

UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

MÉMOIRE DE MAÎTRISE  
PRÉSENTÉ À  
L'UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN ÉCONOMIE FINANCIÈRE

PAR ALEXA LABEL

LE RISQUE DE CRÉDIT DANS LE MARCHÉ HYPOTHÉCAIRE  
DU CANADA

MAI 2020

## Table des matières

CHAPITRE 1: INTRODUCTION.....	3
CHAPITRE 2 : REVUE DE LA THÉORIE ET DE LA LITTÉRATURE .....	12
A. Probabilité de défaut .....	19
Modèles de probabilité de défaut pour prêt individuel .....	20
Modèles de probabilité de défaut pour portefeuille de prêts.....	25
B. Taux de recouvrement et pertes en cas de défaut.....	27
Modèle utilisé pour la perte en cas défaut.....	28
C. Exposition en cas de défaut.....	31
CHAPITRE 3 : MÉTHODOLOGIE.....	31
Données.....	32
Limitations .....	34
Analyse sommaire de l'échantillon de données .....	35
CHAPITRE 4 : ANALYSE ET RÉSULTATS.....	39
Modèle de régression logistique .....	39
Modèle de régression linéaire de portefeuille .....	40
CHAPITRE 5 : CONCLUSION .....	52
ANNEXES.....	54
Références.....	66

## CHAPITRE 1: INTRODUCTION

La réglementation bancaire assure une certaine stabilité dans les marchés financiers. Chaque système financier national est responsable d'assurer la protection de ses déposants, la sécurité des systèmes de paiements ainsi qu'un contrôle du risque systémique. Avec les mouvements de globalisation observés dans les dernières décennies, il eut un besoin d'avoir une réglementation bancaire internationale afin d'harmoniser le marché global d'où l'institution du Comité de Bâle en 1974. Le Comité de Bâle a pour but premier d'énoncer des recommandations aux banques des pays du G10 actives sur le plan international. Ses normes sont ensuite appliquées par les autorités de surveillance nationale, le Bureau du surintendant des institutions financières (BSIF) au Canada. Le BSIF est d'ailleurs responsable de la réglementation et supervision des institutions financières à travers le Canada. En se basant sur les normes de Bâle, le BSIF établit entre autres les lignes directrices concernant les fonds propres, les limites à respecter, les rapports de gestion ainsi que les pratiques commerciales et financières considérées saines et acceptables pour assurer la stabilité du secteur financier canadien.<sup>1</sup>

Le plus récent des Accords de Bâle, les normes Bâle III, publiées en décembre 2010 sont conçues de manière à améliorer la gestion du risque de crédit par rapport aux normes précédentes Bâle II suivant la crise financière mondiale de 2007 à l'aide de cinq caractéristiques clés principales<sup>2</sup> :

---

<sup>1</sup> Banking Regulation 2017, Getting the Deal Through, McMillan LLP (<https://mcmillan.ca/webfiles/Canadian-Banking-Regulation-2017.pdf>)

<sup>2</sup> Comité de Bâle sur le contrôle bancaire, *Finalisation de Bâle III – En Bref*.  
[https://www.bis.org/bcbs/publ/d424\\_inbrief\\_fr.pdf](https://www.bis.org/bcbs/publ/d424_inbrief_fr.pdf)

- Augmentation du niveau et de la qualité des fonds propres : les banques doivent conserver davantage de fonds propres de meilleure qualité afin de faire face à des pertes imprévues.
- Extension de la couverture des risques : les exigences de fonds propres sont dorénavant calculées sur la base d'une période de 12 mois sous tension.
- Limitation du levier bancaire : un ratio de levier bancaire limitant l'accumulation de dette visant à financer les investissements et activités bancaires
- Amélioration de la liquidité des banques : Un ratio de liquidité à court terme exigeant que les banques détiennent suffisamment d'actifs pour couvrir leurs besoins pendant 30 jours en période de tensions.
- Limitation de la procyclicité : en période de forte croissance économique, les banques affectent des fonds à la constitution de volants de fonds propres à utiliser en périodes de tensions économiques.

Le Comité de Bâle met particulièrement l'emphase sur deux types de précautions à prendre afin de limiter le risque pris par les institutions financières. La première est la détermination du capital minimum requis afin d'assurer le bon fonctionnement de l'institution. La deuxième est la procédure de stress-testing qui permet l'évaluation de la robustesse de l'institution en situation de crise. Les normes imposent donc aux banques un capital réglementaire, établi par rapport au volume des actifs pondérés en fonction du risque (APR) qui doit ensuite être soumis à un exercice de tolérance au stress.

(Formule 1)

Ratio de fonds propres pondérés par leur risque

$$= \frac{\text{Fonds propres réglementaires}}{\text{Actifs pondérés par le risque}}$$

Les fonds propres réglementaires comprennent les actions ordinaires, bénéfiques non distribués et autres instruments de fonds propres sans échéance fixe dite «catégorie 1», puis la dette subordonnée et les réserves générales pour pertes dite «catégorie 2». Les actifs pondérés par le risque comprennent les liquidités, les valeurs mobilières et les prêts. Chaque groupe d'actifs représente un risque différent et doit être pondéré en fonction du risque qu'il représente pour la banque. La valeur de l'actif est donc multipliée par la pondération adéquate (entre 0% et 100%) ce qui permet une contribution de fonds propres moins élevé pour les actifs plus sûrs, et plus élevé pour les actifs plus risqués.

Les actifs pondérés par le risque sont calculés en additionnant trois composantes en fonction du type de risque qu'ils représentent, c'est-à-dire le risque de crédit, le risque opérationnel et le risque de marché.<sup>3</sup>

(Formule 2)

$$\frac{\text{Fonds propres}}{\text{APR de crédit}_{\text{Standard}} + 1.06 \times \text{APR de crédit}_{\text{NI}} + 12.5 \times \text{Risque opérationnel} + 12.5 \times \text{Risque de marché}}$$

où :

---

<sup>3</sup> Normes de fonds propres (NFP) : Chapitre 1 – Vue d'ensemble, BSIF ([http://www.osfi-bsif.gc.ca/fra/fi-if/rg-ro/gdn-ort/gl-ld/Pages/CAR18\\_chpt1.aspx](http://www.osfi-bsif.gc.ca/fra/fi-if/rg-ro/gdn-ort/gl-ld/Pages/CAR18_chpt1.aspx))

**APR de crédit standard :** L'actif pondéré en fonction des risques de crédit, calculé selon l'approche standard.

**APR de crédit NI :** L'actif pondéré en fonction des risques de crédit, calculé selon l'approche fondée sur les notations internes (NI).

**Risque opérationnel :** L'exigence de fonds propres à l'égard du risque opérationnel, qui ne sera pas abordé dans ce papier.

**Risque de marché :** L'exigence de fonds propres à l'égard du risque de marché, qui ne sera pas abordé dans ce papier.

Les institutions financières impliquées dans l'octroi des prêts hypothécaires résidentiels font face au risque de crédit. Elles doivent donc suivre les lignes directrices du BSIF afin de quantifier ce risque et déterminer les fonds propres nécessaires. Une approche par notation d'emprunteur est la pratique courante et acceptée de l'industrie.

### *Approche par les notations*

Basée sur le concept d'une transition binaire avec la présence de défaut versus l'absence de défaut, une approche plus approfondie s'est développée en attribuant des notations de crédit à chaque emprunteur de prêts hypothécaires résidentiels de la même manière que l'on attribue des notes d'examen à un étudiant au-delà du *pass or fail* (réussite ou échec). En effet, il existe plusieurs études qui démontrent que la notation attribuée à un emprunteur est hautement corrélée avec le risque de crédit qu'il représente. Une étude conduite par Moody's en 1997 permet de déterminer qu'au cours des 77 dernières années précédant l'étude, seulement 11% des cas de

changement de notations observées ont été de plus d'une catégorie<sup>4</sup>, nous permettant donc de conclure que la majorité des cas de défaut sont précédés par la dégradation progressive de leur notation. De plus, les emprunteurs avec une meilleure notation avaient une plus grande probabilité de maintenir cette meilleure notation que ceux avec une notation inférieure.

Les normes Bâle III permettent aux institutions financières d'évaluer leur risque de crédit avec l'approche par notation selon deux manières : l'approche standard soit l'utilisation des notations externes d'agences spécialisées, ou l'approche fondée sur les notations internes (*Internal Ratings-Based*) soit l'utilisation des notations produites par l'entité même.

#### *L'approche standard*

L'approche standard indique que les institutions financières doivent utiliser les notations d'une agence de crédit spécialisée externe (OEEC ou organisme externe d'évaluation du crédit) ou une pondération de risque générale prescrite par l'autorité nationale si une notation d'OEEC n'est pas disponible. Cette méthode est la plus simple, la moins coûteuse et considérée comme étant qualitative. Le Comité de Bâle indique que chaque autorité de contrôle national (le BSIF au Canada) est responsable de déterminer l'éligibilité et la reconnaissance des agences de crédit pouvant être utilisées par les institutions financières en se basant sur six critères d'évaluation : l'indépendance, l'objectivité, l'accès international/ la transparence, la crédibilité, la communication et les ressources. À ce jour, le BSIF indique que les seules agences reconnues au Canada sont DBRS, Moody's, S&P puis Fitch<sup>5</sup>. Les OEEC obtiennent

---

<sup>4</sup> Carty, Lea V. *Moody's Rating Migration and Credit Quality Correlation, 1920-1996*. New York : Moody's Investors Service, Global Credit Research, July 1997. Report Number 25097.

<sup>5</sup> BSIF. *Lignes directrices, Normes des fonds propres (NFP)*. Décembre 2016. Chapitre 3 - 3.6.1 Évaluations externes du crédit.

leur notation en utilisant une série de données qui permet à un expert ou un comité de notation de prendre une décision finale sur la notation à attribuer à l'emprunteur. Par exemple, pour une entreprise de gestions de propriétés locatives, la notation pourrait être basée sur les ratios financiers de l'entreprise, la qualité de la gestion et des ressources humaines, la qualité des propriétés et de ses locataires, la structure des baux, la liquidité de son marché et plus encore, d'où la nature qualitative de cette approche. Les notations peuvent donc varier d'un OEEC à un autre. Les normes Bâle indiquent qu'on doit toujours utiliser le même OEEC, ou la notation la plus sévère. Bien que les notations varient entre les différentes OEEC, les concepts sont semblables et chacun divise leur notation en catégorie générale nommée «investment grade», pour les emprunteurs démontrant une certaine stabilité et un risque modéré, et «non-investment grade» pour les emprunteurs à risque élevé. À l'intérieur de ces catégories, nous avons les sous-catégories représentant l'échelle de notation attribuable à chaque emprunteur individuel avec les notations variant de Aaa/AAA à Caa/CCC. La notation attribuée permet ensuite d'attribuer un pourcentage utilisé dans la pondération du risque puis dans l'allocation des fonds propres nécessaires.

Tableau 1 – Pondération de risque pour l'exposition sur les entreprises.

<b>Expositions sur les entreprises</b>						
Pondérations des risques dans les juridictions où l'approche fondée sur les notes est autorisée						
Notation externe de la contrepartie	De AAA à AA-	De A+ à A-	de BBB+ à BBB-	BB+ à BB-	Inférieure à BB-	Non notée
Pondération	20%	50%	75%	100%	150%	100 % 85 % en cas de PME

Concernant les hypothèques résidentielles, les agences d'évaluation de crédit disponible au Canada telles qu'Equifax ou TransUnion ne sont pas reconnues comme des OEEC. Les normes de Bâle III prescrivent donc la pondération à utiliser. Sous Bâle II, il y avait seulement une pondération unique de 35% pour toutes les



expositions à l'immobilier résidentiel. Avec l'implémentation de Bâle III, des approches plus sensibles au risque ont été conçues. Les pondérations varient maintenant en fonction du rapport prêt-valeur (RPV ou Loan-to-value, LTV) des prêts hypothécaires.<sup>6</sup>

Tableau 2 – Pondération en fonction du risque pour l'exposition sur l'immobilier résidentiel

Expositions sur l'immobilier résidentiel								
Fourchettes de ratio LTV	Inférieur à 50%	50 % à 60%	60% à 70%	70% à 80%	80% à 90%	90% à 100%	Supérieur à 100%	Critères non satisfaits
<i>Immobilier résidentiel général</i>								
Pondération selon l'approche fondée sur le prêt total	20%	25%	30%		40%	50%	70%	Pondération de la contrepartie
Pondération selon l'approche fondée sur la segmentation du prêt <sup>2</sup>	20%		Pondération de la contrepartie					Pondération de la contrepartie
<i>Immobilier résidentiel producteur de revenu</i>								
Pondération selon l'approche fondée sur le prêt total	30%	35%	45%	60%	75%	105%	150%	

### *L'approche fondée sur les notations internes (NI)*

L'approche fondée sur les notations internes est plus avantageuse que l'approche standard. Outre sa simplicité, l'approche standard mène au désavantage d'un «one size fits all» et d'une gestion des ressources qui n'est pas à son plein potentiel, tandis que l'approche fondée sur les notations internes de l'institution permet une allocation du capital économique plus personnalisé à l'institution financière ainsi qu'une meilleure compréhension du risque encouru et donc une gestion plus efficace des ressources. Le BSIF s'attend à ce que les institutions dont les fonds propres réglementaires dépassent 5 milliards de dollars canadiens, ou dont 10% du total des

<sup>6</sup> Comité de Bâle sur le contrôle bancaire, *Note récapitulative sur les réformes de Bâle III*. Décembre 2017. L'approche standard pour le risque de crédit. ([https://www.bis.org/bcbs/publ/d424\\_hlsummary\\_fr.pdf](https://www.bis.org/bcbs/publ/d424_hlsummary_fr.pdf))

actifs ou passifs sont à l'étranger, utilisent cette méthode pour leurs activités de crédit significatives.<sup>7</sup> Le principe de l'approche fondée sur les NI est basé sur les éléments clés suivants :

- La classification des expositions (entreprises, emprunteurs souverains, banques, clientèle de détail et actions) en fonction des risques de crédit différents attribuables à chaque catégorie d'actifs.
- Pour chacun des regroupements de types d'exposition, une liste des composantes de risques comprenant la probabilité de défaut, la perte en cas de défaut, l'exposition en cas de défaut et l'échéance effective.<sup>8</sup>
- Les pondérations de risque en utilisant chaque composante. Ceci nous permet d'obtenir une fonction continue comparativement à la fonction discrète obtenue en utilisant l'approche standard avec les pourcentages prédéterminés par catégorie de notation.
- Une démonstration transparente de la satisfaction des exigences minimales. Étant donné que chaque institution utilise une approche qui lui est propre, elle doit clairement démontrer à l'autorité de contrôle nationale qu'elle satisfait les exigences établies dans les normes Bâle.

Lorsqu'une institution choisit de suivre la méthode des notations internes, elle doit estimer les 4 composantes du risque puis intégrer ces estimations dans des formules de pondérations du risque tout en respectant les standards établis pour cette

---

<sup>7</sup> Normes de fonds propres (NFP) : Chapitre 1 – Vue d'ensemble, 1.3.1.1. BSIF ([http://www.osfi-bsif.gc.ca/fra/fi-if/rg-ro/gdn-ort/g1-ld/Pages/CAR18\\_chpt1.aspx](http://www.osfi-bsif.gc.ca/fra/fi-if/rg-ro/gdn-ort/g1-ld/Pages/CAR18_chpt1.aspx))

<sup>8</sup> Il y a deux types de systèmes internes de notations : l'approche fondation et l'approche avancée. L'approche fondation permet aux institutions financières d'utiliser leur estimation de PD, et appliquer les estimations de l'autorité de contrôle (BSIF) aux autres composantes de risque (LGD, EAD et M). L'approche avancée permet aux institutions financières d'utiliser leurs propres estimations pour toutes les composantes du risque. Seulement l'approche avancée est disponible pour l'immobilier résidentiel.

méthode<sup>9</sup>. Pour les expositions qui ne sont pas en défaut et qui sont garanties par des hypothèques sur immobilier résidentiel, nous pouvons calculer les pondérations avec la formule suivante.<sup>10</sup>

(Formule 3)

$$K = PCD \times N[(1 - R)^{-0,5} \times G(PD) + (R / (1 - R))^{0,5} \times G(0,999)] - PD \times PCD$$

Actifs pondérés par le risque avec  $NI(APR) = K \times 12,5 \times ECD$

où

**K** : exigence de fonds propres

**PCD** : Perte en cas de défaut

**N (X)** : Fonction de distribution cumulative pour une variable aléatoire standard normale (la probabilité qu'une variable aléatoire normale de moyenne zéro et de variance 1 soit inférieure ou égale à x).<sup>11</sup>

**R** : Corrélation de 0,15

**G(z)** : Fonction cumulative inverse pour une variable aléatoire standard normale (la valeur de x telle que  $N(x)=z$ )<sup>12</sup>

**PD** : Probabilité de défaut

**ECD** : Exposition en cas de défaut

---

<sup>9</sup> Guidelines for Capital Adequacy Requirements (CAR) Chapter 6 – Credit Risk – IRB Approach (6.2.2)

<sup>10</sup> Guidelines for Capital Adequacy Requirements (CAR) Chapter 6 – Credit Risk – IRB Approach (6.4.1)  
[http://www.osfi-bsif.gc.ca/fra/fi-if/rg-ro/gdn-ort/gl-ld/Pages/CAR18\\_chpt6.aspx#6.8](http://www.osfi-bsif.gc.ca/fra/fi-if/rg-ro/gdn-ort/gl-ld/Pages/CAR18_chpt6.aspx#6.8)

<sup>11</sup> La fonction cumulative normale est disponible sur Excel en tant que fonction NORMSDIST.

<sup>12</sup> La fonction cumulative inverse est disponible sur Excel en tant que fonction NORMSINV.

