN_i de l'entreprise (i) est défini par :

$$SN_i = \frac{S_i}{Max(S_i)} \times 100 \quad (28)$$

Avec S_i est le score de l'entreprise (i) calculé par la formule (26).

3.2. La probabilité de défaut par classe

Pour déterminer la probabilité de défaut dans la classe (PD_C), il faut déterminer la PD de l'entreprise (i) sachant que cette entreprise est dans la classe C , de ce fait :

$$PD_C = P(Y = 0/i \in classe\ C)$$

Le calcul empirique de cette probabilité est donné par la formule suivante :

$$PD_C = \frac{nombre\ d'entreprise\ en\ défaut\ dans\ la\ class\ C}{Le\ nombre\ total\ des\ entreprise\ dans\ la\ classe\ C} \quad (29)$$

4. La perte attendue

Pour calculer la perte attendue (EL) définie par la formule (1) on utilise la PD déterminée par la formule (29) ainsi que la LGD et l' EAD définies par le législateur comme suit :

4.1. La perte en cas de défaut (LGD)

La perte en cas de défaut (LGD) doit être estimée par des modéles internes. Toutefois, elle est fixée à 45% par l'accord Bâle II pour l'approche IRB fondation.

4.2. L'exposition en cas de défaut (EAD)

Le facteur de conversion en crédit (FCC) permet de définir l'exposition en cas de défaut EAD.

De ce fait, l' EAD est donnée par la formule mathématique suivante :

$$EAD = V_B + FCC \times V_{HB} \quad (30)$$

Avec :

- V_B : l'encour enregistré au Bilan.
- V_{HB} : l'engagement de financement non utilisé, comptabilisé en Hors Bilan

Dans le cadre de l'approche IRB fondation, le facteur de conversion en crédit des expositions sur les entreprises(FCC) est fixé à 75%.

5. Résultats empiriques

5.1. Description des données

Pour l'étude empirique, nous avons utilisé une base de données composée de 375 PME. Une entreprise est classée PME par Bank Al Maghrib si son chiffre d'affaires est limité entre 10 et 175 millions de MAD.

Les entreprises du portefeuille utilisé sont réparties selon leur statut (saine ou en défaut) dans le tableau 5 suivant :

Tableau 5: Répartition du PTF

	Nombre	%
PTF sain	344	91,73%
PTF en défaut	31	8,27%
Total	375	100%

Source : Auteurs

5.2. Choix des variables

5.2.1. Analyse univariée

Les variables explicatives sont déterminées par l'analyse univariée des variables $X_j, j = 1, \dots, 9$ et $q_k, k = 1, \dots, 3$. Le test de Wald et le pouvoir discriminant de chaque variable sont illustrés dans le tableau 6 suivant :

Tableau 6: Analyse univariée.

intitulé	DDL	Khi² (Wald)	p=Pr(x > Wald)	AUC	AR
X_1 = Chiffre d'Affaires (CA)	1	0,055	0,815	54,1%	8,2%
X_2 = Nombre d'employés	1	5,968	0,015	64,4%	28,8%
X_3 = Age de l'entreprise	1	2,523	0,112	51,1%	2,2%
$X_4 = \frac{Résultat\ net}{Fonds\ propres}$	1	5,038	0,025	61,1%	22,2%
$X_5 = \frac{Frais\ financiers}{Chiffre\ d'Affaires(CA)}$	1	6,075	0,014	63,6%	27,2%
$X_6 = \frac{Dettes\ nette}{fonds\ propres}$	1	0,014	0,906	52,6%	5,2%
$X_7 = \frac{Fond\ de\ Roulement(FDR)}{Actif\ circulant}$	1	0,407	0,524	57,6%	15,2%
$X_8 = \frac{resultat\ net}{fonds\ propres}$	1	4,176	0,041	60,3%	20,6%

$X_9 = \frac{Chiffre\ d'Affaires(CA)}{Actif\ immobilisé\ net+BFR}$	1	4,545	0,033	60,7%	21,4%
q_1 : Taux de défaut secteur	1	5,599	0,018	64,9%	29,8%
q_2 : Expérience PDG.	1	35,387	< 0,0001	77,6%	55,2%
q_3 : La forme de la société.	1	0,281	0,596	58,2%	16,4%

