

Mineur Recherche

RAPPORT

Sujet

Forecast of banks 'default and firms ' returns using their respective historical financial statements



Abstract

The bank's transparency isn't sufficient to determine their risk management compared to behaviors that could lead to bank failures. The work done by Allen N. Berger and Christa HS Bouwman allow us to conclude regarding the importance of capital growth on the risks of bankruptcy, and the different forms it may take. We can insist on the particular importance of liquidity, whose strong presence considerably reduces the risks of bankruptcy.

Using accounting data from French banks between 1900 and 1940, we sought to identify the risky behavior of these banks. The first step was to extract and make this data usable thanks to the python programming language. Then we had to determine the comparisons of economic equations made to know their respective risk of bankruptcy. Please note that the parameters of these equations come from the given file. The choice of comparisons was made using a graphical analysis of the difference in francs of the selected sectors. The larger it is, the more potentially visible the risk. These sectors are capital ratio, liquidity, deposit risk and securities. The other parameters were not used and did not fit into any sector specified by our mentors

The predictive algorithmic model built allow us to find the risk of bankruptcy, between 0 and 1, of one equation compared to another, thanks to logistic regression. We were able to conclude, like Mr. Berger and Mrs. Bouwman, that the importance of capital is the first parameter that can lead to a safety of financial establishments. We also found that certain parameters, such as securities, were more likely to result in bank failure.

Résumé

La transparence des banques n'est pas suffisante pour déterminer la gestion des risques de ces dernières vis-à-vis des comportements pouvant mener à des faillites bancaires. Le travail réalisé par Allen N.Berger et Christa HS Bouwman a permis de conclure quant à l'importance croissante du capital sur les risques de faillites, et les différentes formes qu'il peut prendre. Nous pouvons noter l'importance toute particulière de la liquidité, dont une forte présence réduit considérablement les risques de faillite.

A l'aide des données comptables des banques françaises entre 1900 et 1940, nous avons cherché à dresser le comportement à risque de ces dernières. Il a dans un premier temps fallu extraire et rendre exploitable ces données à l'aide du langage de programmation python.

Puis il a s'agit de déterminer les comparaisons d'équations économiques faites pour connaître leur risque de faillite respectives. Précisons que les paramètres de ces équations proviennent du fichier transmis.

Le choix des comparaisons s'est fait à l'aide d'une analyse graphique sur l'écart en francs des secteurs choisis. Plus il est important plus le risque est potentiellement visible. Ces secteurs sont le capital ratio, la liquidité, le risque de dépôt et les securities.

Les autres paramètres n'ont pas été utilisés ne rentrant dans aucun secteur précisé par le mentor.

Le modèle algorithmique prédictif construit permet de déterminer le risque de faillite, compris entre 0 et 1, d'une équation comparée à une autre, et ce grâce à la régression logistique. Nous avons pu en conclure, comme M.Berger Mme.Bouwman que l'importance du capital est le premier paramètre pouvant mener à une sauvegarde des établissements. Nous avons également noté que certains paramètres, comme les securities, avaient une plus grande probabilité d'entraîner une faillite bancaire.

Sommaire

- 1) Contexte
 - a. Introduction
 - b. Problématique
- 2) Etat de l'art
- 3) Objectifs
- 4) Réalisation du modèle
 - a. Exploitation des données extraites
 - b. Construction du modèle algorithmique
- 5) Résultats et discussion
- 6) Conclusion et perspectives
- 7) Références bibliographiques

1) CONTEXTE

A) Introduction

Les établissements financiers sont dans l'obligation de rendre public leur bilan comptable et compte d'exploitation à des fins de transparence. Ceci permet de rassurer le public et de rendre le contrôle plus simple.

Toutefois cette transparence peut être considérée insuffisante. En effet si nous prenons l'exemple de la crise bancaire la plus récente, la crise des subprimes de 2008, nous savons qu'elle a été dû à l'éclatement de la bulle du marché immobilier américain, notamment suite à la gestion et au contrôle inexistant des différents produits bancaires. Aucune transparence n'est faite sur les échanges réalisés par et entre les banques.