## CHAPITRE 2 : REVUE DE LA THÉORIE ET DE LA LITTÉRATURE

Plusieurs facteurs rendent la modélisation des prêts hypothécaires dans le contexte du risque de crédit plus complexe que beaucoup d'autre modélisation de groupes d'actifs. En autres, plusieurs différents scénarios sont nécessaires afin de capter adéquatement le comportement d'un prêt hypothécaire dans plusieurs états de l'économie; le comportement de chaque prêt individuel n'est pas homogène, beaucoup de facteurs influent sur le comportement des prêts qui peuvent être difficiles à capter, puis différentes périodes d'analyse sont requises. De plus, l'octroi des prêts constitue une activité importante pour la majorité des institutions financières au Canada et un rôle prédominant dans notre économie. Une mauvaise gestion du risque de crédit menacerait la stabilité des marchés financiers.<sup>13</sup>

Pour bien comprendre l'analyse de ce risque de crédit, nous pouvons étudier ses composantes.<sup>14</sup>

- *La perte (P) : le montant de la perte*
- *La probabilité de défaut (PD) : la probabilité qu'un emprunteur fasse défaut dans une période prédéterminée*
- *La perte en cas de défaut (PCD) : la portion d'une exposition qui sera perdue en cas de défaut*
- *L'exposition en cas de défaut (ECD) : l'exposition représentée par le montant non payé en cas de défaut*
- *L'échéance effective (EE) : les échéances contractuelles des expositions*

---

<sup>13</sup> Ligne directrice sur l'octroi des prêts hypothécaires résidentiels, AMF, mars 2018.  
[https://lautorite.qc.ca/fileadmin/lautorite/consultations/institutions-depots/2018-01-19/20171214\\_lignedirectrice-octroi-prets-hypothecaires-residentiels-suivimodifications\\_fr.pdf](https://lautorite.qc.ca/fileadmin/lautorite/consultations/institutions-depots/2018-01-19/20171214_lignedirectrice-octroi-prets-hypothecaires-residentiels-suivimodifications_fr.pdf)

<sup>14</sup> Ebrary.net. [En ligne] [https://ebrary.net/9510/business\\_finance/irb\\_foundation\\_advanced\\_methods](https://ebrary.net/9510/business_finance/irb_foundation_advanced_methods)

Le fondement de toute modélisation de risque de crédit débute avec une formule assez intuitive. La perte probable d'un prêt  $i$  appartenant à une catégorie  $k$  se calcule par le produit de la PD, de la PCD et de l'ECD au moment du défaut  $t$ .

(Formule 4)

$$P_{i,k,t} = PD_{i,k,t} \times PCD_{i,k,t} \times ECD_{i,k,t}^{15}$$

Suivant l'étude de ces concepts fondamentaux, nous pouvons les intégrer dans une approche de l'étude du portefeuille globale en y ajoutant les autres types de risques pour comprendre le concept de capital économique, réserves de fonds propres et la distribution des pertes. Plusieurs études ont été menées sur le sujet dans les dernières années. Voici les principales recherches qui ont été publiées<sup>16,17</sup>:

**Von Furstenberg (1968-1970)** : Détermine que plusieurs caractéristiques du prêt peuvent être utilisées pour déterminer si le prêt fera défaut, incluant le RPV au moment de l'octroi du prêt, la durée, l'échéance, l'âge et le revenu de l'emprunteur. Il poursuit ensuite ses recherches en utilisant le modèle logistique avec les variables de RPV, de la durée et le qualificatif de nouveau prêt ou prêt existant. Il utilise l'équation simplifiée suivante :

$$(\ln D - \ln E)_t = \ln a_0 + a_2 \ln t + a_3 t^2$$

où  $D$  représente la dette,  $E$  l'équité et  $t$  la durée qui lui permet de conclure que ce sont les facteurs les plus significatifs dans la modélisation du taux de défaut.

---

<sup>15</sup> Modélisation du stress testing du risque de crédit appliqué à un portefeuille de prêts aux particuliers, Samir Saissi Hassani, HEC Montréal

<sup>16</sup> Predicting loss given default for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. Mindy Leow and Christophe Mues, International Journal of Forecasting.  
<https://www.sciencedirect.com/science/article/pii/S0169207011000598>

<sup>17</sup> <http://tampub.uta.fi/bitstream/handle/10024/102533/1513083673.pdf>

**Curley et Guttentag (1974)** : Utilise des analyses de simulations et sensibilité afin de déterminer l'impact de la probabilité de prépaiement sur les flux financiers futurs à l'aide de deux approches. La première approche consistait à faire une moyenne de taux de prépaiement pour différentes catégories d'amortissement des prêts (20 ans, 25 ans et 30 ans). Le désavantage de cette méthode étant principalement qu'ils incluaient tout facteur exogène responsable des prépaiements tels que l'augmentation des taux d'intérêt. La deuxième approche consistait à bâtir un modèle de régression dans lequel le taux de prépaiement était expliqué par la durée, l'année d'initiation du prêt, et la relation entre le taux d'intérêt du prêt et le taux d'intérêt du marché dans l'année d'initiation du prêt.

**Vandell (1978)** : Étudie le risque de défaut pour les instruments alternatifs aux hypothèques conventionnelles. Il estime un modèle reliant des variables associées à l'emprunteur, la propriété, et aux instruments alternatifs aux hypothèques conventionnelles au risque de défaut au fil des années. Puis il fait une simulation du risque de défaut sous différents instruments alternatifs en observant le comportement du modèle estimé sous différentes conditions hypothécaires. Les variables utilisées sont le revenu de l'emprunteur au fil des années,  $Y_t$ , le fardeau de la dette hypothécaire c'est-à-dire la portion du revenu utilisée pour le paiement hypothécaire,  $Q_t/Y_t$ , et le ratio d'équité-prêt,  $E_t/V_t$ . Le modèle utilisé est le suivant :

$$\text{Log } D_t = \log C + b_1 \log Y_t + b_2 \log(Q_t/Y_t) + b_3 \log(E_t/V_t) + b_4 \log t$$

Où  $C$ ,  $b_1$ ,  $b_2$ ,  $b_3$  et  $b_4$  sont des constantes, et  $b_1 < 0$ ,  $b_2 > 1$ ,  $b_3 < 0$  et  $b_4 > 0$ .

Il trouve que dans une situation typique d'un emprunteur avec un revenu moyen, le risque de défaut est semblable pour les différents types d'instruments hypothécaires. De plus, pour les emprunteurs à faible revenu, ceux qui choisissent un plus grand

amortissement ou un taux d'intérêt plus élevé, le risque de défaut semble augmenter de manière comparable. Par contre, pour les situations où la contribution d'équité qui semble être le facteur dominant dans le risque de défaut diminue avec une mise de fond ou valeur de la propriété inférieure, le risque s'avère à varier dépendamment de l'instrument.

**Campbell et Dietrich (1983)** : Détermine que les variables macro-économiques affectent également le risque de défaut, tel que le taux de chômage local. Ils utilisent un modèle d'optimisation basé sur le choix de l'emprunteur. Ils présument que celui-ci maximise sa fonction d'utilité définie à l'aide d'un vecteur de choix mutuellement exclusifs et un vecteur de variables d'état exogènes. Le choix d'utilité maximal peut être représenté comme étant une fonction de probabilité de ces variables d'état :

$$P(s_i|X) = f_i(X)$$

Où la somme de probabilités est égale à 1. Les choix de l'emprunteur sont (1) le défaut, (2) le retard de paiement, (3) le prépaiement et (4) la continuité du paiement à jour. Les variables déterminant le choix d'utilité maximal et ainsi limitant la forme de la fonction de probabilité sont le ratio prêt-valeur,  $L_t/V_t$ , le fardeau de la dette hypothécaire c'est-à-dire la portion du revenu utilisée pour le paiement hypothécaire,  $P_t/Y_t$ , la relation entre le taux d'intérêt de l'hypothèque et le taux du marché,  $(r_t-r_0)/r_t$ , le taux de chômage, le qualificatif d'une nouvelle maison ou maison existante puis l'âge du prêt hypothécaire. Ils trouvent que le défaut est en effet relié aux variables utilisées.

**Green et Shoven (1986)** : Étudie ce qui détermine la probabilité qu'une hypothèque soit payée à un certain moment ou un certain âge, et plus précisément, l'impact des

taux d'intérêt sur le prépaiement. Ils utilisent un modèle de risques proportionnels, où :

La probabilité du prépaiement  
d'une hypothèque d'âge ( $a$ ) si  
les facteurs exogènes sont  
 $x_1, \dots, x_n$  au temps  $t$

$$= \lambda(a) \cdot \pi(x_1, \dots, x_n)$$

Où  $\lambda(a)$  est le niveau du risque de référence ou la probabilité de base qu'une proportion de la population puisse effectuer le prépaiement dans des conditions homogènes et stationnaires. Puis le second facteur  $\pi(x_1, \dots, x_n)$  est plus grand ou plus petit que 1 dépendamment de si les caractéristiques exogènes de l'hypothèque  $x_1, \dots, x_n$  augmente ou diminue la probabilité. Les auteurs définissent  $t$  comme étant la date entre l'initiation et la liquidation de l'hypothèque, puis utilisent l'âge du prêt, le nombre de prêts qui ont été prépayés pour un âge donné, et le nombre de prêts à risque d'être prépayés pour un âge donné. Ils déterminent que les taux d'intérêt du marché jouent en effet un rôle clé dans la probabilité du prépaiement.

**Asarnow et Edwards (1995)** : Étudient la PCD pour de prêts commerciaux, sa distribution et les facteurs qui influencent sa variabilité. Ils utilisent la formule de base de la PCD c'est-à-dire :

$$PCD_i = 100 \times [(W_i + ID_i - IC_i - R_i - MSC_i)/IDA_i]$$

Où chaque variable est la somme des valeurs présentes du  $i^{\text{ème}}$  défaut, et où  $W_i$  représente la perte suivant un défaut,  $ID_i$  la perte en intérêts sur le solde restant,  $IC_i$  le montant du paiement d'intérêt,  $R_i$  le montant du paiement du principal qui ne sera pas recouvert,  $MSC_i$  toute autre dépense ou revenu, et  $IDA_i$  le montant initial en défaut. Ils déterminent que la PCD varie en fonction du type de transaction, et qu'on observe une PCD plus importante pour des prêts sans collatéral.



**Ambrose, Capone et Deng (2001)** : Évaluent les valeurs d'options de vente d'un emprunteur qui doivent être atteintes afin de causer un défaut, ainsi que le moment optimal pour des prix et des taux d'intérêt donnés, et les effets de ses options sur la sévérité de la perte. Ils utilisent la formule de base suivante :

$$D_t = L_t - V_t + C_t + P_t$$

Où  $D_t$  représente la valeur présente du défaut,  $L_t$  le solde courant du prêt,  $V_t$  la valeur de la propriété,  $C_t$  le coût d'excès de l'hypothèque, et  $P_t$  la valeur des paiements due sur l'hypothèque. Ils concluent que l'option de défaut est choisie lorsque la valeur présente est positive et supérieure à la valeur présente d'un futur défaut. Ils attribuent également la surutilisation des options de ventes ou l'exercice de l'option à un moment non optimal aux coûts de transactions associés aux changements de vie tels qu'un divorce ou à la présence d'ajout de dettes subordonnées.

**Calem et LaCour-Little (2004)** : Évaluent la réforme des normes de Basel en développant une méthodologie basée sur le risque afin de déterminer un niveau de capital requis pour les prêts hypothécaires. Pour ce faire, ils modélisent une simulation de la distribution de pertes en utilisant une approche non paramétrique ainsi qu'un modèle de survie pour la performance des prêts. Ils estiment ainsi la probabilité de défaut et le taux de récupération sur un échantillon de prêts en défaut. Le taux de récupération est déterminé en utilisant une régression spline afin de considérer la relation non linéaire entre les RPVs (RPV initial à l'octroi du prêt et le RPV au moment du défaut) et le taux de récupération ce qui leur donne un  $R^2 = 0.25$ . Finalement, ils trouvent qu'un niveau de capital approprié varie énormément avec les caractéristiques du prêt, surtout le RPV, ainsi que la diversification géographique

de l'ensemble du portefeuille. Ils concluent donc qu'une réglementation standardisée qui offre peu de flexibilité en termes de diversification du risque mène à une divergence entre le niveau de capital réglementaire et le niveau de capital économique requis.

**Qi et Yang (2009) :** Modélisent des pertes en utilisant des caractéristiques de prêts en défaut, particulièrement ceux à haut RPV, avec des données de compagnies d'assurances hypothécaires privées. Ils utilisent la formule suivante afin de déterminer la PCD :

$$PCD = 100 \times \frac{A + B + C + D - E}{A}$$

Où A représente le solde impayé au moment du défaut, B les intérêts courus pour 3 mois, C les dépenses encourues avec le défaut, D les dépenses d'entretien de la propriété et E la valeur résiduelle de la propriété après les frais encourus. Ils évaluent ensuite la corrélation entre la PCD et certaines variables clés. Ils démontrent que la PCD s'explique d'ailleurs par le RPV, la dépréciation de la propriété, le montant du prêt, le type d'emprunteur, le but du financement et le temps écoulé entre l'initiation et le défaut. Ils régressent la PCD avec le RPV au moment du défaut et trouvent un  $R^2 = 0.6$ , attribuable au fait qu'ils avaient accès à la valeur de la propriété au moment du défaut. Ce modèle ne serait donc pas acceptable sous les normes de Bâle puisqu'elles stipulent que le modèle choisi doit pouvoir être applicable à tous les prêts, et non seulement à ceux en défaut, mais démontre tout de même une valeur explicative assez élevée entre les deux variables.

**Leow et Mues (2012)** : Modélisent la PCD en utilisant deux modèles. Ils évaluent d'abord la valeur ajoutée d'un modèle de probabilité de reprise avec plus que le RPV comme variable explicative. Ils utilisent en fait trois variables, le RPV, un indicateur démontrant si le prêt a déjà été en défaut puis l'âge du prêt et démontrent que sa performance est supérieure au modèle à une variable explicative. Ils utilisent ensuite un modèle de décote afin de produire des estimations de PCD. Ils combinent les deux pour déterminer un pourcentage de perte estimé et trouvent que ce modèle à deux étapes obtient de meilleurs résultats de performance avec un coefficient de détermination plus élevé.

#### A. Probabilité de défaut

La probabilité de défaut calcule la probabilité, pour une période prédéterminée, qu'un emprunteur ne sera pas capable d'effectuer ses paiements.

Suivant l'approche de notations internes, les données utilisées sont surtout basées sur l'emprunteur (source et stabilité du revenu, note de crédit, sommaires des actifs et passifs), le prêt demandé (taux d'intérêt disponible, rapport prêt-valeur) puis les critères de la propriété (emplacement, condition, valeur). Il y a plusieurs modèles utilisés dans l'industrie pour créer sa propre approche interne, voici les principaux.

## Modèles de probabilité de défaut pour prêt individuel<sup>18</sup>

### 1. Modèle de régression linéaire pour le risque de défaut

Le modèle de régression linéaire établit une relation simpliste entre le risque de défaut d'un prêt et des variables exogènes qui pourrait avoir un effet sur ce risque.

$$\text{default risk} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (1)$$

où :

$\alpha$  est une constante,

$X_1, X_2, \dots, X_k$  représentent les variables exogènes, qui permettent de déterminer le risque de défaut

$\beta_1, \beta_2, \dots, \beta_k$  représentent les coefficients, qui établit l'importance que chaque variable exogène peut avoir sur le risque de défaut

$\varepsilon$  est un terme d'erreur indépendant.

Ici,  $x$  peut prendre la valeur de 0 ou 1. Si l'hypothèque est à jour,  $x$  prend la valeur de 0. Si l'hypothèque est en défaut,  $x$  prend la valeur de 1. Le risque de défaut pour chaque prêt individuel n'est donc pas mesuré par la probabilité du défaut, étant donné que le prêt est en défaut ou il ne l'est pas. La probabilité de défaut se calcule sur l'ensemble des prêts et non sur un prêt individuel. Les variables exogènes pourraient être basées sur certains facteurs tels que le ratio prêt-valeur, la note de crédit et le revenu de l'emprunteur par exemple. En alignant les données à la formule (1), nous pouvons estimer la valeur des coefficients  $\beta_k$ , puis en utilisant ces estimations, déterminer un risque de défaut pour un prêt donné.

---

<sup>18</sup> Tous les modèles présentés dans ce papier proviennent du papier de recherche suivant: Residential Mortgage Probability of Default, Models and Methods by Mingxin Li, Risk Surveillance and Analytics, Financial Institutions Commission of British Columbia

Même si ce modèle est simple à utiliser, il y a quelques désavantages. Puisqu'il détermine le risque de défaut par l'état du prêt, le risque de défaut peut seulement prendre la valeur de 0 ou 1 ce qui limite notre analyse des résultats et ne nous permet pas de déterminer la probabilité du défaut. De plus, si nous voulions stresser nos résultats en appliquant des valeurs plus écartées et que ceci nous donnait une valeur autre que 0 ou 1, les résultats ne pourraient pas être interprétés.

## 2. Modèle logistique

Le modèle logistique est une transformation de l'équation (1) qui nous permet de déterminer la probabilité de défaut comme une fonction logistique basée sur nos variables exogènes. Il transforme donc notre résultat d'une valeur discrète à une valeur continue dans un intervalle entre 0 et 1.

$$P(\text{loan status} = 1) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (2)$$

Où :

$P(\text{loan status} = 1)$  est la probabilité qu'un prêt soit en défaut.

Les coefficients  $\beta$  établissent encore l'importance de chaque variable exogène, mais cette fois sur le logarithme naturel (ln) de la probabilité de défaut. Plus notre terme de droite dans l'équation est élevé, plus  $P(\text{loan status} = 1)$  se rapproche de 1 et donc plus la probabilité que notre prêt soit en défaut est élevée. Au contraire, plus notre terme de droite est petit, plus  $P(\text{loan status}=1)$  se rapproche de 0 et donc plus la probabilité que notre prêt soit en défaut est basse. Ceci nous permet donc de déterminer l'impact direct des coefficients sur la probabilité de défaut. Une probabilité donnée nous permettrait ensuite de classer nos prêts hypothécaires en

catégories de prêt avec défaut probable (par exemple, tous prêts avec valeur de 0.5 et plus) puis prêt avec défaut peu probable (par exemple, tous prêts avec valeur inférieure à 0.5).

