

HEC MONTRÉAL

**Modélisation du stress testing
du risque de crédit appliqué à un portefeuille
de prêts aux particuliers**

Par

Saïssi Hassani, Samir

Sciences de la gestion

(Économie financière appliquée)

Mémoire présenté en vue de l'obtention

du grade de maître ès sciences

(M.Sc.)

Juillet 2007

© Samir Saïssi Hassani, 2007

Sommaire

L'objectif de cette étude est de construire un modèle de stress testing appliqué au risque de crédit d'un portefeuille de prêts aux particuliers. Contrairement à ce qui était généralement admis jusqu'à récemment, la modélisation du risque de crédit des prêts aux particuliers n'est pas un cas particulier des modèles du risque de crédit pour les entreprises. Le problème de ce genre de portefeuille est la présence importante d'une composante du risque spécifique par rapport au risque systématique. Pour les entreprises, c'est généralement le contraire. Nous développerons notre modèle à la base de Wilson (1997a, b), de sorte à capter la composante spécifique par des variables idiosyncratiques des prêts et des individus eux-mêmes. D'autre part, la composante systématique est captée par des facteurs macroéconomiques pertinents. Pour ce faire, nous ferons appel aux fonctions de survie de Cox (1975) dont l'utilisation est grandement simplifiée par Shumway (2001). Nous explorerons également l'initiative fondatrice de Gouréroux et al. (2006) pour modéliser le taux de recouvrement. Des simulations Monte Carlo nous permettront d'évaluer les prédictions des pertes, aussi bien dans le cas de l'exploitation normale de notre institution financière qu'en cas de crises économiques hypothétiques. Nos résultats montrent que la perte prévisionnelle, calculée pour l'échantillon dont nous disposons, ne dépasse pas 8% par année dans le pire des scénarios.

Table des matières

Sommaire	i
Liste des tableaux	iv
Liste des figures	v
Remerciements	vi
1 Introduction	1
2 Revue de la littérature	5
2.1 Les bases de la modélisation du risque de crédit	5
2.2 Les fondements des modèles théoriques	9
2.3 Les principaux modèles industriels utilisés	11
2.4 Les approches stress testing du risque de crédit	14
2.5 Les préliminaires de la modélisation à retenir	17
3 Les données	18
4 Le modèle	28
4.1 Distribution PD	28
4.2 Distributions <i>LGD</i> et <i>EAD</i>	35

<i>TABLE DES MATIÈRES</i>	iii
5 Résultats et analyse	40
5.1 La distribution <i>PD</i> à la Shumway	40
5.2 La distribution <i>LGD</i> à la Gouriéroux et al.	44
5.3 Simulations Monte Carlo	48
6 Conclusion	52
Bibliographie	54
A Graphiques	60
B Étude du système VAR(p)	63

Liste des tableaux

3.1	Statistiques descriptives de <i>LGD</i> et <i>EAD</i>	23
3.2	Évolution de la fréquence des défauts et des pertes	24
3.3	Description des variables macroéconomiques retenues	26
3.4	Statistiques descriptives des variables macroéconomiques retenues . .	27
4.1	Signes prédits pour modéliser la probabilité de défaut <i>PD</i>	35
4.2	Signes prédits pour modéliser <i>LGD</i>	39
5.1	Estimation du Logit multi-périodique	41
5.2	Corrélations entre les variables explicatives	42
5.3	Comparaison des modèles Logit multi-périodiques	45
5.4	Fréquence historique et prédictions des défauts des modèles 1,2 et 3 .	46
5.5	Estimation conjointe Tobit <i>LGD/PD</i>	47
5.6	Pertes trimestrielles prévisionnelles du portefeuille	50
B.1	Prévisions des valeurs macroéconomiques	68
B.2	Prévisions suite à une récession (un écart-type négatif)	68
B.3	Prévisions suite à un choc sur le chômage (un écart-type positif) . .	68