Source : Auteurs

L'analyse univariée montre que X_1, X_3, X_6, X_7 et q_3 ne sont pas significatives selon le test de Wald du fait que (p>0,05) et n'ont pas de pouvoir discriminant significatif du fait que l'aire sous la courbe de ROC est inférieure à 0,6. Toutefois, nous allons utiliser pour la modélisation, parmi ces variables non explicatives, celles qui permettent d'augmenter le pouvoir discriminant du modèle.

Soit, M_1 le modèle composé des variables significatives $(X_2, X_4, X_5, X_8, X_9, q_1, q_2)$, De ce fait, nous allons calculer le pouvoir discriminant du modèle M_1 en ajoutant les variables non significatives $(X_1, X_3, X_6, X_7, Q_3)$. Les résultats sont résumés dans le tableau 7 suivant :

Tableau 7 : Pouvoir discriminant des modèles en fonction des variables utilisées

Modèle	M_1	M_2 $(M_1 + X_1)$	M_3 $(M_1 + X_3)$	M_4 $(M_1 + X_6)$	M_5 $(M_1 + X_7)$	M_6 $(M_1 + q_3)$
AUC	78,5%	78,9%	78 ,5%	78,9%	78,4%	78,3%
Modèle	M_7 $(M_2 + X_3)$	M_8 $(M_3 + X_6)$				
AUC	79%	79,2%				

Source : Auteurs

Les pouvoirs discriminant des modèles montrent que les variables X_7 et q_3 ne participent pas à l'augmentation du pouvoir discriminant du modèle M_1. De ce fait, nous avons exclu ces deux variables de la construction du modèle que nous avons utilisé pour modéliser le défaut et déterminer la fonction score.

5.3. Analyse multivariée et fonction de score

5.3.1. Analyse multivariée

Soit $V_m, 1 \leq m \leq 10$ tel que $(V_1, V_2, V_3, \dots, V_{10}) = (X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9, q_1, q_2)$.

L'estimation des paramètres de la régression logistique entre le défaut et les variables retenues $V_m, 1 \leq m \leq 10$ ainsi que les tests de signification se présentent comme suit :

5.3.1.1. Estimation des paramètres et test de Wald.

Pour maximiser la fonction de vraisemblance, nous avons utilisé l'algorithme de Newton-Raphson. L'estimation des paramètres ainsi que le test de Wald sont présentés dans le tableau 8 suivant :

Tableau 8 : Estimation des paramètres du modèle et test de Wald

	β_i	Ecartype	Khi² de Wald	Pv(Pr > Khi²)
Constante	β_0=-2,416	1,469	2,703	0,100
V_1	β_1=-0,005	0,014	0,148	0,701
V_2	β_2=-0,005	0,006	0,708	0,400
V_3	β_3=-0,003	0,007	0,157	0,692
V_4	β_4=0,009	0,009	1,177	0,278
V_5	β_5=0,010	0,006	2,646	0,104
V_6	β_6=-0,003	0,007	0,166	0,684
V_7	β_7=0,001	0,009	0,013	0,909
V_8	β_8=-0,007	0,008	0,766	0,381
V_9	β_9=0,008	0,006	1,928	0,165
V_{10}	β_{10}=0,071	0,014	26,858	< 0,0001

Source : Auteurs

5.3.1.2. Test de vraisemblance

La valeur de la statistique RV définie précédemment est de 39,872. Elle est supérieure à 18,307 (χ^2_{10} au seuil de 0,05). En conséquence, l'hypothèse H_0 est rejetée et les variables retenues sont globalement significatives et permettent de déterminer un modèle meilleur.