Le modèle logistique nous permet donc de mieux comprendre la probabilité de défaut que le modèle de régression linéaire. Nous devons par contre nous assurer que le modèle logistique est compatible avec l'ensemble de nos données. Dans une fonction logistique, la probabilité de défaut doit avoir une relation monotone avec chaque variable explicative afin de pouvoir être compatible. Si nous prenons par exemple la structure des échéances qui représente la relation entre le taux de défaut et le terme de l'hypothèque, von Furstenberg (1968-1970) démontre qu'il y a une hausse du taux de défaut dans les premières années suivie d'un déclin les années suivantes. Pour corriger une incompatibilité de la sorte, nous pourrions utiliser le terme de l'hypothèque au carré. De plus, ce modèle utilise seulement un horizon dans le temps fixe prédéterminé ce qui implique que pour considérer plusieurs variations du temps, nous devons modifier notre échantillon et estimer de nouveau nos coefficients ce qui n'est pas idéal. Par contre au fil des années, le modèle logistique demeure l'un des modèles les plus utilisés pour la modélisation du risque de crédit<sup>19</sup>. Par exemple, Kutty (1990) présente un modèle de régression logistique afin de prédire la probabilité de défaut de la dette des pays en développement. Westgaard et al (2001) utilise également le modèle afin de prédire la probabilité de défaut basé sur des variables financières. Gurny (2013) estime la probabilité de défaut des banques américaines.

---

<sup>19</sup> Zhang, Qingfen, "MODELING THE PROBABILITY OF MORTGAGE DEFAULT VIA LOGISTIC REGRESSION AND SURVIVAL ANALYSIS" (2015). Open Access Master's Theses. Paper 541. <https://digitalcommons.uri.edu/theses/541>

### 3. Modèle de l'analyse de survie

Le modèle de l'analyse de survie étudie la durée pour laquelle un prêt demeure dans un état donné. Il y a trois états possibles pour un prêt, c'est-à-dire à jour, en défaut ou prépayé. Un prêt débute toujours à jour, puis peut demeurer dans cet état ou peut changer d'état avec le temps. Ce modèle nous permet de déterminer la relation entre l'état, le temps et les autres variables explicatives afin d'estimer la durée pour laquelle un prêt demeure à jour.

$$h(t) = h_0(t)e^{(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)} \quad (3)$$

où  $h(t)$  représente le risque de défaut, décrit ici comme la probabilité conditionnelle qu'un prêt survive dans un état à jour jusqu'au temps  $t$ , puis changer d'état dans la période suivante  $t+1$ ,  $x$

$t$  représente le temps

$h_0(t)$  représente le risque de base, c'est-à-dire le risque lorsque toutes les variables sont à 0,

$X_1, X_2, \dots, X_k$  représentent les variables exogènes.

$\beta_1, \beta_2, \dots, \beta_k$  représentent les coefficients, qui établissent l'importance que chaque variable exogène peut avoir sur le risque de défaut

Pour un échantillon de données classé par durée du prêt, avec leur statut et les variables de temps, nous pouvons estimer la valeur des coefficients à l'aide de technique de maximum de vraisemblance. La variable de risque de base  $h_0(t)$  est commune à tous les prêts et ne change pas l'estimation des coefficients. Une fois les coefficients estimés, nous pouvons déterminer des probabilités de défaut.

Le modèle d'analyse de survie est couramment utilisé dans le domaine étant donnée la considération du temps dans la probabilité de défaut ce qui représente un avantage significatif comparativement au modèle logistique.

#### 4. Modèle d'optimisation

Le modèle d'optimisation permet de modéliser la dynamique économique lors du processus de défaut. Le modèle pose l'hypothèse que chaque emprunteur a comme objectif de maximiser sa richesse et son utilité ou minimiser ses coûts, et que ses choix (rester à jour avec ses paiements, refinancer ou faire défaut) sont toujours faits en fonction du meilleur rendement de cet objectif.

$$P_t(H_t, r_t) = \min[P_t^d(H_t, r_t), P_t^r(H_t, r_t), P_t^w(H_t, r_t)] \quad (4)$$

Où  $P_t$  représente le coût de l'emprunteur au temps  $t$ ,

$P_t^d$  représente le coût de l'emprunteur s'il choisit de faire défaut,

$P_t^r$ , s'il choisit de refinancer,

$P_t^w$ , s'il choisit de rester à jour avec ses paiements,

$H_t$  représente la valeur de la propriété,

$r_t$  représente le taux d'intérêt.

On utilise le modèle en générant des valeurs de propriété et taux d'intérêt à l'aide d'un processus stochastique, puis on définit les équations  $P_t^d$  (la valeur de la propriété ajoutée au coût de faire défaut),  $P_t^r$  (le paiement du prêt ajouté au solde impayé du prêt et le coût de refinancement) et  $P_t^w$  (le paiement du prêt ajouté au coût de demeurer à jour avec les paiements). L'équation est répétée pour chaque intervalle de temps  $t$ .



Ce modèle estime que la probabilité de défaut dépend de plusieurs facteurs économiques, comparativement aux trois premiers modèles qui estimaient que la probabilité de défaut dépendait des différentes variables explicatives. Le modèle peut alors incorporer l'aspect des différents cycles et conjoncture économique sans utiliser des données historiques.

### Modèles de probabilité de défaut pour portefeuille de prêts

Cette section aborde un portefeuille de prêt comme étant une donnée afin de déterminer un taux de défaut du portefeuille, comparativement aux quatre modèles précédents qui abordaient chaque prêt individuel comme étant une donnée afin de déterminer le risque de défaut.

#### 5. Modèle de régression linéaire pour le taux de défaut

Le modèle de régression linéaire cherche à établir une relation entre le taux de défaut et les variables explicatives, tel que vu dans le premier modèle de prêts individuels. Pour un portefeuille de prêts, le taux de défaut est calculé comme étant le nombre de prêts en défaut divisé par le nombre total de prêts. Nous interprétons ensuite le taux de défaut comme mesure déterminante du risque de défaut du portefeuille de prêts.

$$\text{default rate} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (5)$$

où :

$\alpha$  est une constante,

$X_1, X_2, \dots, X_k$  représentent les variables exogènes, qui permettent de déterminer le risque de défaut

$\beta_1, \beta_2, \dots, \beta_k$  représentent les coefficients, qui établit l'importance que chaque variable exogène peut avoir sur le risque de défaut

$\varepsilon$  est un terme d'erreur indépendant.

Puisqu'on utilise ici un portefeuille de prêts, nous devons déterminer comment intégrer nos variables exogènes relatives aux prêts individuels. Par exemple, pour intégrer le ratio prêt-valeur qui est une variable considérable dans l'analyse de défaut, nous pourrions utiliser une moyenne pondérée de chaque ratio individuel relatif au montant du prêt. Par ailleurs, nous pourrions classer les portefeuilles basés sur les caractéristiques des prêts individuels telles que le ratio prêt-valeur ou leur terme.

Ce modèle nous permet d'utiliser le taux de défaut comme mesure du risque de défaut, ce qui ne peut pas être fait avec le modèle de régression linéaire de prêts individuels. De plus, la construction de portefeuille nous permet de réduire le problème de colinéarité entre les variables explicatives de données non regroupées.

#### 6. Modèle de régression linéaire de log du ratio de probabilité.

Ce modèle se base d'abord sur l'hypothèse que la probabilité de défaut est une fonction des variables explicatives, tel que  $\pi = f(x)$ . Par contre, nous ne pouvons pas supposer une relation linéaire entre la variable dépendante et les variables explicatives lorsque la variable dépendante est une probabilité. Nous pouvons cependant utiliser le logarithme naturel du ratio de probabilité comme alternative afin de poser une équation linéaire en fonction de nos variables explicatives.

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (6)$$

où :

$X_1, X_2, \dots, X_k$  représentent les variables exogènes, qui permettent de déterminer le risque de défaut

$\beta_1, \beta_2, \dots, \beta_k$  représentent les coefficients, qui établit l'importance que chaque variable exogène peut avoir sur le risque de défaut

$\varepsilon$  est un terme d'erreur indépendant

$\pi$  représente la probabilité moyenne de défaut des prêts du portefeuille, ou le nombre de prêts en défaut par rapport au nombre de prêts total. Nous devons également déterminer comment intégrer nos variables explicatives, et pouvons le faire de la même manière qu'au modèle 5.

Ce dernier modèle est communément utilisé dans l'industrie pour modéliser les taux de défaut pour des portefeuilles de prêts.

### B. Taux de recouvrement et pertes en cas de défaut

Le taux de recouvrement et la perte en cas de défaut (PCD) représentent des aspects importants dans la gestion du risque de crédit. On calcule le taux de recouvrement en déterminant la valeur du principal et des intérêts accumulés pouvant être recouverts suivant le défaut, comme pourcentage de la valeur d'exposition en cas de défaut. On calcule la perte en cas de défaut comme étant 1 moins le taux de recouvrement.

Les recherches empiriques sur le sujet sont encore vagues pour plusieurs raisons. Il est difficile d'établir une définition homogène du moment précis d'un défaut avec autant de différents groupes d'actifs, de pays, d'industrie, etc. La majorité de la littérature existante concerne le secteur commercial étant donnée la disponibilité de l'information publique. L'information sur les prêts hypothécaires résidentiels est détenue par chaque banque et n'est donc pas aussi accessible. De plus, le processus de recouvrement après un défaut peut s'étendre sur une longue période de temps et

chaque pays fait face à des lois différentes sur les processus de recouvrement, ce qui complique la compilation de données à travers le temps. Cela étant dit, il existe deux définitions généralement acceptées pour déterminer la valeur considérée dans le taux de recouvrement :

- La valeur de l'actif immédiatement après qu'un défaut survienne (trading price recovery)
- La valeur du recouvrement considérant la valeur liée à une revente (ultimate recovery)

Dans le cas du marché hypothécaire, chaque définition présente des inconvénients. La première n'est pas toujours employable puisqu'un marché permettant d'établir une valeur n'est pas toujours disponible. Par contre, lorsqu'un tel marché est disponible, cette définition permet d'établir une valeur de recouvrement immédiatement après le défaut. La deuxième présente le désavantage d'être complexe, surtout dans une situation avec un marché stagnant et un actif illiquide. La récupération peut également être extrêmement longue et le montant doit être actualisé afin de refléter la perte de valeur dans le temps.

### Modèle utilisé pour la perte en cas défaut

Plusieurs recherches publiées portent sur le secteur commercial et leurs modèles ne seraient donc pas directement applicables à une étude de prêts hypothécaires résidentiels.

Leow et Mues ont cependant développé une approche intéressante pour les prêts hypothécaires résidentiels.<sup>20</sup> Leur approche est basée sur l'utilisation de deux modèles ou composantes. La première est un modèle de probabilité de reprise (revente de la propriété servant de collatéral) développé avec une régression logistique et la deuxième un modèle de décote en utilisant la régression des moindres carrés ordinaires. La décote représente ensuite le taux d'actualisation à appliquer au prix de revente estimé, supposant une reprise. Les deux modèles sont ensuite combinés afin de déterminer un pourcentage estimé de PCD. Ils démontrent que ces deux modèles combinés performant mieux qu'une approche avec un seul modèle des moindres carrés ordinaires, tel qu'utilisé dans plusieurs publications.<sup>21</sup>

### Modèle de probabilité de reprise

Ils utilisent ce modèle afin de prédire la probabilité qu'une propriété servant de collatéral pour un prêt hypothécaire en défaut soit revendue afin de compenser une portion de la perte. Leur méthodologie permet d'obtenir un modèle R1 avec quatre variables significatives (RPV initial, un indicateur binaire déterminant si le prêt a déjà été en défaut, l'âge du prêt ainsi que le type de propriété. Dans un modèle R2, ils remplacent le RPV initial et l'âge du prêt par le RPV au moment du défaut. Ils créent également un modèle R0 dans lequel ils utilisent comme seule variable explicative le RPV au moment du défaut afin de comparer les résultats avec R1 et R2.

---

<sup>20</sup> [https://ac.els-cdn.com/S0169207011000598/1-s2.0-S0169207011000598-main.pdf?\\_tid=220671f0-3a1d-4ad4-813e-03226dc91568&acdnat=1535481982\\_f54a2e40ffec7b38e9d91eebf5443b00](https://ac.els-cdn.com/S0169207011000598/1-s2.0-S0169207011000598-main.pdf?_tid=220671f0-3a1d-4ad4-813e-03226dc91568&acdnat=1535481982_f54a2e40ffec7b38e9d91eebf5443b00)

<sup>21</sup> <https://www.sciencedirect.com/science/article/pii/S0169207013000447>

Leurs mesures de performance leur permettent de déterminer que le modèle R2 utilisant les 4 variables explicatives incluant le RPV de défaut est celui qui performe le mieux. Il sera donc inclus dans la suite de leur démarche.

### Modèle de décote

Ce modèle s'applique ensuite seulement aux observations qui ont été revendues. La décote est donc définie comme étant le ratio du prix de revente sur la valeur de la propriété au moment du défaut. Leur méthodologie permet d'obtenir un modèle H1 avec sept variables significatives (RPV initial, l'indicateur binaire de défaut précédent, l'âge du prêt, ratio de valeur de propriété et moyenne régionale, type de propriété, âge de la propriété et région). Dans un modèle H2, ils remplacent le RPV initial et l'âge du prêt par le RPV au moment du défaut.

Leurs mesures de performance leur permettent de déterminer que le modèle H1 est celui qui performe le mieux. Ils estiment ensuite l'écart type de la décote afin d'en déterminer la variabilité.

### Modèle combiné

Ils débutent par déterminer la probabilité de reprise pour ensuite appliquer le modèle de décote et d'écart type à chaque observation afin d'obtenir  $H_j$  et  $\sigma_j$ . Ils estiment ensuite la distribution de chaque décote avec une distribution normale. Si la valeur de la décote est plus grande que le RPV au moment du défaut, alors nous pouvons supposer que la valeur de revente couvrirait le montant du prêt et qu'il n'y aurait donc pas de pertes. Nous pouvons donc conclure :

$$E(\text{shortfall percent} \mid \text{repossession}) = \int_{-\infty}^{\text{DLTV}} p(h)(\text{DLTV} - h)dh \quad (7)$$

où

Shortfall percent : pourcentage de perte

Repossession : valeur de revente

$p(h)$ : la fonction de densité de probabilité de  $h$

Puis,

$$E(\text{loss} \mid \text{default}) = \left[ \begin{array}{l} E(\text{shortfall percent} \mid \text{repossession}) \\ \times \text{indexed valuation} \\ \times P(\text{Repossession} \mid \text{default}) \end{array} \right] + [c \times (1 - P(\text{Repossession} \mid \text{default}))] \quad (8)$$

Ils calculent la PCD en divisant la perte attendue (Expected LGD démontré ci-haut à la formule 8) par le solde estimé du prêt au moment du défaut.

### C. Exposition en cas de défaut

L'exposition en cas de défaut indique à combien peut s'élever l'exposition du prêt, et donc le solde impayé. Pour une majeure partie de produits de crédit, l'ECD peut s'avérer assez difficile à modéliser. Pour des prêts hypothécaires résidentiels, le calcul est cependant assez simple étant donnée la nature prédéterminée de l'échéancier de remboursement du prêt. Afin de déterminer l'exposition, nous pourrions simplement débiter avec le montant initial du prêt, et en déterminer le solde pour les différentes valeurs de  $t$ .

## CHAPITRE 3 : MÉTHODOLOGIE

Pour la section empirique de ce projet de mémoire, nous avons tenté d'étudier la probabilité de défaut et les variables qui l'affectent, considérant les données publiques que nous étions en mesure d'obtenir. Nous avons débuté avec le modèle logistique pour déterminer s'il est optimal dans le contexte de la probabilité de défaut

étant donné que le modèle propose de tester la régression dont la variable expliquée est dichotomique (à jour ou en défaut).

$$P(\text{loan status} = 1) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (2)$$

Ensuite, nous avons testé le modèle de régression linéaire de portefeuille en utilisant les moyennes de nos données afin de déterminer s'il nous permet de tirer de meilleures conclusions que le modèle logistique.

$$\text{default risk} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (5)$$

### *Données*

Le BSIF indique que chaque institution financière qui souscrit des prêts hypothécaires a la responsabilité de respecter certaines lignes directrices en matière de souscription prudentielle. Il énonce trois principes clés portant sur la décision en matière de crédit hypothécaire résidentiel et sur le processus de souscription. Le premier porte sur l'identité de l'emprunteur, ses antécédents et sa volonté d'acquitter ses dettes dans les délais impartis. Le deuxième porte sur la capacité de l'emprunteur d'acquitter ses dettes dans les délais impartis. Le troisième porte sur la valeur ou la sûreté sous-jacente du bien immobilier et le processus de gestion de cette sûreté. Afin d'avoir une analyse complète de la probabilité de défaut, il serait idéal d'avoir accès à une banque de données qui nous permettrait de bien capturer ces trois principes clés dont les banques doivent tenir compte lors de leur souscription. Nous pourrions donc utiliser les caractéristiques de l'emprunteur telles que l'âge, la race, le sexe, l'état civil, puis sa cote de crédit pour bien capturer le premier principe. Pour capturer le deuxième principe, nous pourrions utiliser le ratio d'amortissement brut de la dette (ABD), le ratio d'amortissement total de la dette (ATD), le ratio prêt-



valeur (RPV) et le taux d'intérêt. Pour capturer le troisième principe, nous pourrions utiliser la valeur de la propriété, puis finalement inclure des données macro-économiques telles que le taux de chômage et le taux directeur de la banque du Canada.

Les lois d'accès à l'information ne nous ont par contre pas permis d'accéder à toute cette information au Canada. En effet, seulement les données agrégées très limitées existent, et nous avons malheureusement seulement pu obtenir certaines données recherchées d'une base de données publique de Fannie Mae aux États-Unis. Idéalement, nous aurions eu accès à un programme spécialisé pour le traitement des données tel que SAS ou R afin de pouvoir maximiser le potentiel de notre analyse. Due à des limitations de ressources, ceci n'a pas été possible et nous avons dû utiliser Excel qui n'offre pas la même capacité, puissance et flexibilité. Notre échantillon a donc été manipulé de manière à le rendre optimal pour Excel. Ceci nous a donné un échantillon de 425 674 données mensuelles avec (1) le statut des prêts hypothécaires à jour ou en défaut, (2) le taux d'intérêt à l'initiation, (3) le solde impayé ainsi que (4) la maturité ou l'âge du prêt. Ces données étaient disponibles de 2009 à 2018. Nous y avons ajouté (5) le taux de chômage du Canada également disponible sur une base mensuelle afin d'inclure une composante macro-économique reliée au Canada étant donné que notre étude et sujet de ce mémoire porte sur le risque de crédit du marché hypothécaire canadien. Les données utilisées sont les meilleures alternatives étant donnée la disponibilité de l'information. Nous avons donc déterminé s'il était possible de déterminer des corrélations ou autres conclusions à partir des données trouvées.