Table des figures

2.1	Distributions typiques des rendements de marché et de crédit	6
3.1	Répartition des prêts par cote de risque	19
3.2	Répartition des défauts par cote de risque (en %)	22
3.3	Répartition des défauts par trimestre	22
3.4	Distribution historique de <i>LGD</i>	24
4.1	Distribution des pertes historiques	37
5.1	Influence de l'âge des prêts sur les probabilités de défaut	43
5.2	Comparaison des prédictions des probabilités de défaut	46
A.1	Simulation des conditions économiques ordinaires	60
A.2	Simulation d'une récession économique	61
A.3	Simulation d'une crise de chômage	62
B.1	Évolution des séries macroéconomiques	69
B.2	Effet des chocs sur les variables macroéconomiques (sur 3 ans)	70
B.3	Effet des chocs sur les variables macroéconomiques (sur 12 mois)	71

Remerciements

Je tiens à remercier mon directeur de recherche, M. Georges Dionne, pour son soutien direct et ponctuel, tout au long de ce travail. Ses encouragements ont très certainement contribué, également, à me valoir une place sur le Tableau d'Honneur de l'École et les bourses d'excellence que j'ai obtenues.

Je voudrais aussi remercier tous mes professeurs dont j'ai eu, à la fois, le plaisir et l'honneur d'avoir compté parmi leurs étudiants à HEC Montréal.

Chapitre 1

Introduction

À la prochaine entrée en vigueur de l'accord de Bâle II à partir de 2007, les institutions financières se forcent de finaliser leurs modèles internes de gestion des risques. Les recherches de modélisation en risque de crédit, déjà peu nombreuses par rapport à celles du risque de marché, se sont presque exclusivement concentrées sur le crédit aux entreprises publiques. Le crédit du détail, englobant les PME avec le crédit aux particuliers, a longtemps été considéré comme un cas particulier du premier, où il convenait d'appliquer une version adaptée des développements retenus. De ce fait, les travaux sur le crédit de détail sont encore plus rares. Tout récemment, les quelques pionniers qui se sont intéressés au sujet ont trouvé que ce type de crédit présentait des spécificités différentes. Notre travail s'inscrit dans ce contexte. Aussi, nous proposons-nous de construire et d'appliquer un modèle de stress testing conforme aux exigences de Bâle II, pour une institution canadienne. Cela s'adresse à son portefeuille d'hypothèques, de prêts à la consommation et de marges de crédit aux particuliers, cependant, nous ne traiterons que les prêts à la consommation dans ce travail.

Commençons par définir le stress testing et situer son importance dans le processus de gestion des risques des institutions financières. Le stress testing est une procédure de simulation de l'impact des crises économiques et financières sur les portefeuilles

détenus par les institutions. Les recommandations du comité de Bâle visent à limiter les risques que ces institutions prennent. Deux types de précautions sont de mise. La première est la détermination du capital minimum requis à garder par l'institution pour une exploitation ordinaire de tous les jours. La deuxième est la procédure du stress testing qui permet d'évaluer les vulnérabilités des investissements de l'institution en cas de crises. L'évaluation de la sensibilité des portefeuilles se fait avec des scénarios hypothétiques. Pour Bâle II, un scénario de désastre doit être peu probable mais plausible. L'article 435 du nouvel accord précise les exigences du stress testing appliqué au risque de crédit. Dans sa nouvelle version, l'accord de Bâle revoit les trois types de risque : le risque de marché, celui de crédit et le risque opérationnel. Le risque opérationnel gagne relativement de l'intérêt et de la précision par rapport à la version précédente de Bâle, mais c'est essentiellement le risque de crédit qui bénéficie d'une attention toute particulière, puisqu'il représente à lui seul 70% du risque total des institutions.

Pour simplifier, le comité de Bâle propose toujours un modèle dit standard qu'il convient d'appliquer tel quel, à moins de disposer d'un modèle interne que l'institution préfère adopter. Dans ce cas, il lui faudra prouver sa capacité à l'appliquer sous trois conditions. La première est que l'institution dispose de tous les moyens pour le mettre en oeuvre. La seconde est qu'elle sache le tester ex-post, afin de mesurer les échecs constatés par rapport aux prédictions du modèle (back testing). Enfin, la troisième est, justement, de développer une méthodologie cohérente pour le stress testing afin de compléter le dispositif global de la maîtrise des risques.