5.3.1.3. Test de validation et de performance du modèle

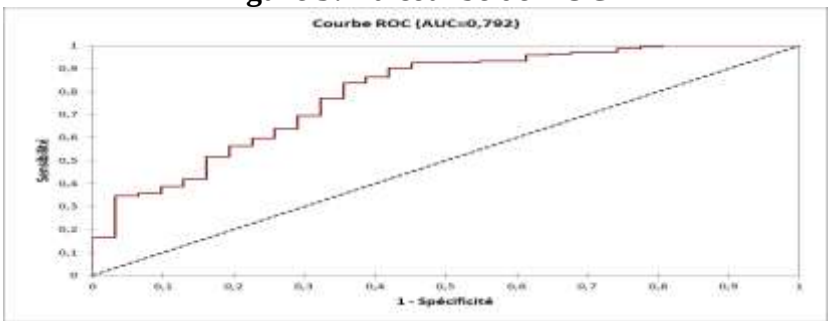
❖ **Test de validation**

La valeur de la statistique SHL valide le modèle et montre qu'il s'ajuste avec l'échantillon étudié. En effet, elle est égale à 6,173, soit, une pv de 0,628 qui est supérieure au seuil de décision de 0,05.

❖ **La performance du modèle**

La figure 3 représente la courbe de ROC qui montre que le modèle choisi permet d'avoir une classification satisfaisante des entreprises, soit, une valeur de l'AUC égale à 0,792.

Figure 3: La courbe de ROC



Source: Auteurs

5.3.2. La fonction score

L'estimation des paramètres permet de déterminer la fonction score. De ce fait, le score $S_1(i)$ de chaque entreprise (i) s'écrit comme suit :

$$S_1(i) = \beta_0 + \beta_1 V_{i1} + \beta_2 V_{i,2} + \beta_3 V_{i,3} + \beta_4 V_{i,4} + \beta_5 V_{i,5} + \beta_6 V_{i,6} + \beta_7 V_{i,7} + \beta_8 V_{i,8} + \beta_9 V_{i,9} + \beta_{10} V_{i,10}$$

En conséquence, la variable Y_i peut être estimée par la probabilité :

$$p_{1i} = \frac{e^{S_1(i)}}{1 + e^{S_1(i)}}$$

Avec :

$$S_1(i) = -2{,}416 - 0{,}005 * V_{i,1} - 0{,}005 * V_{i,2} - 0{,}003 * V_{i,3} + 0{,}009 * V_{i,4} + 0{,}010 * V_{i,5} - 0{,}003 * V_{i,6} - 0{,}001 * V_{i,7} - 0{,}007 * V_{i,8} + 0{,}008 * V_{i,9} + 0{,}071 * V_{i,10}$$

De ce fait, la fonction score est exprimée comme suit:

$$S_1 = -2{,}416 - 0{,}005 * V_1 - 0{,}005 * V_2 - 0{,}003 * V_3 + 0{,}009 * V_4 + 0{,}010 * V_5 - 0{,}003 * V_6 - 0{,}001 * V_7 - 0{,}007 * V_8 + 0{,}008 * V_9 + 0{,}071 * V_{10}$$

5.4. La grille de notation

La fonction score a permis de répartir les entreprises de l'échantillon par classe de notation. La structure du PTF par classe de notation est résumée dans le tableau 9 suivant :

Tableau 9: La grille de notation et la répartition de l'échantillon des entreprises

CLASSE	SCORE	NBRE D'ENTREPRISE	%ENTREPRISE
A	[90-100]	8	0,02
B	[80-90[	62	0,17
C	[70-80[	121	0,32
D	[60-70[	92	0,25
E	[50-60[	46	0,12
F	[40-50[	18	0,05
G	[30-40[	19	0,05
H	[0-30[	9	0,02

Source: Auteurs

5.5. La probabilité de défaut par classe

La répartition du portefeuille par classe de notation a permis de déterminer les entreprises saines et celles en défaut dans chaque classe. La probabilité de défaut *PD* qui est égale à la proportion des entreprises en défaut dans chaque classe est résumée dans le tableau 10 suivant :

Tableau 10: Probabilité de défaut par classe de notation

Classe de notation	Nbre entreprise saine	Nbre entreprise en défaut	PD
A	6	0	0,03%
B	62	1	1,6%
C	115	4	3,4%
D	84	6	6,7%