### *Limitations*

Puisque nous avons dû manipuler nos données afin de les rendre compatibles avec les capacités d'Excel, nous avons quelques limitations qui doivent être soulignées étant donnée leur importance dans la considération des résultats de notre analyse.

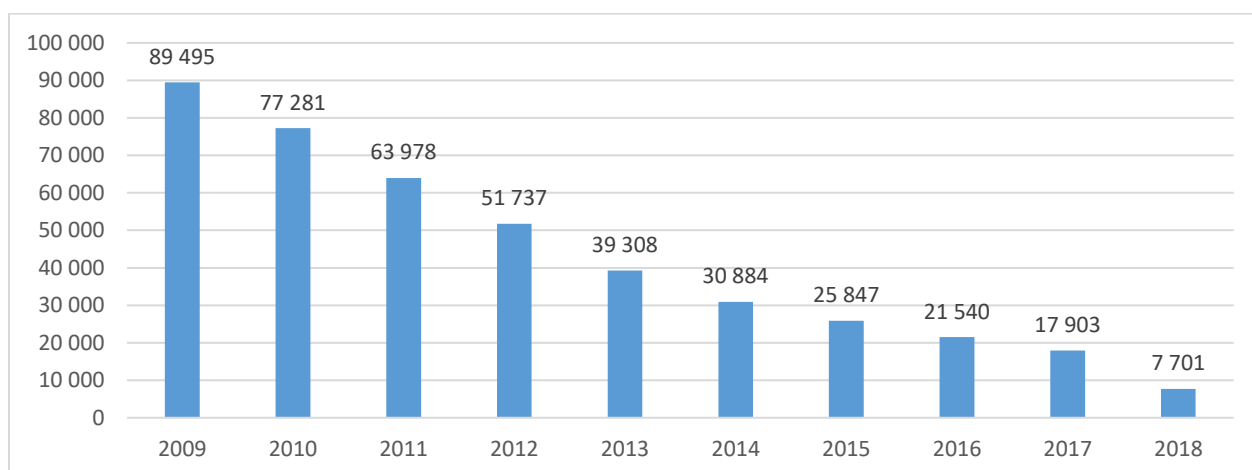
- Notre revue littéraire nous permet de conclure que la plupart des études s'entendent pour affirmer que le ratio prêt-valeur ainsi que certains autres facteurs reliés à l'emprunteur tels que sa cote de crédit, son revenu et taux d'endettement jouent un rôle clé dans la détermination de la probabilité de défaut. Puisque ces données ne nous sont pas accessibles avec la limitation de nos ressources, nous compromettons fort probablement notre habilité à développer un modèle nous permettant d'expliquer nos résultats.
- Nous avons initialement environ 1 million d'observations dans notre fichier pour un total de 8870 prêts qui incluait de l'information détaillée sur le statut du prêt, c'est-à-dire à jour ou en retard de moins de 30 jours, en retard de 30-59 jours, en retard de 60-89 jours, en retard de 90-119 jours et ainsi de suite. Pour en faciliter la manipulation, nous avons seulement gardé les données avec un statut à jour ou en retard de moins de 30 jours et en retard de 30-59 jours. Ceci nous a laissés avec 425 674 données. Nous avons pu conclure que des 8870 prêts, seulement 229 prêts ou 2.58% sont tombés en défaut de 30-59 jours puis sont revenus à jour le mois suivant. Le reste des prêts qui sont tombés en défaut y sont restés pour plus 2 mois. Nous n'avons pas été capables de déterminer combien de prêts des 8870 sont tombés en défaut et combien sont toujours restés à jour de 2009-2018. Pour notre analyse, nous avons supposé que les prêts en défaut de 30-59 jours sont considérés comme un bon indicatif d'un défaut pour un modèle de crédit.
- Nous n'avons pas pu garder les identifiants uniques associés à chaque observation qui nous auraient permis de suivre l'évolution mensuelle de

chacun des 8870 prêts de 2009-2018. Chaque observation est donc traitée comme étant une observation individuelle ou un prêt unique. Nous ne pouvons pas suivre l'évolution de chaque prêt individuel avec considération de ses caractéristiques telles que le montant initial du prêt et le taux d'intérêt. De plus, nos observations suivent la performance des prêts initiés en 2004, de 2009 à 2018. Nous avons débuté notre échantillon en 2009 pour limiter le nombre de données à traiter. Nous omettons donc la performance de ces prêts de 2004 à 2009, ce qui pourrait être une période critique dans l'analyse de défaut. De plus, nous débutons notre analyse en 2009 qui se trouve à être à la fin d'une crise économique et période de récession aux États-Unis. Elle est fort probablement non représentative d'une moyenne stable long terme de conditions économiques.

#### *Analyse sommaire de l'échantillon de données*

Ce premier graphique démontre le nombre total d'observations inclus dans notre échantillon. En regardant cette distribution, nous pouvons conclure que notre échantillon de données représente des prêts initiés en 2009 puis leur suivi de performance au fil des années. Le nombre d'observations diminue puisque le nombre de prêts actifs est réduit au fil des années lorsque les prêts sont refinancés, repayés ou tombent en défaut pour plus de 60 jours.

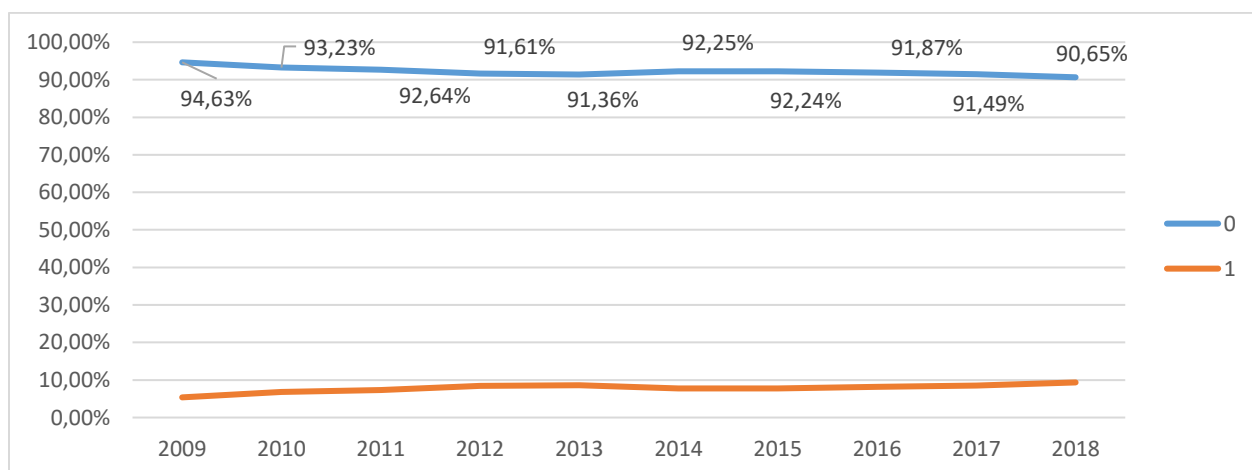
Graphique A – Total d’observations pour chaque année



1. Le statut des prêts hypothécaires à jour ou en retard de moins de 30 jours (0) ou en défaut (1)

Le graphique suivant représente la proportion des prêts à jour par rapport à ceux en défaut pour la totalité de notre échantillon de 425 674 données mensuelles. Nous pouvons remarquer que la tendance se maintient entre 90-95% de 2009 à 2018.

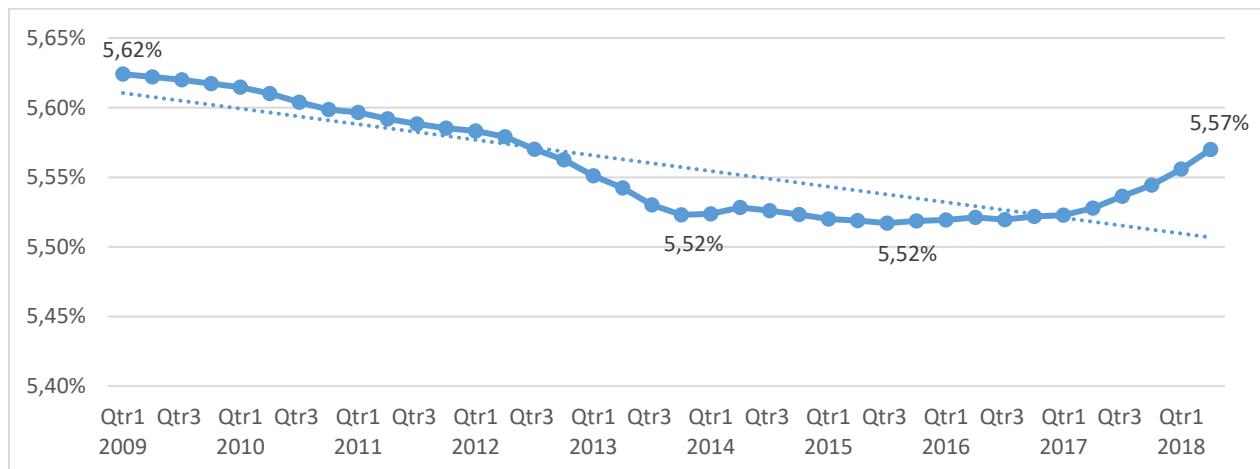
Graphique B – Proportion des prêts à jour



## 2. Taux d'intérêt à l'initiation

Le graphique suivant représente la moyenne du taux d'intérêt à l'initiation des prêts inclus dans notre échantillon. Nous pouvons remarquer qu'il y a une tendance à la baisse de 2009 à 2018 avec le taux moyen débutant à 5.62% pour atteindre au taux plancher de 5.52% fin 2013 à fin 2015, pour ensuite remonter à 5.57% en 2018. Ceci pourrait être dû au fait que ce sont les prêts avec des taux d'intérêt plus bas à l'initiation en 2004 qui sont restés à jour jusqu'en 2018. La variation est par contre minime sur une moyenne pour la taille de notre échantillon.

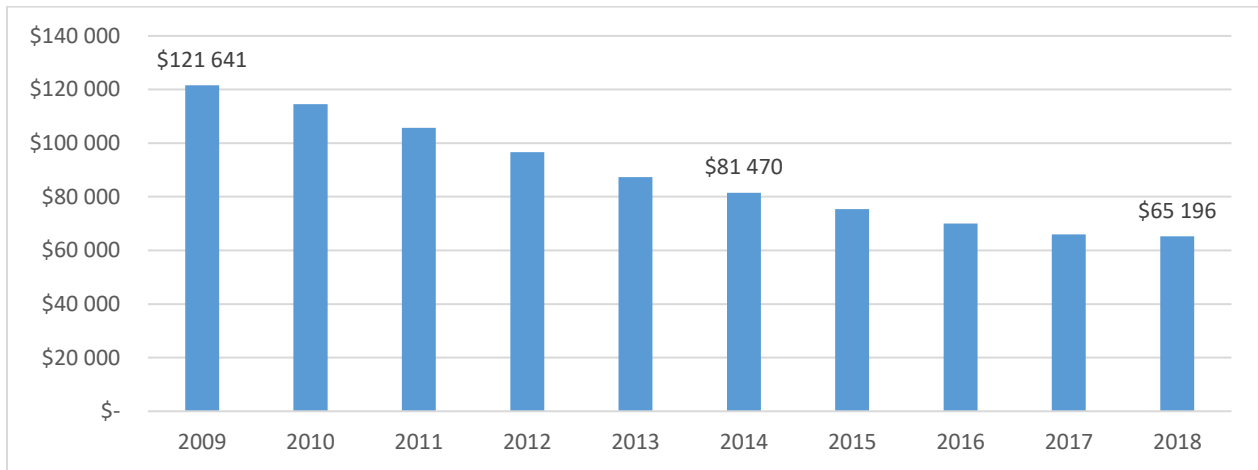
Graphique C – Moyenne du taux d'intérêt à l'initiation



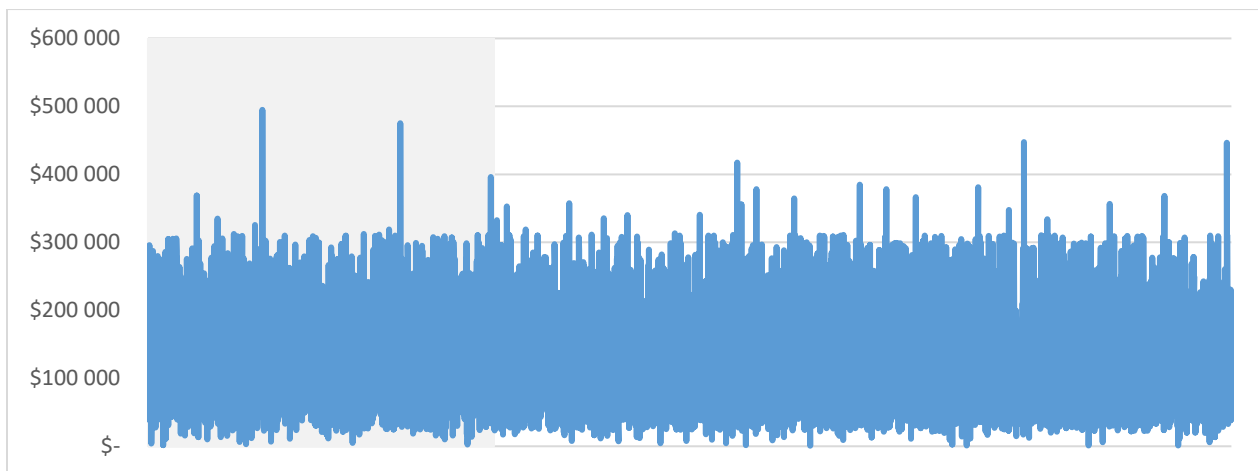
## 3. Solde impayé

Le premier graphique ci-dessous représente la moyenne du solde impayé de nos observations au fil du temps. Puisque nous suivons l'évolution, il est normal que le solde diminue. Le deuxième graphique représente la distribution du solde initial des prêts en janvier 2009. On remarque que la majorité des prêts sont sous 300,000\$ et donc notre moyenne du premier graphique semble bien représenter notre échantillon.

Graphique D – Moyenne du solde impayé des observations



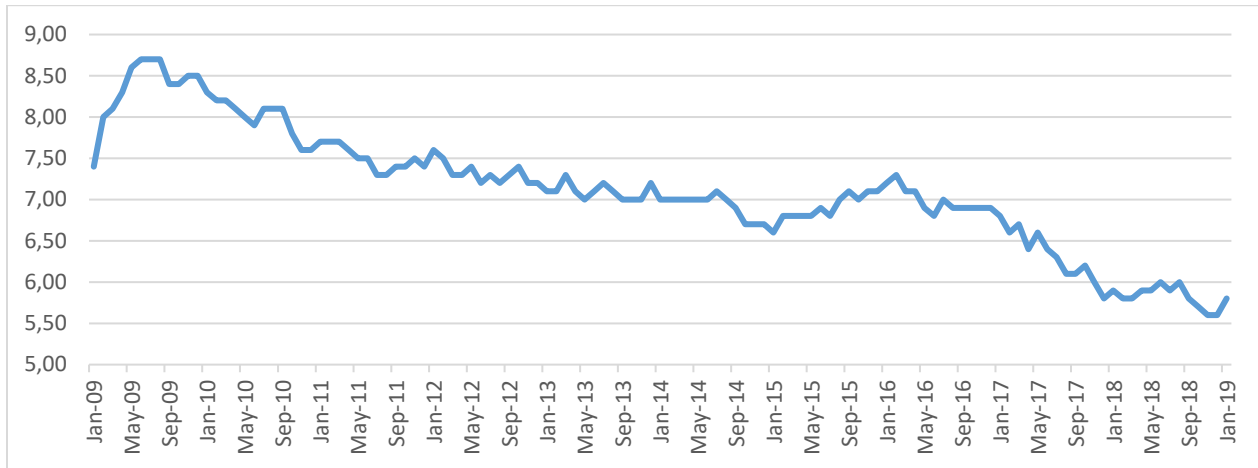
Graphique E – Distribution des soldes initiaux des observations en janvier 2009



#### 4. Taux de chômage au Canada

Le graphique suivant démontre le taux de chômage au Canada de 2009 à 2019. En regardant cette distribution, nous pouvons conclure que le taux de chômage a une tendance à la baisse sur la période de temps analysée. Étant donné la crise économique de 2009 et la période de temps analysée, il est possible que notre échantillon représente seulement une partie d'un cycle économique.

Graphique F – Taux de chômage au Canada de 2009 à 2019



## CHAPITRE 4 : ANALYSE ET RÉSULTATS

Tel que démontré plus haut, nous avons donc 425 674 données comportant le statut du prêt (à jour ou en défaut) de 2009 à 2018 à travers les États-Unis comme variable expliquée  $Y$  puis le taux d'intérêt  $X_1$ , le solde impayé  $X_2$ , l'âge du prêt  $X_3$  et le taux de chômage  $X_4$  comme variables explicatives que nous avons utilisées dans nos modèles ci-bas.

### *Modèle de régression logistique*

De la même manière que montré dans le projet de mémoire «Modeling the Probability of Mortgage Default via Logistic Regression and Survival Analysis» de Zhang, Qingfen (2015)<sup>22</sup> ainsi que plusieurs recherches citées dans la revue du modèle logistique plus haut, nous avons débuté notre analyse avec le modèle logistique puisqu'il est couramment utilisé pour modéliser la probabilité de défaut ainsi que parce qu'il nous permet de tester une régression dont la variable expliquée est dichotomique.