Cette crise soulève également un deuxième point, la non utilisation des données financières passées. Les informations comptables des banques, au minima en France, peuvent remonter jusqu'à 1900. Même si aucune régulation n'existait à cette période, il reste utile de comparer l'évolution des chiffres des banques de l'époque afin de déterminer si des tendances précédents une crise bancaire ne se dessinent pas.

L'intérêt de ces données est leur fort potentiel de prédiction de comportements à risque de banques pouvant mener à leur faillite. Leur exploitation peut avoir lieu grâce aux nouvelles technologies.

Nous pouvons citer le machine learning et le langage de programmation python qui permettent de pouvoir réaliser ce travail.

Il va tout particulièrement intéressant d'effectuer cette recherche sur un ensemble de banques des années 1900, avec l'absence de contrôles rendant ces comportements plus visibles.

B) Problématique

Dès lors il va être intéressant de se demander quel a pu être le comportement à risque de banques entre 1900 à 1940 ?

Pour répondre à cette problématique nous nous appuyons sur la documentation fournie par M.RIVA Angelo et M.JUN KIM Jae Jun.

Elle consiste en un fichier excel regroupant les données comptables des banques françaises de cette période.

2) ETAT DE L'ART



How does capital affect bank performance during financial crises? ☆



Allen N. Berger^{a,b,c,*}, Christa H.S. Bouwman^{b,d,1}

^a University of South Carolina, Moore School of Business, 1705 College Street, Columbia, SC 29208, USA

^b Wharton Financial Institutions Center, University of Pennsylvania, Philadelphia, PA 19104, USA

^c Center for Economic Research (CentER)—Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

^d Case Western Reserve University, Weatherhead School of Management, 10900 Euclid Avenue, 362 PBL, Cleveland, OH 44106, USA

La solution au sujet de recherche va s'appuyer sur l'article suivant « How does capital affect bank performance during financial crisis ? », écrit par Allen N.Berger et Christa H.S.Bouwman.

Dans leur approche les deux chercheurs ont décomposé les banques étudiées en 3 catégories en fonction de la taille de leur capital. Les petites, moyennes et grandes banques en fonction de l'importance de leur capital, en dollars, respectivement inférieur à 1 milliard, compris entre 1 et 3 milliards et supérieur à 3 milliards.

Lors de tests de variations de capitaux des banques pour déterminer leurs effets sur des banques lors de périodes de crise, ils ont utilisé la méthode de la régression logistique.

Les chercheurs ont retenu deux facteurs dans l'étude de la prépondérance du capital, la quantité de fonds propres possédés par la banque et le recours ou non à une M&A assistée.

M.Berger et Mme.Bouwman ont choisi d'étudier les paramètres des banques lors de 10 crises financières de 1984 à 2010.

Pour avoir une comparaison réaliste des crises et des temps normaux, ils ont créé des fausses crises.

Pour éviter les erreurs de variation, les régressions contiennent des variables de contrôle.

Elles sont classée en 7 catégories. Risques et opacités, taille et filet de sécurité, ownership, compétitivité, location et rentabilité.

Les chercheurs ont pu remarquer que les grandes banques ont un taux de survie supérieur, suivie des petites et des moyennes. On explique ce résultat par le lien entre les petites et les grandes. En effet de nombreuses banques dont le capital est relativement faible sont locales. Approchées par de plus importantes institutions, elles ont accepté des fusions ou des rattachements à des groupes ce qui explique leur solidité apparente malgré un capital au départ des crises plus faible que celui des banques de taille moyenne.