E	41	5	10,9%
F	17	3	15,0%
G	14	5	26,3%
H	5	7	58,3%
TOTAL	344	31	

Source: Auteurs

5.6. La perte attendue

L'encours des crédits utilisés par le portefeuille étudié (V_B) s'élève à 3.475,60 millions MAD pour un engagement de financement ($V_B + V_{HB}$) de 4.075 millions MAD. La répartition des encours et des engagements par classe de notation est résumée dans le tableau 11 comme suit :

Tableau 11: Répartition de V_B et ($V_B + V_{HB}$) par classe de notation (en millions MAD)

CLASSE DE NOTATION	VB_0	Structure en%	$(V_B + V_{HB})$	Structure en%
A	27,6	0,79%	40	0,98%
B	281,5	8,10%	322	7,90%
C	641,5	18,46%	765	18,77%
D	1182,4	34,02%	1350	33,13%
E	672,3	19,34%	873	21,42%
F	225,2	6,48%	247	6,06%
G	265,1	7,63%	286	7,02%
H	180	5,18%	192	4,71%
Total	3475,6	100,0%	4075	100,00%

Source: Auteurs

Le montant de la perte attendue par classe de notation, est déterminé par la formule (1) sachant que la perte en cas de défaut (LGD) est égale à 45% et le FCC est égal à 75%.

Le montant de la perte attendue totale (EL) et sa répartition par classe de notation sont résumés dans le tableau 12 comme suit :

Tableau 12: Répartition du montant de la perte attendue (en millions MAD)

NOTATION	VB_0	$(V_B + V_{HB})$	PD	LGD	EAD	EL
A	27,6	40	0,03%	45%	36,9	0,00
B	281,5	322	1,6%	45%	311,875	2,25
C	641,5	765	3,4%	45%	734,125	11,23
D	1182,4	1350	6,7%	45%	1308,1	39,44
E	672,3	873	10,9%	45%	822,825	40,36
F	225,2	247	15,0%	45%	241,55	16,30
G	265,1	286	26,3%	45%	280,775	33,23
H	180	192	58,3%	45%	189	49,58
Total	3475,6	4075			3925,15	192,40

Source: Auteurs

Le tableau précédent montre que l'impact de la perte attendue concernant le PTF étudié, sur le résultat s'élève à 192,40 millions MAD.

Conclusion

Le risque de crédit représente le volet le plus important dans la gestion des risques bancaires. En effet, les banques doivent observer en continue l'adéquation de leurs fonds propres avec les risques encourus pour couvrir la perte inattendue.

Par ailleurs, les banques doivent constituer une provision pour couvrir la perte attendue dans le cadre de la norme Ifrs 9 qui vient en déduction du résultat ce qui impact directement le niveau de fonds propres de base (Core one). De ce fait, son calcul doit être minutieux pour se conformer à la réglementation bancaire.

Le système de notation impact directement le calcul de la perte attendue du fait qu'il fournit les différentes composantes de calcul qui sont : la probabilité de défaut (PD), la perte en cas de défaut (LGD) et l'exposition en cas de défaut (EAD) par classe de notation et par contrepartie. De ce fait, le montant de la perte attendue dépend de façon directe de l'approche de modélisation utilisée.

Dans cet article, nous avons construit un outil de notation qui a un bon pouvoir discriminant en se basant sur la modélisation du défaut par la régression logistique, et ce, en utilisant les données quantitatives et qualitatives.

L'impact de la perte attendue s'élevé à 192,40 millions de MAD, soit, 5,53% de l'encours de crédit ce qui donne lieu à une provision additionnelle pour risques et charges déduite du résultat de l'exercice.

La norme IFRS 9 représente une nouvelle contrainte pour les banques du fait qu'elle a imposé la couverture de la perte attendue en plus de la perte inattendue couverte par les fonds propres.

BIBLIOGRAPHIE

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

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

Basel Committee on Banking Supervision (2006), "International Convergence of Capital Measurement and Capital Standards", Bank for International Settlements.