<sup>22</sup> Zhang, Qingfen, "MODELING THE PROBABILITY OF MORTGAGE DEFAULT VIA LOGISTIC REGRESSION AND SURVIVAL ANALYSIS" (2015). Open Access Master's Theses. Paper 541. <https://digitalcommons.uri.edu/theses/541>

$$P(\text{loan status} = 1) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (2)$$

Pour ce faire, nous avons utilisé la transformation avec le maximum de vraisemblance. Un logiciel statistique nous aurait permis d'estimer un logit directement, mais puisque nous n'en avons pas nous avons utilisé des manipulations manuelles dans Excel. Malheureusement, nos coefficients déterminés à partir du Solver étaient nuls ou très proches de 0 ce qui nous empêchait de pousser notre analyse du modèle logistique plus loin. Nous avons donc décidé d'analyser nos variables plus en profondeur afin de déterminer où pourrait être notre problème, en débutant par une représentation graphique des variables dans le temps. Nous avons pu en conclure que c'est dû à des limitations dans notre manipulation de données que notre modèle ne fonctionnait pas. Dans la manipulation de nos données, nous avons dû enlever les identifiants uniques qui nous permettaient de suivre chaque prêt au fil du temps. Ceci a pour effet que chaque observation est traitée comme étant un prêt individuel lorsqu'en réalité ils ne le sont pas. L'ensemble de nos données sont en fait des blocs d'observations en ordre chronologique pour chacun des 8870 prêts ce qui rend notre modèle utilisant les données individuelles invalides. Voir l'annexe A pour tous les détails de la méthodologie et les résultats. Pour régler le problème, nous avons décidé de regrouper nos données en moyenne et tester le modèle de régression linéaire de portefeuille.

#### *Modèle de régression linéaire de portefeuille*

Au lieu d'utiliser les données de notre échantillon de manière individuelle comme démontré dans la première méthode, nous avons utilisé une approche de portefeuille qui consiste à grouper nos données. Pour ce faire, nous avons regroupé nos données par mois.

$$\text{default rate} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (5)$$



Au lieu d'utiliser le risque de défaut comme variable expliquée (*default risk*), nous avons utilisé un taux de défaut (*default rate*) déterminé en utilisant le nombre de défauts observés divisé par le nombre d'observations total sur une base mensuelle ce qui nous donnait en moyenne un taux de défaut de 7.73%. Ceci adresse notre contrainte de variable expliquée binaire qui nous limitait dans nos choix de modèles lorsqu'on traitait les données individuellement. Ensuite, nous avons déterminé les moyennes de chacune de nos variables explicatives toujours sur une base mensuelle, sauf pour le taux de chômage qui était déjà une moyenne de base mensuelle. La moyenne totale du taux d'intérêt était de 5.56%, la moyenne totale du solde impayé était de 89,508\$, la moyenne de l'âge du prêt était de 116 mois et la moyenne du taux de chômage était de 7.47%. En regroupant nos données de cette manière, nous avons pu diminuer nos observations à 114, c'est-à-dire une par mois de janvier 2009 à juin 2018. Nous avons ensuite tenté de déterminer les valeurs des coefficients  $\beta_k$  avec la méthode des moindres carrés du Solver d'Excel. Voici nos résultats :

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.871530023							
R Square	0.759564582							
Adjusted R Square	0.750741263							
Standard Error	0.005697406							
Observations	114							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	4	0.011177561	0.002794	86.08605	7.79368E-33			
Residual	109	0.003538187	3.25E-05					
Total	113	0.014715748						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-3.352176101	0.37468826	-8.94657	1.07E-14	-4.094796052	-2.60955615	-4.094796052	-2.60955615
Taux d'intérêt	0.760222742	0.081192654	9.363196	1.21E-15	0.599301543	0.921143942	0.599301543	0.921143942
Solde impayé	-6.06544E-06	5.86342E-07	-10.3445	6.95E-18	-7.22755E-06	-4.90332E-06	-7.22755E-06	-4.90332E-06
Âge du prêt	-0.002475905	0.000267878	-9.24265	2.28E-15	-0.003006831	-0.001944979	-0.003006831	-0.001944979
Taux de chômage	0.004708275	0.001522415	3.092636	0.00252	0.001690899	0.007725652	0.001690899	0.007725652

## *Discussion*

Nous avons d'abord débuté en regardant le coefficient de corrélation  $R$  ainsi que le coefficient de détermination  $R^2$ . Notre coefficient de corrélation nous donne une idée de la relation linéaire entre notre variable expliquée et nos variables explicatives, 1 étant une relation positive parfaite. Notre coefficient de détermination nous permet de déterminer en pourcentage quelle variation de notre statut du prêt est expliquée par les variables explicatives utilisées. Le plus près de 100% le mieux. En regardant le coefficient de corrélation, on remarque qu'il nous donne une valeur de 0.87 puis le coefficient de détermination une valeur de 75%. Ceci nous permet de conclure à première vue que notre modèle est relativement acceptable. La relation entre nos variables explicatives et notre variable expliquée est de 0.87 ce qui est assez fort, puis 75% de nos valeurs sont expliquées par nos variables explicatives. Notre erreur type est de 0.0057 ce qui nous informe de la distance moyenne entre nos points et la ligne de régression. Ensuite nous pouvons regarder les valeurs de nos coefficients.

- Le taux d'intérêt a un coefficient positif de 0.76 qui nous indique l'ampleur du changement moyen pour une augmentation d'une unité du taux d'intérêt. Intuitivement, il semble adéquat que notre modèle nous indique que plus le taux d'intérêt augmente, plus le taux de défaut augmente puisque théoriquement le taux d'intérêt augmenterait les paiements et serait initialement déterminé en partie sur la cote de crédit d'un emprunteur. Un emprunteur de moindre qualité avec une probabilité de défaut plus élevée aurait un taux d'intérêt plus élevé.
- Le solde impayé a un coefficient négatif de 6.07. Intuitivement, il ne semble pas adéquat que notre modèle nous indique que plus le solde impayé augmente, plus le taux de défaut diminue puisque plus le solde est élevé plus les paiements sont élevés. Par contre, on pourrait peut-être considérer que plus

le solde est élevé, plus les emprunteurs avaient de bons revenus stables et plus les prêts étaient de meilleures classes ou qualité. Toute chose égale par ailleurs, la relation devrait cependant tout de même être positive. La relation est également très forte, c'est-à-dire 6.07 changements d'unité moyenne à la baisse du taux de défaut pour 1 changement d'unité du solde impayé à la hausse ce qui semble être un problème.

- L'âge du prêt a un coefficient négatif de 0.0024. Intuitivement, il semble adéquat que notre modèle nous indique que plus l'âge du prêt augmente, plus le taux de défaut diminue puisque les prêts font en général défaut dans les premières années de leur initiation. De plus, le solde diminuerait sur un prêt plus l'âge du prêt augmente et donc le taux de défaut diminuerait théoriquement en conséquence. La relation est cependant très faible à 0.0024 ce qui indique que l'âge n'aurait pas un gros impact, ce qui est possible.
- Le taux de chômage a un coefficient positif de 0.0047. Intuitivement, il semble adéquat que notre modèle nous indique que plus le taux de chômage augmente, plus le taux de défaut augmente puisque le nombre d'emprunteurs sans emploi et donc possiblement dans l'incapacité de payer leur prêt augmente également. La relation est cependant très faible à 0.0047 ce qui indique que le taux de chômage n'aurait pas un gros impact ce qui semble étonnant.

### *Analyse de la variance*

Nous pouvons maintenant faire une analyse des variances et tester l'égalité de nos variables avec le test de Fisher.

Nos hypothèses seront donc les suivantes :

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \beta_k \neq 0, \text{ pour au moins une valeur de } k$$

Le tableau ANOVA dans Excel nous indique déjà nos statistiques clés. Nous pouvons voir que notre  $F=86.09$  et que notre valeur de  $p$  est très près de 0. Puisque  $0 < 0.05$ , nous pouvons rejeter notre  $H_0$  et conclure qu'au moins l'un de nos coefficients est différent de 0. Nous pouvons également regarder nos t-stat directement et voir que tous nos paramètres sont significativement différents de 0. Par contre, nos t-stats semblent très élevées.

Nous avons alors choisi d'évaluer la présence d'autocorrélation dans les résidus de la régression linéaire. Pour ce faire, nous avons utilisé le test de Durbin-Watson et la formule suivante :

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

où  $e$  représente les résidus ou la différence entre le taux de défaut prédit par notre régression et le taux de défaut observé. La somme de la différence au carré des résidus de nos données nous donne 0.003439 puis la somme des résidus au carré nous donne 0.003538. Notre statistique  $d$  équivaut donc à 0.972066. Afin de déterminer s'il y a présence d'autocorrélation, nous avons utilisé les hypothèses suivantes :

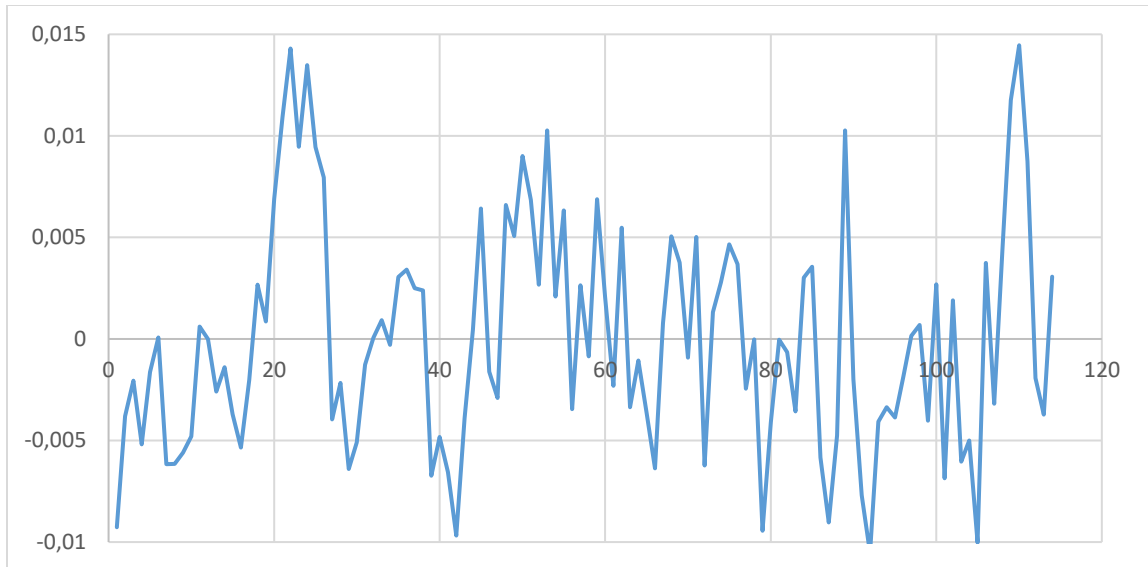
$$H_0 : \rho \leq 0$$

$$H_1 : \rho > 0$$

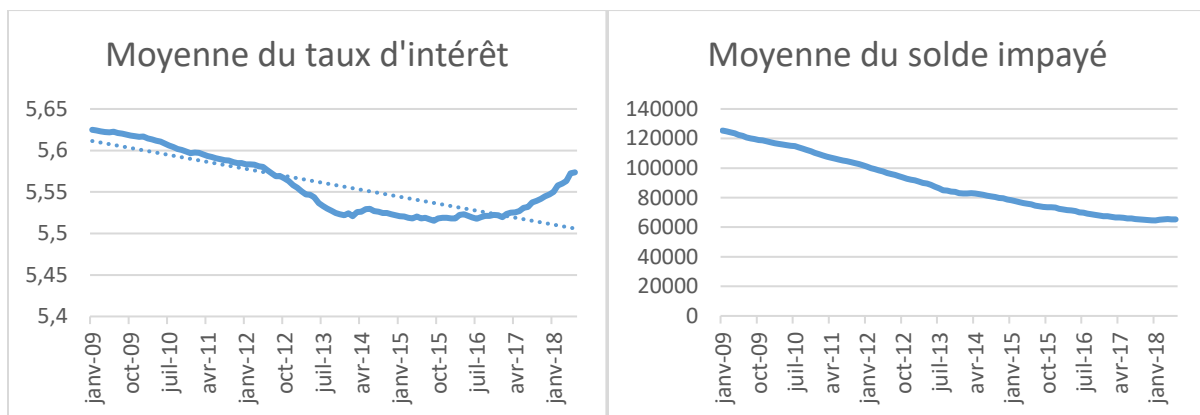
Si  $d < d_L$ , alors on rejette  $H_0 : \rho \leq 0$

Si  $d > d_U$ , alors on ne rejette pas  $H_0$

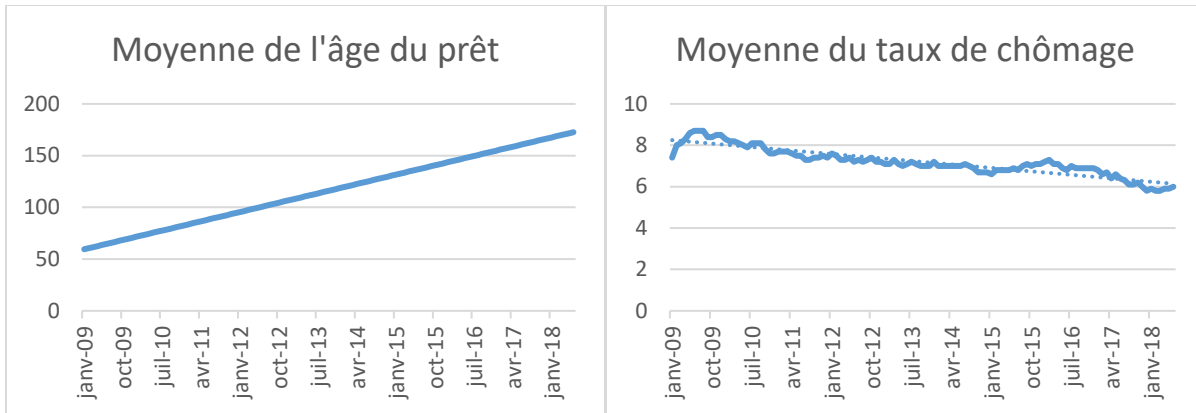
D'après la table de Durbin-Watson en utilisant un niveau d'alpha de 0.05, notre  $k=4$  et notre  $n=114$ , nous obtenons un  $d_L$  de 1.461 et un  $d_U$  de 1.625. Puisque notre  $d < d_L$ , nous avons rejeté l'hypothèse nulle et conclut qu'il y a présence d'autocorrélation positive. Le graphique suivant est une représentation des résidus. On peut également observer graphiquement la présence d'autocorrélation dans nos résidus.



Il y a une règle générale qui dit que si notre coefficient de détermination  $R^2$  obtenu dans la régression est plus grand que la statistique DW, alors nous pouvons être en présence d'une série non stationnaire<sup>23</sup>. Ce n'est pas le cas ici, mais nous avons tout de même choisi de représenter nos variables en fonction du temps afin d'évaluer visuellement la stationnarité étant donné nos tstat élevées. Nous remarquons clairement que la stationnarité n'est pas respectée, ce qui peut peut-être être réglé en utilisant la première différence.

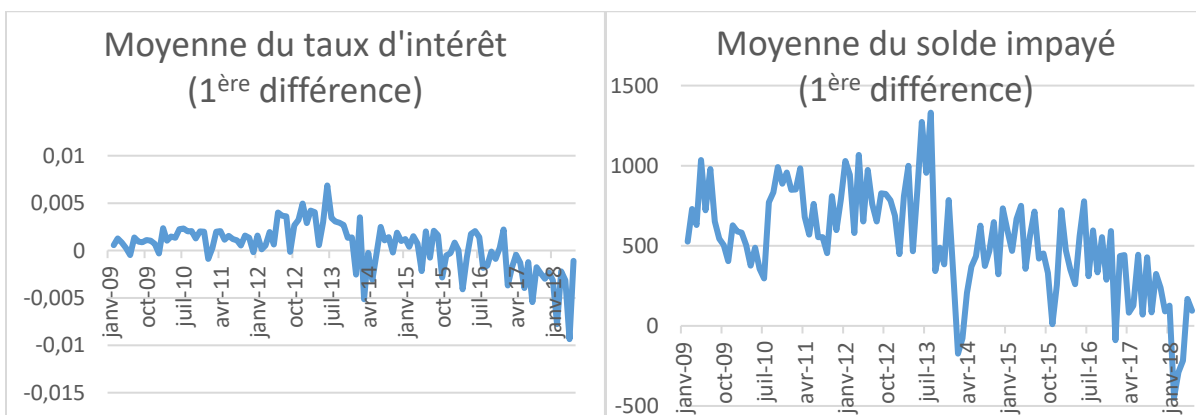


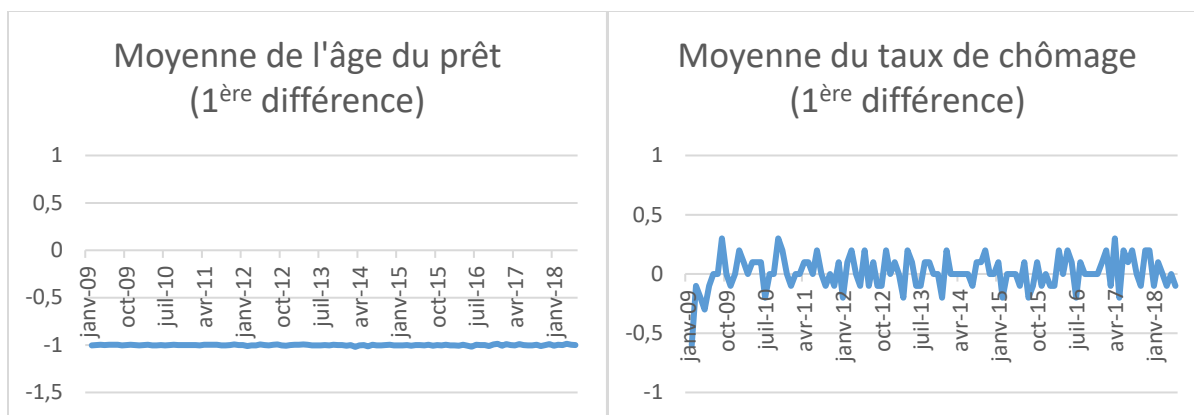
<sup>23</sup> <http://cruncheconometrix.blogspot.com/2018/02/how-to-perform-stationarity-test-in.html>



Avant d'évaluer et corriger notre problème de stationnarité, nous avons également fait une série de tests d'hypothèse sur nos données qui doivent tenir dans la régression linéaire multiple, c'est-à-dire tester la relation linéaire entre chaque variable explicative et la variable expliquée, tester la multicolinéarité entre variables explicatives et tester si nos résidus suivaient une loi normale avec une moyenne de 0. Les résultats de ces tests se trouvent en annexe B.