3) **OBJECTIFS**

- Exploiter les données d'un fichier csv
- Implémenter des équations économiques
 - Capital Ratio
 - Securities
 - Liquidité
 - Risque de Dépôts
 - Risque de Crédits
- Construire un algorithme de prédiction => Régression Logistique
- Obtenir des résultats dont les conclusions sont similaires à celles de Allen N Berger et Christa HS Bouwman
- Déterminer l'importance du rôle des différents paramètres économiques sur le risque de faillite des banques
- Obtenir un code réutilisable et facilement compréhensible

4) REALISATION DU MODELE

A) *Exploitation des données extraites*

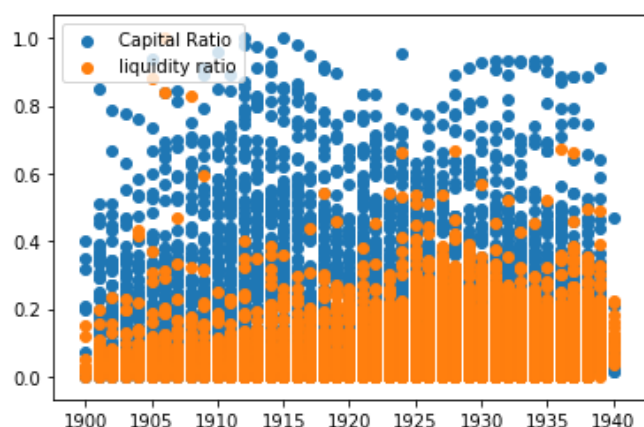
L'exploitation des données passe d'abord par le choix des équations.
Pour déterminer les équations comparables, nous avons choisis d'effectuer une analyse graphique. Plus la différence entre 2 équations est relevée, plus la comparaison sera pertinente. Si la différence n'est pas visible la comparaison n'est pas retenue.

Le procédé de programmation nous ayant mené au choix des équations est décrit dans la partie B.

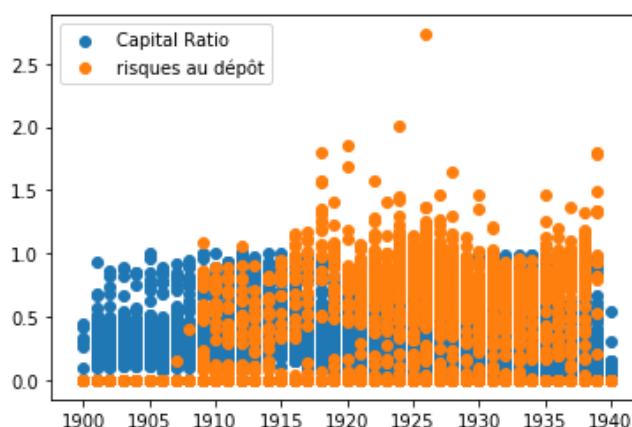
Voici l'ensemble des équations selon leur secteur :

Capital Ratio	$\frac{\text{capital}}{\text{Total Assets}}$	$\frac{\text{capital} + \text{capital surplus}}{\text{Total Assets}}$	$\frac{\text{paid in capital}}{\text{Total Assets}}$	$\frac{\text{paid in capital} + \text{capital surplus}}{\text{Total Assets}}$
Liquidity Ratio	$\frac{\text{cash}}{\text{Total Assets}}$	$\frac{\text{cash} + \text{commercial portefolio}}{\text{Total Assets}}$		
Risque de Dépôts	$\frac{\text{on demand}}{\text{Total Assets}}$	$\frac{\text{on demand} + \text{on time}}{\text{Total Assets}}$		
Securities				
Risque de Crédits	$\frac{\text{tot credits}}{\text{Total Assets}}$			

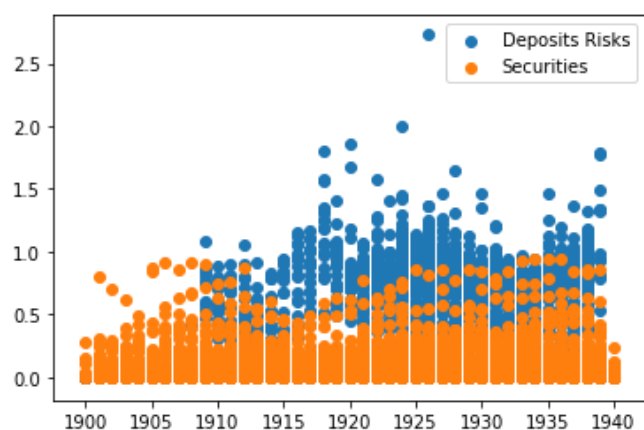
Voici les graphes retenus et les comparaisons déduites.