D'un point de vue technique, pour identifier le capital requis à un horizon h donné, il convient de déterminer la distribution des pertes et profits du portefeuille. Nous pouvons mesurer le risque en déterminant les pertes aux derniers quantiles, ce qui donne le montant de perte maximale que court l'institution au degré de confiance de

99%, par exemple. Cette mesure peut être faite à l'aide de concepts simples tels que celui de la VaR (Valeur à Risque). Pour le stress testing, il s'agit aussi de calculer aussi des montants à risque des portefeuilles, mais dans des configurations de crises. Ces montants calculés ne sont pas en principe pour les garder dans l'institution, mais plutôt pour prendre des actions préventives en termes de couvertures ou de réduction des positions sensibles du portefeuille. L'objectif de cette action est de réduire l'exposition globale au risque (Jorion, 2005).

Une fois ce point clarifié, signalons déjà la difficulté de choisir des scénarios de crise. Pour Bâle II, les articles 434 et 435 incitent suffisamment pour bâtir les simulations de crise par rapport à des facteurs précis, avec une formulation économique et financière explicite. D'un autre côté, historiquement, chacune des crises possède sa propre spécificité. Aucune crise ne ressemble à aucune autre par le passé. De ce fait, s'il est toujours important de simuler les crises antérieures, ce n'est pas suffisant pour tirer les bonnes conclusions. Actuellement, ce qui semble se généraliser dans les études est l'idée de recourir aux chocs unitaires et individuels de facteurs macroéconomiques afin d'en déduire et d'étudier l'impact pour les comprendre. Une fois les corrélations entre ces différents facteurs identifiées, on peut alors confectionner et simuler des chocs plus complexes. Dans ce sens, nous comprendrons que les praticiens privilégient les modèles de crédit qui explicitent directement les relations avec les facteurs macroéconomiques. Conceptuellement, le calcul du capital requis et celui des pertes du stress testing sont similaires dans le sens que les deux se basent sur le même modèle de crédit, et découlent des prédictions en hors échantillon. Cependant, si la détermination du capital requis suppose une continuité dans l'évolution du contexte économique, le stress testing envisage plutôt des chocs qui sortent de l'ordinaire.

La plupart des modèles que nous examinerons sont bâtis pour les prêts aux entreprises. Comme nous allons le voir, il n'est pas évident de les transposer purement et

simplement aux prêts aux particuliers. En effet, la principale spécificité est la présence d'une composante idiosyncratique trop importante pour ne pas en tenir compte par rapport à la composante systématique du risque de crédit global. Dans notre démarche, nous dégagerons progressivement des faits qui nous conduiront à choisir le modèle de Wilson (1997a,b) comme base de développement. Nous proposerons une variante de ce modèle adaptée à notre problématique. En fait, pour tenir compte des deux composantes du risque de crédit, spécifique et systématique, de notre portefeuille, nous aurons besoin du concept de fonctions de survie de Cox (1975). Nous verrons comment nous le mettrons à contribution pour traiter notre modèle et interpréter ses résultats, en utilisant les travaux de Shumway (2001). Une deuxième partie de la modélisation s'intéressera au taux de recouvrement en cas de défaut. Nous y explorerons les premières idées fondatrices de Gouriéroux et al. (2006) pour tenter une modélisation conjointe des taux de pertes en cas de défaut (complément des taux de recouvrement) avec les probabilités de défaut de notre portefeuille. Enfin, nous finirons par la construction des distributions des pertes à l'aide de simulations par la méthode de Monte Carlo. Avant cela, nous allons passer en revue la littérature orientée vers la modélisation du risque de crédit.

Chapitre 2

Revue de la littérature

Dans ce chapitre, nous allons tenter d'identifier les spécificités du risque de crédit quant à sa modélisation et ses fondements théoriques dans les deux premières sections. Ceci devra nous permettre de voir, dans la troisième section, comment ces bases sont approchées dans la pratique par les concepteurs des modèles industriels les plus utilisés par les institutions financières. Nous discuterons, ensuite, des propriétés fondamentales de la modélisation du stress testing du risque de crédit proprement dit, et nous terminerons ce chapitre avec des préliminaires que nous retenons pour notre modèle final. Commençons, maintenant, par les particularités du risque de crédit et de sa modélisation.