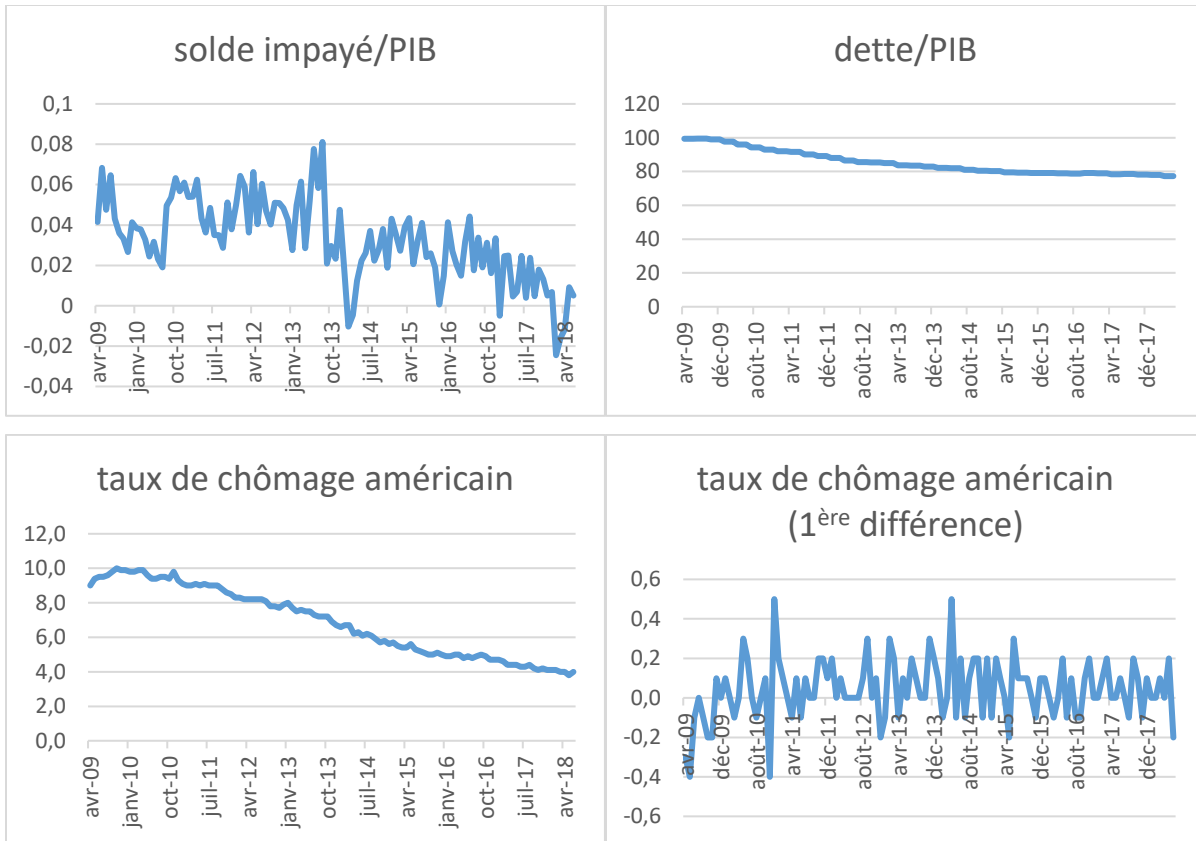
Afin de procéder avec notre analyse et régler le problème de stationnarité, nous avons transformé les données en utilisant la première différence. Voici nos nouveaux graphiques en fonction du temps:





On ne voit plus clairement de problème de stationnarité, par contre nous avons choisi d'enlever la variable moyenne de l'âge du prêt puisqu'elle est maintenant constante. Nous avons donc refait notre régression avec la première différence de la moyenne du taux d'intérêt, du solde impayé et du taux de chômage.

Nos résultats étaient assez faibles donc nous avons choisi de refaire l'exercice avec la deuxième et troisième différence, puis de combiner la première et troisième différence étant donné que ceux-ci avaient des résultats plus élevés que la deuxième différence. Voir résultats en annexe C. Afin de tenter d'améliorer notre modèle pour une dernière fois, nous avons ensuite décidé d'ajouter à notre analyse une variable d'endettement des ménages divisé par le PIB aux États-Unis, le solde impayé du prêt divisé par le PIB, puis remplacer notre taux de chômage du Canada par le taux de chômage aux États-Unis. Nous avons finalement gardé le solde impayé divisé par le PIB et la première différence du taux de chômage américain étant donnée la non-stationnarité du taux de chômage en niveau. Nous avons exclu la dette divisée par le PIB étant donnée la non-stationnarité et le fait que puisque nos variables étaient trimestrielles étendue afin d'être transformée en variables mensuelles, la première différence ne fonctionnait pas.



Après avoir fait plusieurs tests pour évaluer nos coefficients et nos t-stats les plus significatifs, nos variables pour notre modèle final sont donc la moyenne du solde impayé en première différence, la moyenne du taux d'intérêt en première différence, le solde impayé divisé par le PIB américain puis le taux de chômage américain en première différence. Voici les résultats de notre régression :



SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.561406							
R Square	0.315177							
Adjusted R Square	0.289813							
Standard Error	0.009227							
Observations	113							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>gnificance F</i>			
Regression	4	0.004232	0.001058	12.42625	2.38E-08			
Residual	108	0.009195	8.51E-05					
Total	112	0.013426						
	<i>Coefficient</i>	<i>Standard Err</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>lower 95.0%</i>	<i>pper 95.0%</i>
Intercept	0.079036	0.001946	40.60989	3.04E-67	0.075179	0.082894	0.075179	0.082894
Moyenne du solde impayé (1ère différence)	0.000193	3.07E-05	6.291652	6.86E-09	0.000132	0.000254	0.000132	0.000254
Solde impayé/PIB américain	-3.20383	0.493521	-6.49179	2.66E-09	-4.18208	-2.22559	-4.18208	-2.22559
Moyenne du taux d'intérêt (1ère différence)	-0.06352	0.436583	-0.14549	0.884592	-0.9289	0.801863	-0.9289	0.801863
Taux de chômage américain (1ère différence)	0.002631	0.002227	1.181661	0.239936	-0.00178	0.007045	-0.00178	0.007045

### Discussion

En regardant le coefficient de corrélation, on remarque qu'il nous donne une valeur de 0.56 puis le coefficient de détermination une valeur de 31%. Ce ne sont pas des résultats très conclusifs pour une régression linéaire. La relation entre nos variables explicatives et notre variable expliquée est assez faible à 0.56, puis seulement 31% de nos valeurs sont expliquées par nos variables explicatives. Nos coefficients sont assez petits, sauf pour le solde impayé divisé par le PIB. Finalement nos t-stats sont toutes plus grandes que 1.98 à l'exception du taux d'intérêt, mais les statistiques plus élevées semblent tout de même encore problématiques. On peut en conclure que notre modèle n'est pas idéal. D'après la revue littéraire et les modèles analysés, nous pouvons conclure qu'il manque fort probablement de variables liées au prêt telles que le ratio prêt valeur, la cote de crédit de l'emprunteur, son ratio d'endettement, etc.

Nous avons tout de même décidé de conclure la partie empirique de ce projet avec une analyse de quelques hypothèses de régression linéaire, c'est-à-dire

l'autocorrélation dans les résidus en utilisant la statistique de Durbin-Watson ainsi que de la normalité des résidus en utilisant la statistique de Jarque Béra.

Le test de Durbin-Watson utilise la formule suivante :

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

où  $e$  représente les résidus ou la différence entre le taux de défaut prédit par notre régression et le taux de défaut observé. La somme de la différence au carré des résidus de nos données nous donne 0.009195 puis la somme des résidus au carré nous donne 0.004554. Notre statistique  $d$  équivaut donc à 0.4953. Afin de déterminer s'il y a présence d'autocorrélation, nous avons utilisé les hypothèses suivantes :

$$H_0 : \rho \leq 0$$

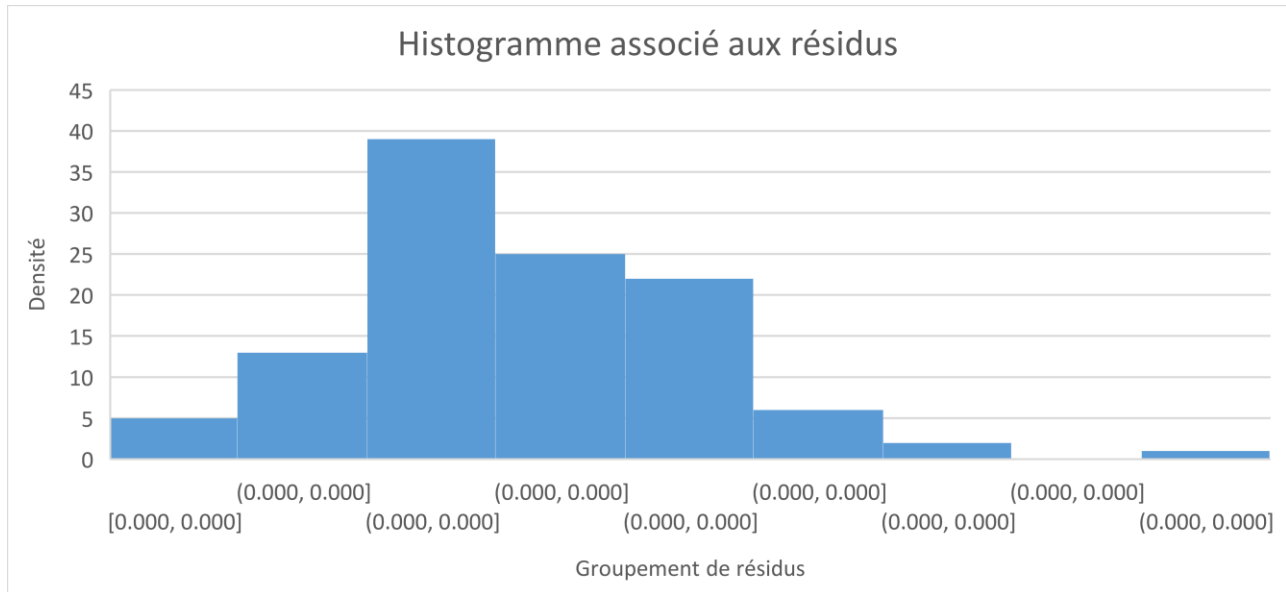
$$H_1 : \rho > 0$$

Si  $d < d_L$ , alors on rejette  $H_0 : \rho \leq 0$

Si  $d > d_U$ , alors on ne rejette pas  $H_0$

D'après la table de Durbin-Watson en utilisant un niveau d'alpha de 0.05, notre  $k=4$  et notre  $n=113$ , nous obtenons un  $d_L$  de 1.461 et un  $d_U$  de 1.625. Puisque notre  $d < d_L$ , alors on rejette l'hypothèse nulle et on conclut qu'il y a présence d'autocorrélation positive ce qui pourrait être un bon signe que nous avons omis des variables importantes (tel que celles liées à l'emprunteur).

Ensuite, les résidus devraient être distribués selon une loi normale de moyenne 0. Nous pouvons d'abord évaluer la normalité avec une représentation graphique des résidus. En regardant le graphique ci-bas, il serait possible que nous puissions conclure que nos résidus suivent une loi normale.



Nous avons donc testé cette hypothèse avec un test de normalité en utilisant la statistique Jarque Béra qui suit asymptotiquement une loi  $\chi^2$  à 2 degrés de liberté, où nos hypothèses sont les suivantes :

$H_0$  : les données suivent une loi normale

$H_1$  : les données ne suivent pas une loi normale

Et,

$$JB = \frac{n - k}{6} \left( S^2 + \frac{(K - 3)^2}{4} \right)$$

Où  $n$  est le nombre d'observations (113),  $k$  le nombre de variables explicatives (4),  $S$  le coefficient d'asymétrie calculé à l'aide de la fonction de statistique descriptive dans Excel (0.4468) puis  $K$  le coefficient d'aplatissement Kurtosis également calculé dans Excel (0.64513). Ceci nous donne une statistique JB de 28.812. Avec un seuil de signification de 5%, ceci nous donne une valeur  $p$  de 0.00001. Ceci nous permet de conclure que nous pouvons rejeter  $H_0$  et que nos données ne suivent pas une loi normale. Une solution ici pourrait être une transformation de Box Cox.

## CHAPITRE 5 : CONCLUSION

Plusieurs de mes lectures, comme le rapport publié par le Federal Reserve Bank of St. Louis sur l'effet de la démographie des emprunteurs sur le retard de paiement<sup>24</sup> conclue que la démographie d'un emprunteur nous permet d'observer des tendances importantes dans le taux de défaut. De plus, nous avons vu que les données utilisées pour déterminer la probabilité de défaut sont surtout celles basées sur l'emprunteur telles que la source et stabilité du revenu, note de crédit, sommaires des actifs et passifs et donc l'évaluation de l'avoir net, sur le prêt demandé tel que le taux d'intérêt disponible et le rapport prêt-valeur puis les critères de la propriété telle que l'emplacement, la condition et la valeur. De plus, Campell et Dietrich (1983) avaient déterminé que certaines variables macro-économiques affectaient le risque de défaut tel que le taux de chômage local. Peut-être qu'utiliser un taux de chômage plus granulaire tel qu'un taux de chômage par État associé à un identifiant géographique aurait amélioré le modèle. Nos limitations telles que la limitation de programme statistique assez puissant pour extraire les données nécessaires ne nous a donc pas permis d'obtenir un bon modèle. La meilleure solution aurait assurément été de pouvoir utiliser un programme de traitement de données plus puissant qu'Excel et utiliser une base de données à l'initiation du prêt jointe à celle que nous avons utilisé qui permettait le suivi de leur statut. Il est évident que cette approche nous aurait fort probablement permis de tirer de meilleures conclusions.

J'en conclus donc que sans ces variables, il est difficile d'observer des corrélations et encore moins des causalités entre les variables explicatives utilisées dans ce projet et le statut d'un prêt. Il est évident que les institutions financières ont accès à

---

<sup>24</sup>[https://www.stlouisfed.org/~media/files/pdfs/hfs/assets/2016/emmons\\_ricketts\\_demographics\\_delinquency\\_1018\\_2016.pdf?la=en](https://www.stlouisfed.org/~media/files/pdfs/hfs/assets/2016/emmons_ricketts_demographics_delinquency_1018_2016.pdf?la=en)

beaucoup plus de ressources et de données pour leur permettre de développer de bons modèles à utiliser dans la gestion du risque de crédit et la probabilité de défaut.

## ANNEXES

### A- Modèle de régression logistique

Méthodologie telle que montrée sur le site suivant :

<https://blog.learningtree.com/how-to-apply-logistic-regression-using-excel/>

Nous avons débuté par estimer notre équation de régression en utilisant des coefficients arbitraires pour nos  $\beta$  de 0.0001 en comprenant que ces valeurs arbitraires seront remplacées avec nos valeurs optimales avec le Solver d'Excel. Nous avons ensuite calculé l'exponentielle de l'équation que nous venions de trouver afin de trouver notre  $e^{\alpha+\beta_1X_1+\beta_2X_2+\dots+\beta_kX_k}$ . Puis nous avons calculé notre probabilité de défaut, pour finalement calculer un log de vraisemblance. Finalement, nous avons utilisé le Solver dans Excel pour déterminer nos coefficients.

1. Calcul appliqué à toutes nos données pour le logit avec coefficients arbitraires :

current loan delinquency status	Current interest rate	Current actual UPB	Loan age in months	Chomage	LOGIT	exp	P(X)	LL
0	0.055	146944.25	58	0.074	=SUM\$2+\$M\$3*B2+\$M\$4*C2+\$M\$5*D2+\$M\$6*E2			
0	0.055	146710.41	59	0.074	14.6770539	2366810.675	0.9999995775	-14.67705432
0	0.055	146475.5	60	0.074	14.6536629	2312091.074	0.9999995675	-14.65366333
0	0.06375	118093.51	59	0.074	11.81536478	135315.5551	0.9999926099	-11.81537217
0	0.06375	117933.16	60	0.074	11.79942978	133176.3908	0.9999924912	-11.79943728
0	0.06375	76669.72	59	0.074	7.672985775	2149.489766	0.9995349896	-7.673450894
0	0.06375	76562.27	60	0.074	7.662340775	2126.729803	0.9995300155	-7.66281087
0	0.04875	119350.24	59	0.074	11.94103628	153435.6094	0.9999934827	-11.94104279

2. L'exponentiel de notre régression :

current loan delinquency status	Current interest rate	Current actual UPB	Loan age in months	Chomage	LOGIT	exp	P(X)	LL
0	0.055	146944.25	58	0.074	14.7003379	=EXP(F2)		
0	0.055	146710.41	59	0.074	14.6770539	2366810.675	0.9999995775	-14.67705432
0	0.055	146475.5	60	0.074	14.6536629	2312091.074	0.9999995675	-14.65366333
0	0.06375	118093.51	59	0.074	11.81536478	135315.5551	0.9999926099	-11.81537217
0	0.06375	117933.16	60	0.074	11.79942978	133176.3908	0.9999924912	-11.79943728
0	0.06375	76669.72	59	0.074	7.672985775	2149.489766	0.9995349896	-7.673450894

### 3. Probabilité :

current loan delinquency status	Current interest rate	Current actual UPB	Loan age in months	Chomage	LOGIT	exp	P(X)	LL		
2	0	0.055	146944.25	58	0.074	14.7003379	2422566.08	=G2/(1+G2)	-14.70033831	b0= 0.0001
3	0	0.055	146710.41	59	0.074	14.6770539	2366810.675	0.9999995775	-14.67705432	b1= 0.0001
4	0	0.055	146475.5	60	0.074	14.6536629	2312091.074	0.9999995675	-14.65366333	b2= 0.0001
5	0	0.06375	118093.51	59	0.074	11.81536478	135315.5551	0.9999926099	-11.81537217	b3= 0.0001
6	0	0.06375	117933.16	60	0.074	11.79942978	133176.3908	0.9999924912	-11.79943728	b4= 0.0001
7	0	0.06375	76669.72	59	0.074	7.672985775	2149.489766	0.9995349896	-7.673450894	
8	0	0.06375	76562.27	60	0.074	7.662340775	2126.729803	0.9995300155	-7.66281087	
9	0	0.04875	119350.24	59	0.074	11.94103628	153435.6094	0.9999934827	-11.94104279	
10	0	0.04875	118580.22	60	0.074	11.86413428	142078.3969	0.9999929617	-11.86414131	