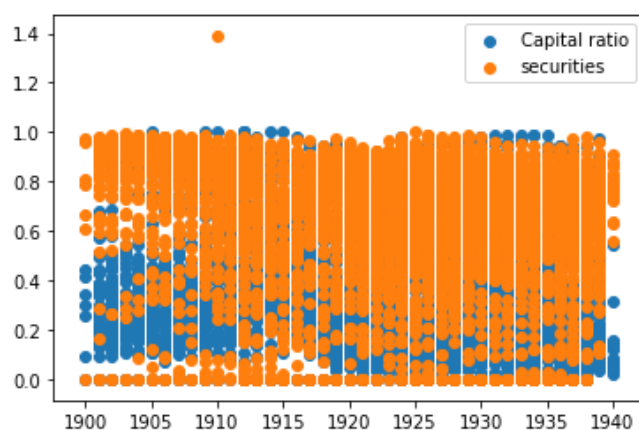
Titre : Comparaison des valeurs de la première équation de capital ratio et de la 1ère de liquidité entre 1900 et 1940 pour toutes les banques.



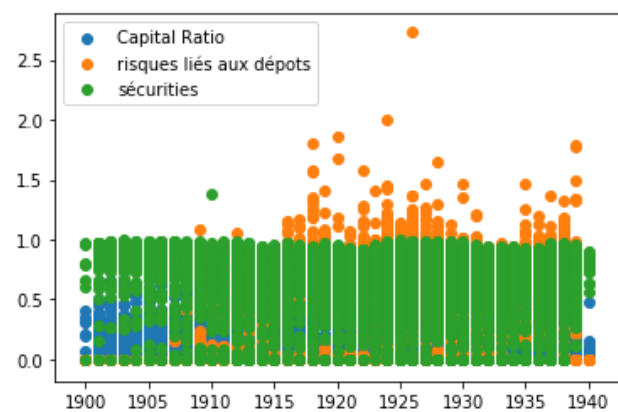
Titre : Comparaison des valeurs de la 2ème équation de capital ratio et de la 1ère de risque de dépôt entre 1900 et 1940 pour toutes les banques.



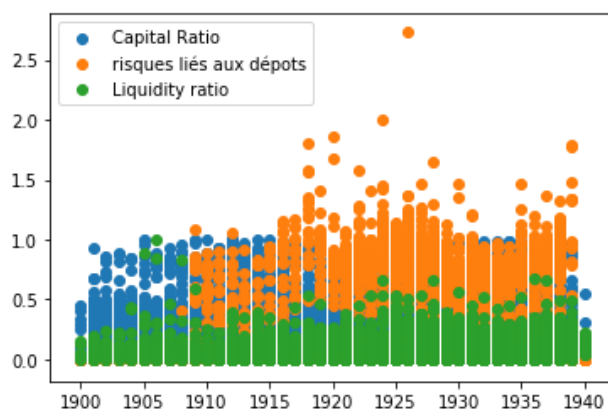
Titre : Comparaison des valeurs des securities et de la 1ère de risque de dépôt entre 1900 et 1940 pour toutes les banques.



Titre : Comparaison des valeurs de la 2ème équation de capital ratio et des securities entre 1900 et 1940 pour toutes les banques.



Titre : Comparaison des valeurs de la 3^{ème} équation de capital ratio, de la 1^{ère} de risque de dépôts et des securities entre 1900 et 1940 pour toutes les banques.