Bewick, V., Cheek L. and Ball L.(2004), "Receiver operating characteristic curves", Critical Care Vol 8 No 6,508-512.

Cohen, B. and Edwards, Jr G.(2017), "The new era of expected credit loss provisioning", BIS Quarterly Review, March 2017, 36-59.

Courdec, F. and Renault, O.(2005), "Times-to-default: life cycle, global and industry cycle impacts", Research Paper 142, FAME-International Center of Financial Asset Management and Engineering, Université de Genève. https://www. researchgate.net/publication/5021538 _Times-To- DefaultLife_ Cycle_ Global_ and_Industry_ Cycle_Impact

Figlewski, S., Frydman, H. and Liang, W.(2012), "Modeling the effect of macroeconomic factors on corporate default and credit rating transitions", International Review of Economics and Finance 21 (2012) 87–105.

Grunert, J., Norden, L. and Weber, M. (2005), "The role of non-financial factors in internal credit ratings", Journal of Banking & Finance, 29: 509–531.

Habachi, M. (2009), " les modèles internes de quantification des risques de crédit et opérationnel dans les banques marocaines: oppurtinités et risque de modèle", thèse de doctorat, Fsjes agdal, Université Mohammad V, Rabat.

Habachi, M. and Benbachir, S. (2018), "Assessing the Impact of Modelling on the Expected Credit Loss (ECL) of a Portfolio of Small and Medium-sized Enterprises", Universal Journal of Management 6(10): 409-431. DOI:10.13189/ujm.2018.061005.

Habachi, M. and Benbachir, S.(2019), "Combination of linear discriminant analysis and expert opinion for the construction of credit rating models: The case of SMEs", Cogent Business & Management (2019), 6: 1685926. https://doi.org/10.1080/23311975 .2019.1685926.

Hensher, D.A., Jones, S., William, H. and Greene, W H.(2007), "An Error Component Logit Analysis of Corporate Bankruptcy and Insolvency Risk in Australia", The Economic Record, 83, 86-103.

Hosmer, D. W. and Lemeshow, S. (1980), "A goodness-of-fit test for the multiple logistic regression model", Communications in Statistics, A10, 1043-1069.

Hosmer, D., Lemeshow, S., and Sturdivant, R. (2013), "Applied Logistic Regression", 3rd Ed., John Wiley & Sons, New York.

Hunter, J. and Isachenkova, N.(2002), "A panel analysis of UK industrial company failure", ESRC Centre for Business Research Working Paper 228, Cambridge University. https://www.cbr.cam.ac.uk/fileadmin/user_upload/centre-for-business-research/downloads/working-papers/wp228.pdf

Kristóf, T. and Virag, M.(2017), "Lifetime probability of default modeling for Hungarian corporate debt instruments", 31st European Conference on Modelling and Simulation. 23-26 May 2017, Budapest, Chapter: Finance and Economics and Social Science (FES), Publisher: European Council for Modelling and Simulation. https://www.researchgate.net/publication/317098064_Lifetime_probability_of_default_modeling_for_Hungarian_corporate_debt_instruments.

Ohlson, J. A. (1980), "Financial Ratios and the Probabilistic Prediction of Bankruptcy", Journal of Accounting Research, 18-1,109-131.

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, 4-6, 437-445.

Taffler, R.J. (1982), "Forecasting company failure in the UK using discriminant analysis and financial ratio data", Journal of the Royal Statistical Society, Series A, General, 145(3), 342-358.

Vaněk, T. and Hampel, D.(2017), "The probability of default under IFRS9: Multi-period estimation and macroeconomic forecast", Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis. 2017, V.65, N° 2, 759-776.

Xu, X.(2016), Estimating Lifetime Expected Credit Losses Under IFRS 9, Electronic copy available at: https://ssrn.com/abstract=2758513.

Yildirak, K. and Suer, O. (2013), "Qualitative determinants and credit-default risk: evidence from turkey", АКТУАЛЬНІ ПРОБЛЕМИ ЕКОНОМІКИ №7(145).