2.1 Les bases de la modélisation du risque de crédit

Le risque de crédit est la perte probable que court le prêteur suite au non paiement de l'emprunteur. L'état de défaut de ce dernier est identifié dès qu'il y a retard de paiement de plus de quatre-vingt-dix jours.

Contrairement au risque de marché où, en général, on considère que les distributions des rendements suivent une loi normale, le risque de crédit est caractérisé par

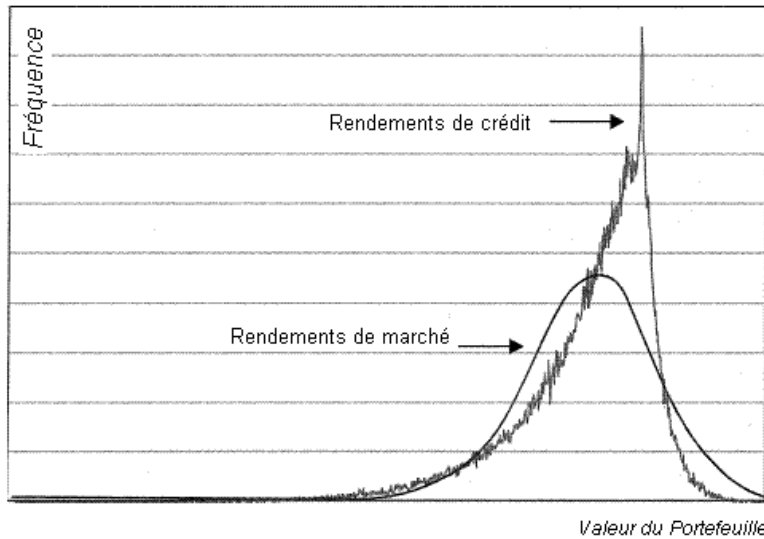


FIG. 2.1 – Distributions typiques des rendements de marché et de crédit
(source : CIBC)

une distribution leptocurtique et franchement asymétrique vers les valeurs négatives, comme le montre l'exemple de la figure 2.1. Continuer à supposer sa normalité revient à sous-estimer le risque de crédit correspondant. En effet, la modélisation de la perte se base sur ce qui se passe dans les derniers quantiles, et donc dépendra de l'information qui réside dans la queue gauche de la distribution.

Le fondement de la modélisation du risque de crédit prend son départ sur une formulation intuitive. La perte probable sur un prêt i , appartenant à une catégorie k , est le produit de la probabilité que l'emprunteur fasse défaut, du montant dû au moment du défaut t , et de la proportion qui sera récupérée par le recouvrement de l'institution. L'équation de base s'écrit

$$L_{i,k,t} = PD_{i,k,t} * LGD_{i,k,t} * EAD_{i,k,t} \quad (2.1)$$

L est le montant de la perte (Losses), PD est la probabilité de défaut (Probability of Default), LGD est le taux de perte en cas de défaut (Loss Given Default), EAD est l'exposition au risque (montant dû) au moment du défaut (Exposure At Default). Par définition, LGD et EAD sont conditionnels à l'état de défaut PD . Dans la lit-

térature, la variable EAD est fréquemment considérée déterministe ou, au mieux, assimilée à une distribution normale. Par contre LGD est le plus souvent considérée stochastique. On aura compris que la pierre angulaire des développements en risque de crédit est la distribution PD . D'ailleurs, beaucoup de modèles ne diffèrent entre eux que par la modélisation de PD comme nous allons voir par la suite. Examinons d'abord quelques aspects de LGD et EAD .