Titre : Comparaison des valeurs de la 2^{ème} équation de capital ratio, de la 1^{ère} de risque de dépôts et de la 2^{ème} de liquidité entre 1900 et 1940 pour toutes les banques.

B) Construction du modèle algorithmique

1) Lecture du fichier Excel

- a. Ouvrir le fichier à l'aide de la bibliothèque pandas
- b. Extraire les données du fichier et les stockées en mémoire dans une variable

2) Création d'une base de données

- a. Instanciation du nombre de lignes que va parcourir l'algorithme
- b. Création des données x et y
 - i. Avec x : contenant les colonnes du fichier dont les données sont utilisées
 - ii. Avec y : contenant les résultats de faillite f1 (1 si faillite bancaire, 0 sinon)

3) Training et Test

- a. Initialisation des équations à comparer en tant que paramètres de type dataframe
- b. Traitement des données
 - i. Remplacer les valeurs NAN et négatives par 0
- c. Création des données de training et de test
- d. Affichage de ces données
- e. Création des matrices de training et de test à l'aide de la bibliothèque numpy
- f. Ajout de la colonne de 1 aux matrices

4) Régression logistique

a. Fonction Hypothèse

- i. On définit l'hypothèse comme étant le produit matriciel de X par θ .

Avec $X = X_{\text{train}}$ et θ le poids de ma fonction, soit la valeur pour laquelle l'erreur doit être minimale

b. Fonction de coût

- i. Elle renvoie la valeur de θ , à partir des paramètres X et Y, ici X_{train} et Y_{train} , comme étant le produit matriciel du produit matriciel de la transposée de X par Y par l'inverse du produit matriciel de la transposée de X par X

```
#Fonction CFS
def CFS(X, Y):
    p1 = np.dot(X.T,X)
    p2 = np.dot(X.T, Y)
    #Theta = np.linalg.pinv(p1).dot(p2)
    Theta = np.linalg.pinv(p1)
    Theta = Theta.dot(p2)
    #Theta = Theta[~np.isnan(Theta)]
    return Theta
```

c. Fonction Erreur

- i. Renvoie la fiabilité du modèle. Plus l'erreur est proche de 0 plus le modèle est fiable.

La fonction erreur instancie la valeur du résultat prédit

- ii. La valeur de l'erreur correspond au mean square error (MSE).

$$MSE(\theta) = \text{Biais}(\theta)^2 + \text{Var}(\theta)$$

```
#Fonction Erreur
def Erreur(x,y,theta):
    ypred = np.dot(x,theta)
    I = len(y)
    # boucle for pour parcourir les matrices
    for i in range(I):
        erreur = (1/I) * np.sum(np.square(ypred-y)) #MSE
    return erreur
```

5) Paramètres optimaux

- a. Instanciation de Theta avec l'appel à la fonction de coût
- b. La définition de la variable renvoyant la valeur prédictive des résultats avec la fonction de l'hypothèse
- c. Définition de la variable renvoyant l'erreur estimée du modèle.

6) Affichage des résultats

- a. A l'aide de la bibliothèque matplotlib.pyplot

7) RESULTATS ET DISCUSSION

Nous allons maintenant présenter et expliquer les résultats obtenus

Erreur : 0.013072892066836423

Yprediction : 0.01533674149|77466|

Titre : *Risque de faillite avec intégration du capital ratio, risques de dépôts et securities.*

Il est important de souligner l'écart entre la valeur prédite et la valeur du résultat actuel.

En effet au moins à environ 0.2, un écart d'environ 0.8 est une marge d'erreur trop importante.

Mais cette prédiction permet tout de même une prédiction du risque de faillite grâce à sa valeur moyenne.

La grande majorité des données est étudiée pour une faillite nulle, aussi celles étudiées dans une période à fort risque de faillite par exemple (la période est ici un paramètre) ont une sont prédites avec une valeur supérieure à la moyenne.