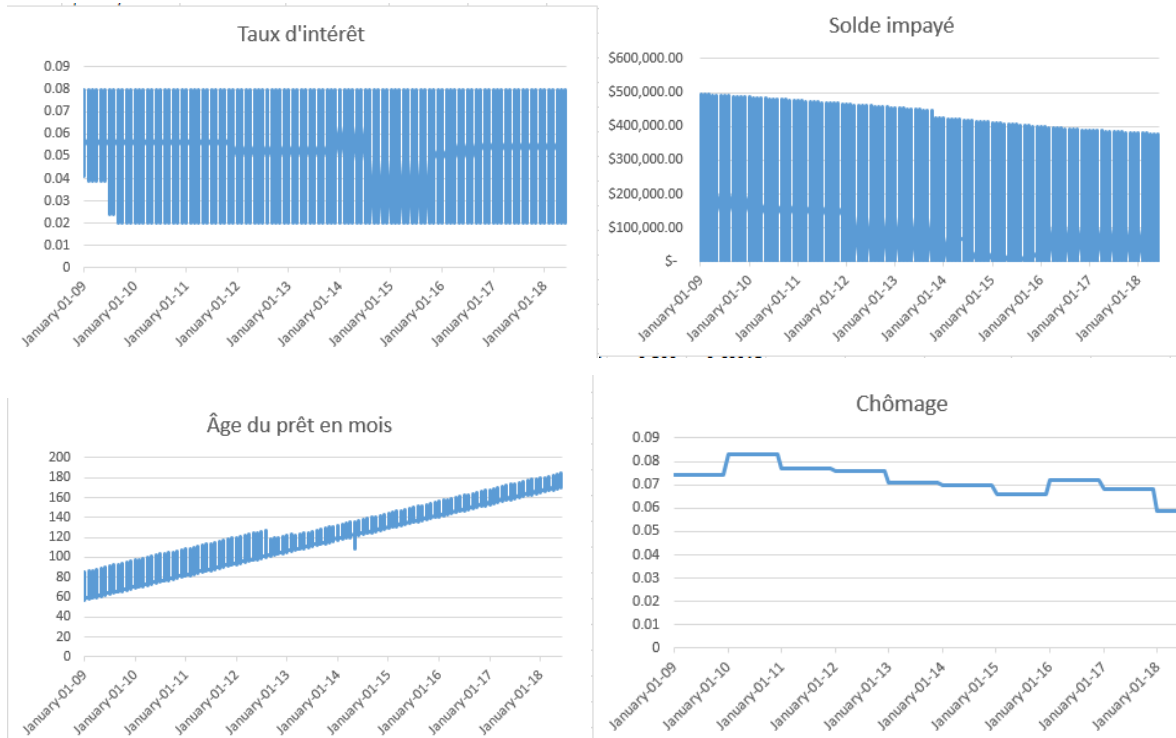
### 4. Log de vraisemblance :

current loan delinquency status	Current interest rate	Current actual UPB	Loan age in months	Chomage	LOGIT	exp	P(X)	LL		
2	0	0.055	146944.25	58	0.074	14.7003379	0.9999995872	=A2*LN(H2)+(1-A2)*LN(1-H2)	-14.67705432	b0= 0.0001
3	0	0.055	146710.41	59	0.074	14.6770539	2366810.675	0.9999995775	-14.65366333	b1= 0.0001
4	0	0.055	146475.5	60	0.074	14.6536629	2312091.074	0.9999995675	-11.81537217	b2= 0.0001
5	0	0.06375	118093.51	59	0.074	11.81536478	135315.5551	0.9999926099	-11.79943728	b3= 0.0001
6	0	0.06375	117933.16	60	0.074	11.79942978	133176.3908	0.9999924912	-7.673450894	b4= 0.0001
7	0	0.06375	76669.72	59	0.074	7.672985775	2149.489766	0.9995349896	-7.66281087	
8	0	0.06375	76562.27	60	0.074	7.662340775	2126.729803	0.9995300155		

### 5. Résultats :

current loan delinquency status	Monthly reporting period	Current interest rate	Current actual UPB	Loan age in months	Remaining months to maturity	Chomage	LOGIT	exp	P(X)	LL		
2	January-01-09	0.055	\$ 146,944.25	58	302	0.074	0	1	0.500	-0.69315	b0=	0
3	February-01-09	0.055	\$ 146,710.41	59	301	0.074	0	1	0.500	-0.69315	b1=	0
4	March-01-09	0.055	\$ 146,475.50	60	300	0.074	0	1	0.500	-0.69315	b2=	0
5	April-01-09	0.055	\$ 146,139.52	61	299	0.074	0	1	0.500	-0.69315	b3=	0
6	May-01-09	0.055	\$ 145,902.00	62	298	0.074	0	1	0.500	-0.69315	b4=	0
7	June-01-09	0.055	\$ 145,663.39	63	297	0.074	0	1	0.500	-0.69315	b5=	0
8	July-01-09	0.055	\$ 145,423.68	64	296	0.074	0	1	0.500	-0.69315		
9	August-01-09	0.055	\$ 145,107.88	65	295	0.074	0	1	0.500	-0.69315		
10	September-01-09	0.055	\$ 144,865.63	66	294	0.074	0	1	0.500	-0.69315	SUM OF LL	-295055
11	October-01-09	0.055	\$ 144,622.27	67	293	0.074	0	1	0.500	-0.69315		

On voit ici que nos coefficients ne sont pas pertinents et donc quelque chose ne fonctionne pas dans notre modèle. Nous avons donc décidé d'analyser nos variables plus en profondeur à l'aide de représentation graphique au fil du temps. Voir ci-dessous :



On voit clairement qu’il y a un problème. Le problème vient de nos limitations. Dans la manipulation de nos données, nous avons dû enlever les identifiants uniques qui nous permettaient de suivre chaque prêt au fil du temps. Ceci a pour effet que chaque observation est traitée comme étant un prêt individuel lorsqu’en réalité ils ne le sont pas. Par exemple, en regardant nos premières observations ci-bas, on remarque clairement en regardant le taux d’intérêt, le solde impayé (UPB) et l’âge du prêt que les observations sont pour un même prêt. Ceci explique pourquoi un modèle qui utilise les données individuelles ne fonctionne pas. Étant donnée la suite chronologique dans nos données, une manipulation telle que la première différence ne fonctionnerait pas non plus.



current loan delinquency status	Monthly reporting period	Current interest rate	Current actual UPB	Loan age in months	Remaining months to maturity	Chomage
0	January-01-09	0.055	\$ 146,944.25	58	302	0.074
0	February-01-09	0.055	\$ 146,710.41	59	301	0.074
0	March-01-09	0.055	\$ 146,475.50	60	300	0.074
0	April-01-09	0.055	\$ 146,139.52	61	299	0.074
0	May-01-09	0.055	\$ 145,902.00	62	298	0.074
0	June-01-09	0.055	\$ 145,663.39	63	297	0.074
0	July-01-09	0.055	\$ 145,423.68	64	296	0.074
0	August-01-09	0.055	\$ 145,107.88	65	295	0.074
0	September-01-09	0.055	\$ 144,865.63	66	294	0.074
0	October-01-09	0.055	\$ 144,622.27	67	293	0.074
0	November-01-09	0.055	\$ 144,252.79	68	292	0.074
0	December-01-09	0.055	\$ 144,006.62	69	291	0.074
0	January-01-10	0.055	\$ 143,759.32	70	290	0.083
0	February-01-10	0.055	\$ 143,510.89	71	289	0.083
0	March-01-10	0.055	\$ 143,261.32	72	288	0.083
0	April-01-10	0.055	\$ 143,010.60	73	287	0.083
0	May-01-10	0.055	\$ 142,758.74	74	286	0.083
0	June-01-10	0.055	\$ 142,505.72	75	285	0.083
0	July-01-10	0.055	\$ 142,251.54	76	284	0.083
0	August-01-10	0.055	\$ 141,996.20	77	283	0.083
0	September-01-10	0.055	\$ 141,738.60	78	282	0.083

Si notre modèle avait fonctionné, nous aurions pu ensuite procéder avec des tests pour déterminer la pertinence du modèle logistique tels que le test de Hosmer-Lemeshow, une table de classification et une analyse de la courbe ROC.<sup>25</sup> Mais puisque ce n'est pas le cas, nous avons évalué la pertinence d'un autre modèle.

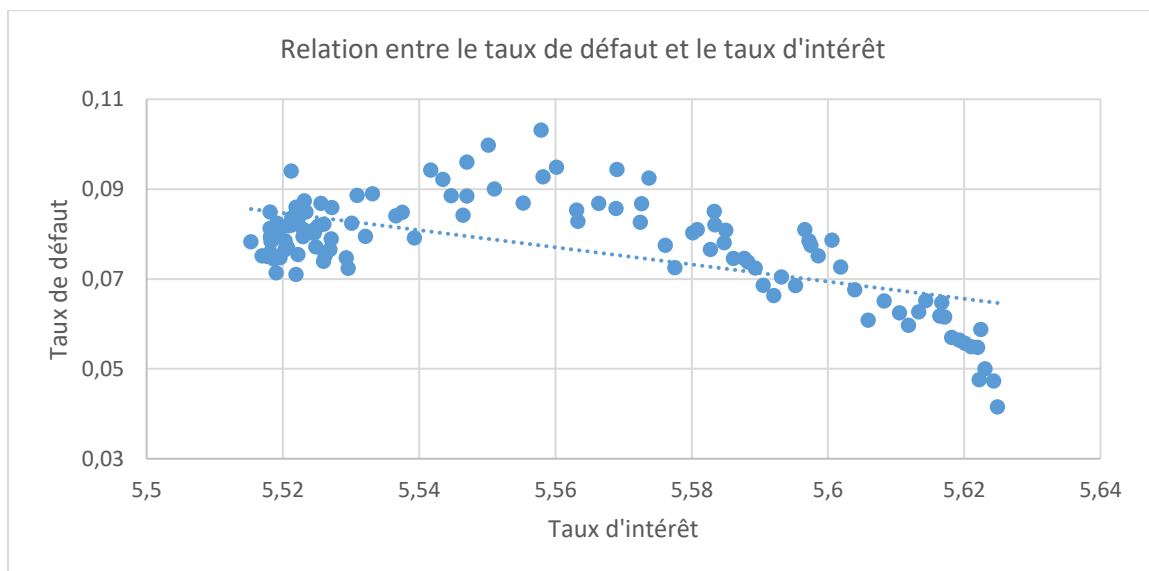
*B- Modèle de régression linéaire de portefeuille (avant la première différence)*

Avant de nous apercevoir que nos variables n'étaient pas stationnaires, nous avons testé certaines hypothèses qui doivent tenir dans une régression linéaire multiple, c'est-à-dire la relation linéaire entre chacune de nos variables explicatives et la

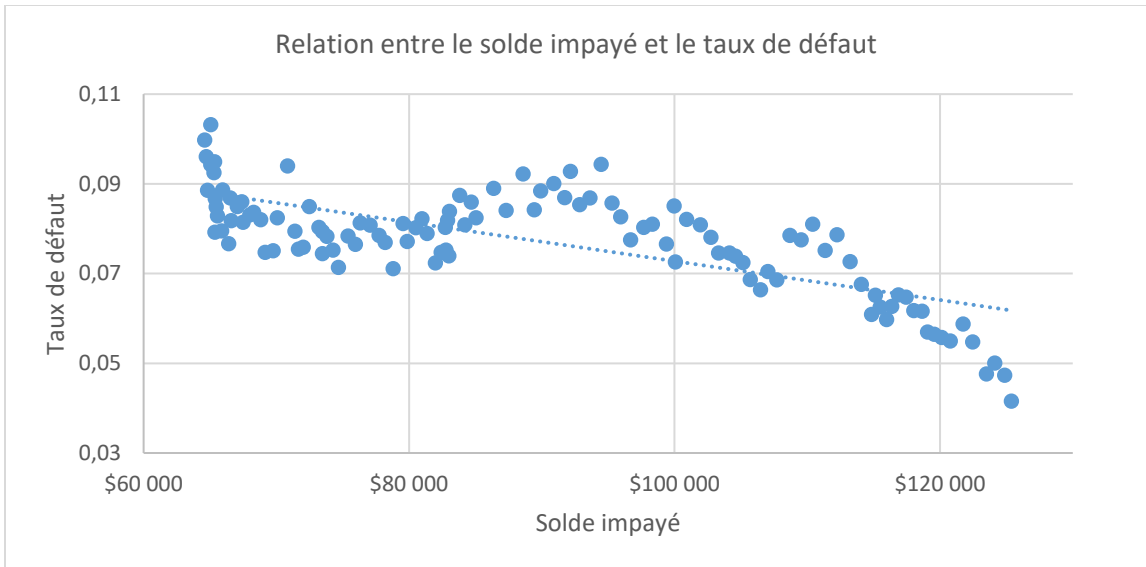
<sup>25</sup> [https://www.medcalc.org/manual/logistic\\_regression.php](https://www.medcalc.org/manual/logistic_regression.php)

variable expliquée, la multicollinéarité et la distribution gaussienne de nos résidus. Nos résultats étaient les suivants:

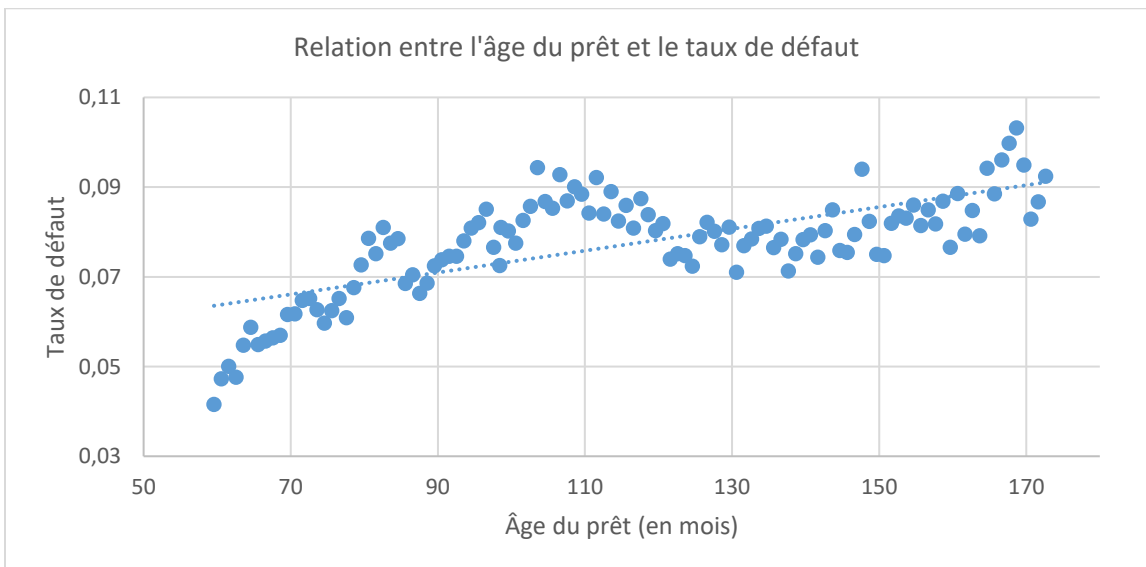
1. Relation linéaire : Chaque variable explicative aurait une relation linéaire avec la variable expliquée si toutes les autres variables explicatives étaient maintenues constantes. Afin de vérifier cette hypothèse, nous avons fait des représentations graphiques de chacune de nos variables explicatives avec notre variable expliquée, le taux de défaut.



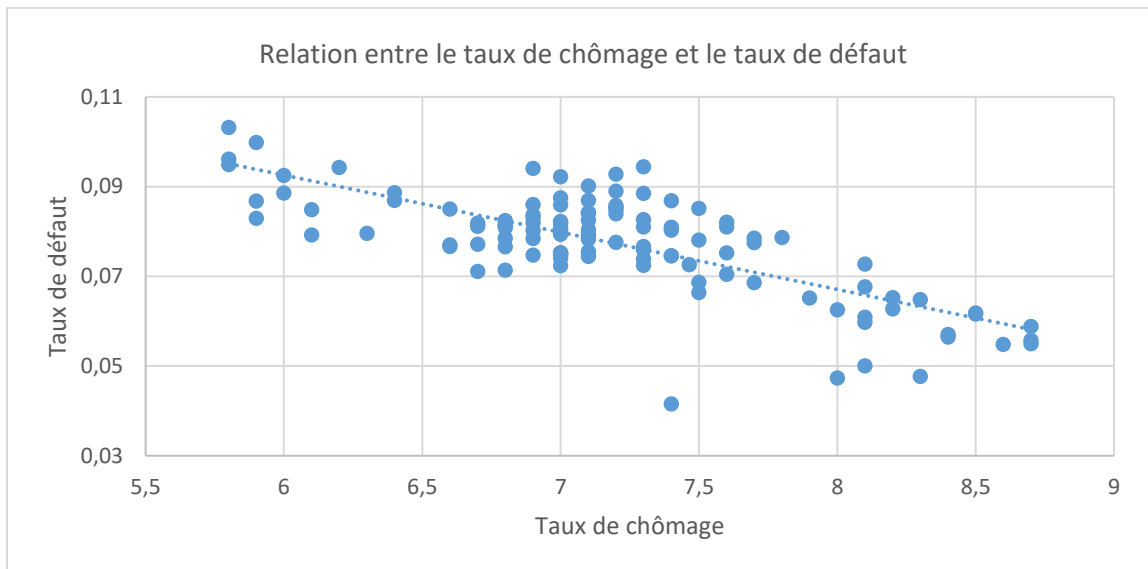
En regardant le graphique ci-haut, nous pouvons voir que bien que les points semblent rapprochés de la droite, ils ont plutôt une forme sinusoïdale. Il semble y avoir une hausse dans le taux de défaut lorsque le taux d'intérêt augmente autour de 5.56%, puis une baisse plus le taux d'intérêt augmente passé ce point. Le taux de défaut et le taux d'intérêt ne semblent donc pas avoir une relation linéaire.