LGD est généralement approchée par une loi Bêta. Il n'y a aucune intuition derrière cette utilisation si ce n'est le fait que l'on veut une distribution à support de densité fixe entre 0 et 1, donc de 0 à 100% comme pourcentages de pertes après défaut. Empiriquement, il est à noter que les moyennes observées de LGD oscillent autour de 40% et peuvent atteindre 60% en temps difficiles de récession. Ces distributions peuvent être bimodales (Asarnow & Edwards, 1995). Dans ce cas, la loi Bêta ne se prête pas à une calibration satisfaisante avec ses deux premiers moments comme il se fait classiquement. Il existe alors une approche alternative. Il s'agit de l'estimation Kernel de la densité (Hu et Perraudin, 2002) mais cela nécessite une transformation des observations avant de les utiliser. Renault et Scaillet (2003) proposent d'estimer le Kernel Bêta directement. Cependant, il y a encore des problèmes dans cette dernière méthode, comme le montre Gouriéroux et Monfort (2006). En parallèle de ces difficultés, il existe des pratiques qui considèrent que LGD est une simple constante fixée à sa moyenne historique, d'autres gardent la loi Bêta mais la calibrent en fixant un pourcentage de perte historique arbitraire. LGD étant un membre multiplicatif de l'équation (2.1), ces approximations sont plutôt surprenantes, voire paradoxales, puisque cela affecte la qualité de la modélisation finale de la distribution des pertes. Vu sous un autre angle, cela illustre la difficulté de la modélisation du crédit et de l'effort de recherche qui reste encore à déployer dans ce champ.

Quant à EAD , cette distribution est encore moins étudiée. Elle est souvent sup-

posée fixe et est forcée à sa moyenne historique, ou au mieux, supposée suivre une gaussienne inverse. Là encore, aucune intuition n'existe sauf qu'il convient d'obtenir des valeurs positives pour les montants issus de *EAD*. La relation entre *EAD* et les conditions économiques n'est pas évidente. Pour le crédit aux entreprises, Baker & Wurgler (2000) ainsi que Collin-Dufresne et Goldstein (2001) trouvent que *EAD* augmente si les conditions économiques sont difficiles, pour des besoins de liquidité. Au contraire, Mei et Saunders (1997) et Lown et Morgan (2001) montrent que les banques resserrent l'octroi des crédits dans les mauvaises conditions, ce qui impliquerait des *EAD* moins importants. Cette relation d'arbitrage complique la modélisation de cette variable pour le stress testing. Par ailleurs, des recherches montrent également que cette distribution est significativement affectée par les caractéristiques des crédits accordés en termes de maturité et des collatéraux présentés.

Par ailleurs, les trois distributions *PD*, *LGD* et *EAD* sont souvent considérées indépendantes par les modèles actuellement utilisés. Beaucoup de travaux tentent de prouver la dépendance entre *PD* et *LGD*. En effet et intuitivement, quand l'économie va mal, par exemple, il y a plus de défauts et moins de possibilités de récupérer l'argent prêté. Ceci est longuement discuté dans la littérature de procyclicité de *PD*, *LGD* et *EAD*, tel que fait par Allen et Saunders (2004). Ces faits sont importants pour les simulations de stress testing que nous comptons effectuer. Dans ce sens, nous retiendrons que, comme pour *PD*, *LGD* est aussi négativement corrélée aux conditions économiques. De plus, Frye (2000) trouve que *LGD* augmente de 20 à 25% dans les conditions économiques sévères. Ce fait chiffré est fort intéressant puisqu'il peut aider à prévoir la borne maximale du déplacement de la distribution *LGD* lors de nos simulations.

C'est dans ce contexte que nous commençons à situer l'article de Gouriéroux et al.

(2006)¹. Les auteurs précisent bien qu'il n'est pas nouveau de penser que PD et LGD sont liées, et que les chercheurs supposent généralement l'inverse seulement par souci de simplification et pour mieux se concentrer sur PD . L'apport de cet article est de voir la distribution LGD d'une nouvelle manière. En fait, les observations que possède une institution sur sa distribution LGD ne sont que la partie observable de LGD relativement aux défauts comptabilisés. Ce constat conduit directement à repenser LGD comme une distribution censurée par l'état de défaut. Par conséquent, une modélisation Tobit est plus appropriée comme réponse à ce genre de problématique. En d'autres termes, derrière la difficulté de modéliser LGD , il y aurait un problème technique important. En effet, il y aurait un biais si l'on essaie de définir LGD par ses seules valeurs observables en cas de défaut. Nous allons en voir quelques facettes dans le chapitre du modèle. Pour le moment, nous examinons les concepts relatifs à la distribution PD .