Il s'agit d'un indicateur que nous considérons fiable.

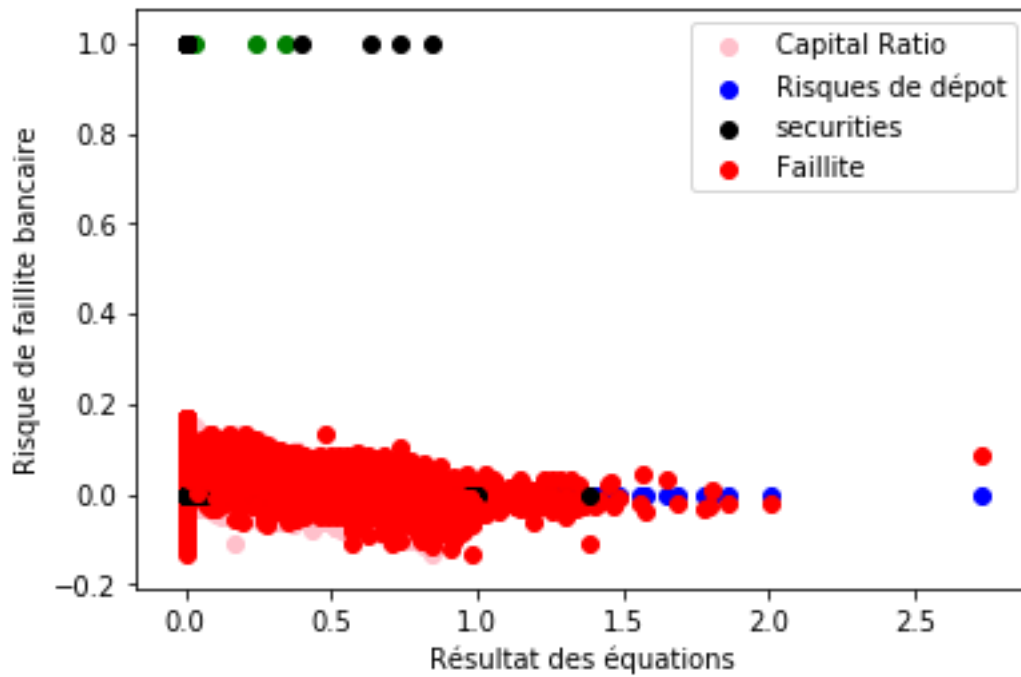
De plus nous pouvons affirmer que pour chaque comparaison dont la valeur prédite est supérieure à 0.1, le risque de faillite est avéré.

Exemple pour la même comparaison que ci-dessus :

```
1 np.amax(Yprediction)
0.16659338546474212

1 np.mean(Yprediction)
0.01533674149871485

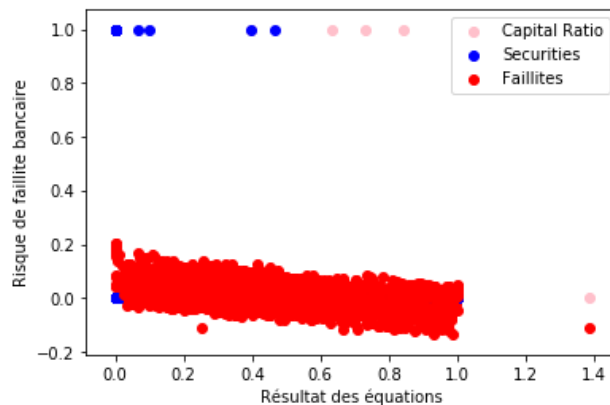
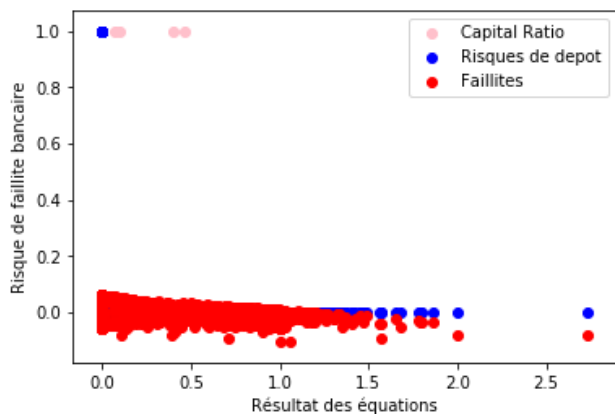
1 print(Yprediction[1007:1010], '\n', Ytrain[1007:1010])
[[0.0203323 ]
 [0.15357021]
 [0.04080659]]
[[0]
 [1]
 [0]]
```