En regardant le graphique ci-haut, nous pouvons voir que bien que les points semblent rapprochés de la droite, ils ont encore plutôt une forme sinusoïdale. Il semble y avoir une baisse du taux de défaut entre 60K\$ et 80K\$, puis une hausse entre \$80K et \$100K, pour ensuite baisser passé le \$100K. Le taux de défaut et le solde impayé ne semblent donc pas avoir une forme linéaire.



En regardant le graphique ci-haut, nous pouvons voir que bien que les points semblent rapprochés de la droite, ils ont encore plutôt une forme sinusoïdale. Il semble y avoir un bas taux de défaut lorsque le prêt est à moins de 90 mois, puis il y a une tendance vers la hausse près de 110 mois, pour ensuite retomber. Le taux de défaut et l'âge du prêt ne semblent donc pas avoir une relation linéaire.



Puis finalement en regardant le graphique ci-haut qui représente les taux de défaut en fonction des taux de chômage mensuels récoltés, nous pouvons voir qu'il semble avoir une relation négative entre le taux de chômage et le taux de défaut, ce qui semble contre-intuitif. La forme ne semble pas être linéaire par contre, étant donné le regroupement de points de données au centre.

Nous pouvons donc conclure que notre première hypothèse de relation linéaire ne semble pas être respectée pour aucune de nos variables explicatives. Il semble d'ailleurs y avoir présence d'hétéroscédasticité dans chacun de nos résidus examinés plus haut. La solution au problème de serait soit d'abandonner la linéarité ou bien de l'améliorer en utilisant une

transformation afin de permettre une meilleure explication. Dans le cadre de ce mémoire, nous n'avons pas cherché à faire les modifications pour corriger le modèle.

2. Multicolinéarité: Afin qu'on puisse utiliser la régression linéaire multiple, nous avons une deuxième hypothèse indiquant qu'il ne doit pas avoir de haute corrélation entre les variables explicatives. Afin de tester la colinéarité entre nos différentes variables, nous avons examiné les facteurs d'inflation de la variance (FIV), une approche qui consiste à déterminer de combien la variance d'un coefficient est augmentée dû à sa relation linéaire avec d'autres coefficients. Nous avons donc simplement régressé chaque variable explicative sur les autres variables sans inclure la variable originale expliquée, c'est-à-dire le taux de défaut, puis utiliser le  $R^2$  pour trouver notre FIV en utilisant la formule suivante :

$$VIF_i = \frac{1}{1 - R_i^2}$$

Il n'y a pas de consensus pour déterminer le niveau de VIF acceptable. Certaines recherches indiquent que nous pouvons accepter les variables avec des FIV inférieures à 3, d'autres à 5<sup>26</sup> et certains vont même jusqu'à 10<sup>27</sup>. L'analyse du FIV m'a donné les chiffres suivants :

	<b>FIV</b>
<b>Taux d'intérêt</b>	38.017
<b>Solde impayé</b>	520.309
<b>Âge du prêt</b>	289.488
<b>Taux de chômage</b>	7.091

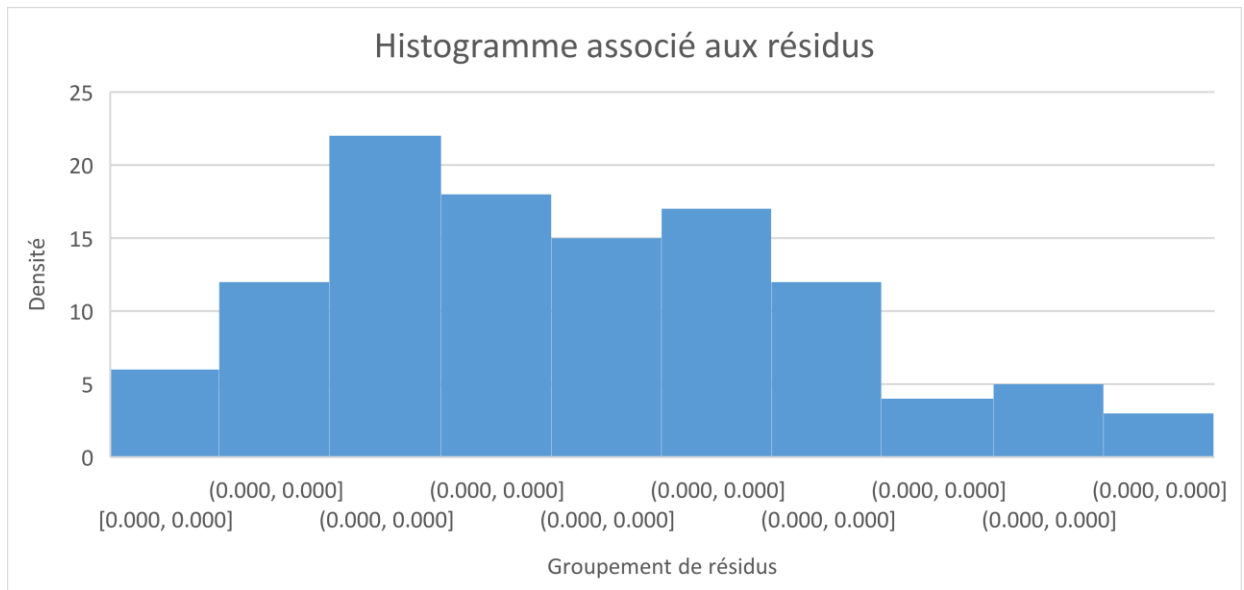
---

<sup>26</sup> Ringle, Christian M., Wende, Sven, & Becker, Jan-Michael. (2015). SmartPLS 3. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>

<sup>27</sup> Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. & Black, W. C. (1995). Multivariate Data Analysis (3rd ed). New York: Macmillan.

Nous pouvons conclure qu'aucune de nos variables explicatives ne semble adéquate étant donné les hauts FIVs mise à part peut-être le taux de chômage si nous considérons les études moins conservatrices indiquant qu'un FIV inférieure à 10 est acceptable. Le taux de chômage était notre seule variable qui venait d'un autre échantillon de donnée. La solution à la multicollinéarité serait d'enlever la ou les variables qui causent le problème, mais dans notre cas, elles semblent en majorité toutes problématiques. Une solution possible pourrait d'enlever le solde impayé puisqu'il a le FIV le plus élevé et apporte donc de l'information hautement redondante à notre modèle. Nous pourrions ensuite réévaluer le modèle afin de voir si cela règle notre problème. Cependant, dans le cadre de ce mémoire nous n'avons pas cherché à faire les modifications pour corriger le modèle.

3. Distribution gaussienne : Notre troisième hypothèse qui doit tenir afin qu'on puisse utiliser la régression linéaire des moindres carrés est celle qui suppose que les résidus sont distribués selon une loi normale de moyenne 0. Nous pouvons d'abord évaluer la normalité avec une représentation graphique des résidus. En regardant le graphique ci-bas, il est possible que nous puissions conclure que nos résidus suivent une loi normale.



Nous avons donc testé cette hypothèse avec un test de normalité en utilisant la statistique Jarque Béra qui suit asymptotiquement une loi  $\chi^2$  à 2 degrés de liberté, où nos hypothèses sont les suivantes :

$H_0$  : les données suivent une loi normale

$H_1$  : les données ne suivent pas une loi normale

Et,

$$JB = \frac{n - k}{6} \left( S^2 + \frac{(K - 3)^2}{4} \right)$$

Où  $n$  est le nombre d'observations (114),  $k$  le nombre de variables explicatives (4),  $S$  le coefficient d'asymétrie calculé à l'aide de la fonction de statistique descriptive dans Excel (0.4950) puis  $K$  le coefficient d'aplatissement Kurtosis également calculé dans Excel (-0.1741). Ceci nous donne une statistique JB de 50.67. Avec un seuil de signification de 5%, ceci nous donne une valeur  $p$  de 0.00001. Ceci nous permet de conclure que nous pouvons rejeter  $H_0$  et que nos données ne suivent pas une loi normale. Une solution ici pourrait être une transformation de Box Cox.

### C- Résultats de régression linéaire avec différences

#### Première différence $x(t)-x(t-1)$

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.297862623			
R Square	0.088722142			
Adjusted R Square	0.0636411			
Standard Error	0.011004119			
Observations	113			
<b>ANOVA</b>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	3	0.001285045	0.00043	3.53742
Residual	109	0.013198878	0.00012	
Total	112	0.014483923		
<i>Coefficients</i>				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.081196915	0.002243393	36.1938	1.4E-62
Average of Current interest rate (lagged)	0.007995865	0.521880544	0.01532	0.9878
Average of Current actual UPB (lagged)	-7.9171E-06	3.97642E-06	-1.99101	0.04898
Average of Chomage (lagged)	0.015168622	0.007550244	2.00902	0.04701

#### Deuxième différence $x(t-1)-x(t-2)$

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.12329			
R Square	0.0152			
Adjusted R Square	-0.01215			
Standard Error	0.01146			
Observations	112			
<b>ANOVA</b>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	3	0.00021876	7.3E-05	0.55567
Residual	108	0.01417278	0.00013	
Total	111	0.01439155		
<i>Coefficients</i>				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.0771	0.00108284	71.2039	1.3E-92
Average of Current interest rate (lagged t2)	-0.40754	0.45860298	-0.88865	0.37616
Average of Current actual UPB (lagged t2)	-2.3E-06	3.837E-06	-0.60126	0.54892
Average of Chomage (lagged t2)	-0.00556	0.00594166	-0.93594	0.35139

#### Troisième différence $x(t-2)-x(t-3)$



SUMMARY OUTPUT					
<i>Regression Statistics</i>					
Multiple R	0.06187				
R Square	0.00383				
Adjusted R Square	-0.0241				
Standard Error	0.01156				
Observations	111				
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	5.5E-05	1.8E-05	0.13706	0.9119
Residual	107	0.0143	0.00013		
Total	110	0.01436			
<i>Coefficients</i>					
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower Bound</i>	<i>Upper Bound</i>
Intercept	0.07704	0.0011	70.1848	2.6E-91	0.0758
Average of Current interest rate (lagged t3)	0.15633	0.27561	0.56721	0.57176	-0.2610
Average of Current actual UPB (lagged t3)	6.6E-07	2.3E-06	0.29168	0.77109	-1.2E-06
Average of Chomage (lagged t3)	0.0006	0.00356	0.16917	0.86598	-0.0023

Première et troisième différence :

SUMMARY OUTPUT					
<i>Regression Statistics</i>					
Multiple R	0.3247958				
R Square	0.1054923				
Adjusted R Square	0.0538861				
Standard Error	0.0111127				
Observations	111				
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	6	0.001514642	0.00025244	2.04418	0.0471
Residual	104	0.012843195	0.00012349		
Total	110	0.014357837			
<i>Coefficients</i>					
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower Bound</i>	<i>Upper Bound</i>
Intercept	0.0807805	0.002457358	32.8729103	9.4E-57	0.0758
Average of Current interest rate (lagged)	-0.0320551	0.620861921	-0.05162997	0.95892	-0.6549
Average of Current actual UPB (lagged)	-7.475E-06	4.45093E-06	-1.67931882	0.09609	-1.5E-05
Average of Chomage (lagged)	0.0205645	0.008818163	2.33205623	0.02162	0.0029
Average of Current interest rate (lagged t3)	0.2213743	0.284219288	0.77888547	0.43782	-0.3461
Average of Current actual UPB (lagged t3)	1.28E-06	2.32755E-06	0.5497586	0.58366	-4.5E-06
Average of Chomage (lagged t3)	-0.0031573	0.003770392	-0.83740135	0.40429	-0.0070

## Références

1. Akindaini, Bolarinwa «Machine Learning Applications in Mortgage Default Prediction» *University of Tampere, Faculty of Natural Sciences (Nov.2017)*
2. Ambrose, Brent W., Capone JR., Charles A. and Deng, Yongheng «Optimal Put Exercise: An empirical Examination of Conditions for Mortgage Foreclosure» *Journal of Real Estate Finance and Economics (2001) pp. 213-234*
3. Asarnow, Elliot and Edwards, David «Measuring loss on Defaulted Bank Loans: A 24-Year Study» *Journal of Commercial Lending (March 1995) pp. 11-23*
4. Autorité des marchés financiers «Ligne directrice sur l’octroi de prêts hypothécaires résidentiels» *mars 2018*
5. Bandyopadhyay, Arindam and Singh, Pratima «Estimating Recovery Rates on Bank’s Historical Loan Loss Data» *Munich Personal RePEc Archive (February 2007) Paper No. 9525*
6. Bera, Anil K., Carlos M. Jarque «Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence» (1981). *Economics Letters 7 (4) : 313–318*
7. BSIF. *Lignes directrices, Normes des fonds propres (NFP)*. Decembre 2016. Chapitre 1 – Vue d’ensemble, Chapitre 3 – 3.6.1 Évaluations externes du crédit, Chapitre 5 – Atténuation du risque de crédit, Chapitre 6 – Risque de crédit – Approche fondée sur les notations internes.
8. Bujang, Mohamad Adam, Sa’at, Nadiah, Tg Abu Bakar Sidik, Tg Mohd Ikhwan and Joo, Lim Chien «Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data» *Malaysian Journal of Medical Sciences (Jul 2018) pp. 122-130*

9. Calem, Paul S. and LaCour-Little, Michael «Risk-based capital requirements for mortgage loans» *Journal of Banking & Finance* (2004) pp.647-672
10. Campbell, Tim S. and Dietrich, J. Kimball. «The Determinants of Default on Insured Conventional Residential Mortgage Loans» *The Journal of Finance* (Dec., 1983) pp. 1569-1581.
11. Capozza, Dennis R., Kazarian, Dick and Thomson, Thomas A. «The Conditional Probability of Mortgage Default» *Real Estate Economics* (1996) V26 pp. 359-389
12. Carty, Lea V. «Moody's Rating Migration and Credit Quality Correlation, 1920-1996». *New York : Moody's Investors Service, Global Credit Research, July 1997. Report Number 25097.*
13. Comité de Bâle sur le contrôle bancaire, *Finalisation de Bâle III- En Bref* (décembre 2017)
14. Curley, Anthony J., Guttentag, Jack M. «The Yield on Insured Residential Mortgage» *National Bureau of Economic Research* (1974) pp.114-161
15. Ebrary.net «IRB Foundation and Advanced Methods»  
[https://ebrary.net/9510/business\\_finance/irb\\_foundation\\_advanced\\_methods](https://ebrary.net/9510/business_finance/irb_foundation_advanced_methods)
16. Emmons, William R., and Ricketts, Lowell R. «The Demographics of Loan Delinquency: Tipping points or tip of the iceberg?» *Center for Household Financial Stability, Federal Reserve Bank of St. Louis* (Oct.2016)
17. Green, Jerry and Shoven, John B. «The Effects of Interest Rates on Mortgage Prepayments» *Journal of Money, Credit and Banking* (Feb., 1986) pp. 41-59.
18. Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. & Black, W. C. «*Multivariate Data Analysis* (3rd ed).» New York: Macmillan (1995)
19. Hale, Galina, Krainer, John and McCarthy, Erin «Aggregation Level in Stress Testing Models» *Federal Reserve Bank of San Francisco, Working paper* 2015-14

20. Hassani, Samir Saïssi «Modélisation du stress testing du risque de crédit appliqué à un portefeuille de prêts aux particuliers» *Mémoire (Juillet 2007) présenté au HEC Montréal*
21. Investopedia. [Online] Definition of probability of default  
<https://www.investopedia.com/terms/d/defaultprobability.asp>.
22. Jarque, Carlos M. & Anil K. Bera «Efficient tests for normality, homoscedasticity and serial independence of regression residuals» (1980) *Economics Letters* 6 (3) : 255–259.
23. Jarque, C. M. & Bera, A. K. «A test for normality of observations and regression residuals» (1987) *International Statistical Review* 55, 163–172.
24. Kersten, Erik «Everything you always wanted to know about Basel II in 15 minutes (a real estate perspective)» *De Nederlandsche Bank Eurosystem*
25. Kutty, Gopalan «Logistic regression and probability of default of developing country debt» *Journal of Applied Economics, Volume 22, 1990 – Issue 12. Pp 1649-1660*
26. LaCour-Little, Michael «Equity Dilution: An Alternative Perspective on Mortgage Default» *Real Estate Economics (September 2004) pp. 359-384*
27. Leow, Mindy and Mues, Christophe «Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data» *International Journal of Forecasting (2012) pp. 183-195*
28. Li, Mingxin «Residential Mortgage Probability of Default Models and Methods» *Risk Surveillance and Analytics, Financial Institutions Commission of British Columbia (October 2014)*
29. McMillan LLP, «Getting the Deal Through Banking Regulation 2017» *Law Business Research Ltd*

30. MedCalc, «Logistic regression»  
[https://www.medcalc.org/manual/logistic\\_regression.php](https://www.medcalc.org/manual/logistic_regression.php)
31. Misina, Miroslav, Tessier, David and Dey, Shubhasis «Stress Testing the Corporate Loans Portfolio of the Canadian Banking Sector» *Bank of Canada Working Paper No. 2006-47*
32. Qi, M. and Yang, X.L. «Loss Given Default of High Loan-to-Value Residential Mortgages» *Journal of Banking & Finance* (2009) pp 788-799
33. Ringle, Christian M., Wende, Sven, & Becker, Jan-Michael. *SmartPLS 3. Bönningstedt: SmartPLS*. (2015) Provenant de <http://www.smartpls.com>
34. Tong, Edward N.C. Mues, Christophe and Thomas, Lyn «A zero-adjusted gamma model for mortgage loan loss given default» *Internal Journal of Forecasting* (Oct-Dec 2013) Vol. 29, Issue 4, pp. 548-562
35. Van Roy, Patrick «Credit Ratings and the Standardised Approach to Credit Risk in Basel II» *European Central Bank (August 2005) Working Paper Series No. 517*
36. Vandell, Kerry D., «Default Risk Under Alternative Mortgage Instruments» *The Journal of Finance* (Dec 1978) Vol. 33 No. 5 pp. 1279-1296
37. Von Furstenberg, George M, «Default Risk on FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis» *The Journal of Finance* (Jun 1969) Vol. 24 No. 3 pp. 459-477
38. Zhang, Qingfen «Modelling the probability of mortgage default via logistic regression and survival analysis» *Open Access Master's Theses* (2015). Paper 541