2.2 Les fondements des modèles théoriques

Dans cette partie, donnons-nous un aperçu sur les bases théoriques et historiques de la modélisation du risque de crédit. Comme déjà signalé, c'est la distribution PD qui est au centre des concepts. Il existe trois grandes familles. Merton (1974) ouvre le jeu avec son école de modèles dits structurels. Sous cette appellation, il convient de comprendre que le modèle interprète le sens économique de l'état de défaut ou de la dégradation de la qualité de crédit en général. Si l'on suppose que la valeur des actifs de l'emprunteur suit un processus aléatoire, l'état de défaut advient quand cette valeur passe en dessous d'un certain seuil. Dans ce cas, la valeur nette de l'emprunteur est nulle si ses actifs sont en deçà de la valeur faciale des dettes. Dans l'autre cas,

¹Cet article a été présenté à la 3ème conférence internationale sur le risque de crédit et opérationnel, HEC Montréal, 12 et 13 avril 2007

la valeur nette est la différence entre les actifs et la dette. Cela équivaut à dire que l'emprunteur possède une option d'achat sur ses propres actifs, avec un prix d'exercice qui est égal à sa dette. Cette similitude est à l'origine de la modélisation structurelle du risque de crédit.

Il existe une modélisation structurelle intéressante pour le cas des prêts aux particuliers. L'idée se base sur le fait que l'individu ne perd pas ses droits sur ses avoirs de la même manière que l'entreprise. Andrade et Thomas (2004) retracent cette discussion. Ils considèrent que c'est essentiellement l'accès au crédit que l'individu risque de perdre. Ils proposent alors de formuler ceci par une variable latente cohérente avec ce raisonnement pour remplacer celle de la valeur de l'actif de Merton (1974), et utilisent les informations disponibles aux bureaux de crédit.

La deuxième famille renferme les modèles dit de scoring depuis Altman (1968). L'approche est actuarielle. L'état de défaut est dû au hasard et aucune interprétation économique de l'état de défaut n'est requise. La troisième catégorie de modèles se dit hybride en englobant un dosage mixte des deux approches en même temps. Zhou (1997), Madan et Unal (1999) sont pionniers dans ce sens. Le premier greffe un processus de saut au modèle structurel classique, les derniers proposent de calibrer des modèles stochastiques exogènes par les informations comptables des entreprises. Duffie et al. (1999) étudient les différences fondamentales entre ces approches. Ce travail montre que c'est une question d'asymétrie d'information. Les modèles structurels supposent, en effet, que toute l'information est révélée par l'emprunteur. À l'opposé, les autres modèles rejettent cette hypothèse, et préfèrent se baser sur des faits statistiques observables. En outre, il est généralement admis que les modèles de scoring sont considérés plus précis à court terme, alors que les modèles structurels reprennent l'avantage à plus long terme.

En parallèle aux efforts de développement de ces écoles, certaines recherches explorent des pistes intéressantes telles que les fonctions de survie de Cox(1975). Ce concept s'adresse initialement aux études cliniques et pharmaceutiques pour modéliser la durée de vie d'un traitement ou d'un état en fonction du temps. Sa complexité en termes de mise en oeuvre est radicalement réduite par les travaux de Shumway (2001) qui lui donnent un second souffle dans la modélisation du risque de crédit. Depuis, ce terrain connaît de plus en plus de résultats importants, tel que nous allons le voir quand nous commencerons à construire notre modèle au chapitre 4.

Pour le moment, revenons à la problématique de notre portefeuille de prêts aux particuliers. Nous disons qu'une modélisation de type scoring serait plus adéquate, sachant que les horizons des prêts à la consommation sont généralement de court terme, et que la notion de valeur de marché n'existe pas pour les individus. Malgré cette différence conceptuelle de taille entre les individus et les entreprises, nous allons compléter cette partie par un tour afin de connaître les modèles industriels utilisés sur le marché pour les entreprises. L'idée est de tenter d'établir des similitudes qui nous seront utiles pour notre type de portefeuille de prêts aux particuliers.