Titre : *Risque de faillite avec intégration du capital ratio, risques de dépôts et securities.*

Ce graphique nous permet de conclure quand à l'importance de la quantité de capital. En effet l'article de M.Berger et Mme.Bouwman insistait sur cette importance, et principalement sur la liquidité d'une banque.

On remarque ici que les faibles quantités de capitaux, comprises au ratio de 0 à 1 environ présentent le plus élevé risque de faillite.



Titre : *Comparaison du risque de faillite entre le capital ratio et les risques de dépôt entre 1900 et 1940 pour toutes les banques.*

Titre : *Comparaison du risque de faillite entre le capital ratio et les securities entre 1900 et 1940 pour toutes les banques.*

Nos résultats nous permettent également de conclure sur le rôle des securities comparé à la liquidité et au capital ratio.

On observe un risque de faillite moyen beaucoup important pour une comparaison aux securities, ainsi qu'une valeur maximale de faillite plus importante.

Entrée [3]: 1 np.amax(Yprediction)

Out[3]: 0.20188178770268572

Entrée [4]: 1 np.mean(Yprediction)

Out[4]: 0.015336741499989532

Entrée [4]: 1 np.amax(Yprediction)

Out[4]: 0.06313051923296742

Entrée [5]: 1 np.mean(Yprediction)

Out[5]: 0.0153367414977466

Titre : *Valeur max et moyenne du risque de faillite pour le capital ratio comparé aux securities.*

Titre : *Valeur max et moyenne du risque de faillite pour le capital ratio comparé aux risques de dépôt.*

8) CONCLUSION ET PERSPECTIVES

Nous aimerions indiquer que travailler sur ce sujet s'est révélé passionnant et n'a eu pour seule limite des contraintes de temps et d'effectif.

Nous avons pu retrouver les résultats de l'article de recherche de référence, ainsi qu'obtenir les rôles croissants des paramètres comparés. Il nous est donc possible de conclure qu'une quantité trop faible de securities aura un risque de faillite plus grand qu'un même ratio de capital.

Certaines fonctions n'ont pas été utilisées dans notre approche. Celles-ci ont pour vocation de simplifier le possible futur travail de continuité sur ce sujet. Nous pouvons notamment citer une fonction `Logistic_Regression()`, devant permettre une régression d'elle-même, et `costFunctionDerivative()`, indispensable au bon fonctionnement de `Logistic_Regression()`.

Le code a donc vocation à être réutilisé et peut être optimisé. Une approche possible serait de comparer les résultats à partir de mêmes paramètres entre les fonctions présentées dans ce rapport, soit une régression logistique, et un fichier py utilisant les algorithmes bayésiens.

Nous aimerions indiquer que travailler sur ce sujet s'est révélé passionnant et n'a eu pour seule limite des contraintes de temps et d'effectif.

Références Bibliographiques

“ How does capital affect bank performance during financial crisis ?”, Allen N.Berger & Christa H.S.Bouwman, 11 Mars 2011

“Banking Crisis : A Review,” Luc Laeven, 2011

“Flight to safety and the credit crunch: a new history of the banking crises in France during the Great Depression”, Patrice Baubeau & Eric Monnet & Angelo Riva & Stefano Ungaro, 2020

<https://github.com/markdregan/Bayesian-Modelling-in-Python>