2.3 Les principaux modèles industriels utilisés

Comme dit précédemment, les principaux développements des modèles ont été réalisés pour le risque de crédit aux entreprises. Crouhy et al. (2000) présentent une revue détaillée de ces modèles. Nous nommons CreditMetrics, MKMV-PortfolioManager, Standard & Poor's Portfolio Risk Tracker, CreditPortfolioView et CreditRisk+. Brannan et al. (2002) montrent que 80% des banques utilisent l'un de ces modèles au moins.

CreditMetrics est le premier à être publié en 1997. Il se base sur les cotations des entreprises du genre de Standard & Poor's. Les migrations entre les sept cotes outre

l'état de défaut sont prises en compte. On parle alors de modèle MM (Mark-to-Market) par opposition aux modèles DM (Default Mode) qui ne s'intéressent qu'aux deux états : défaut et non défaut. Une matrice de transition donne la probabilité de défaut. Une autre matrice donne les corrélations des transitions jointes. Ces deux matrices proviennent des fréquences historiques. Pour un portefeuille de deux obligations, par exemple, il existe $8^2 = 64$ possibilités de migrations conjointes. Dans ce cas, la distribution PD est discrète, avec 8^N points pour un portefeuille de N actifs. La distribution LGD est parfois approchée par une loi Bêta, supposée indépendante de PD . Ce modèle utilise le concept de Merton (1974) pour déterminer les corrélations entre entreprises de cotes différentes. Notons aussi que EAD est déterminé par calcul, en se basant sur les spreads de crédit des structures à termes des entreprises publiques. En effet, la dégradation de la qualité de crédit d'une obligation implique une baisse de sa valeur sur le marché.

Le deuxième modèle est le MKMV-Portfolio Manager. Il est du type DM. LGD est également une loi Bêta indépendante de PD . Dans un sens, ce modèle pourrait être vu comme une version simplifiée du premier. Son principal apport est qu'il est possible de le calibrer avec les probabilités de défaut et la structure de corrélation entre les cours des actions.

Portfolio Risk Tracker de Standard & Poor's est le troisième de notre liste. Il est de type MM en gérant les transitions. Il se veut dynamique par rapport à ses prédécesseurs, en calibrant les matrices de transition à partir des spreads, des cours des actions et des corrélations empiriques des états. Mais sa principale nouveauté est de modéliser directement la relation entre PD et LGD . Comme les deux autres modèles, Portfolio Risk Tracker se base sur les simulations Monte Carlo. Les distributions obtenues reflètent explicitement le caractère asymétrique, avec une queue gauche conforme à ce qui est généralement observé pour le risque de crédit.

Nous arrivons à CreditRisk+ de Credit Suisse Financial Products (1997). Il se contente d'être DM, mais il est le seul des cinq modèles énumérés ici à offrir une expression analytique pour déterminer le risque de crédit du portefeuille. Ceci est une bonne chose pour les portefeuilles à très grands nombres d'actifs où les simulations Monte Carlo seraient très coûteuses. Le modèle décompose le portefeuille en paliers de *LGD* comparables. *PD* est stochastique. Le modèle est calibré par les *EAD* des paliers, les taux de défaut, et enfin les volatilités des taux de défaut. Pour une catégorie donnée, la probabilité de défaut est approchée par une distribution Poisson de moyenne μ elle-même stochastique et distribuée suivant une loi Gamma.

Une critique commune aux modèles DM est qu'en ignorant les transitions de crédit, cela revient à sous-estimer le risque de crédit global. Par exemple, une obligation d'un rating BBB, ne présente que peu de risque de défaut pur. Cependant, le risque de dégradation de la qualité de son émetteur est tout à fait réel. Une autre critique importante relative à certains modèles qui utilisent les spreads des structures à terme. Ces modèles utilisent la totalité des spreads des structures à terme associés au risque de crédit pour évaluer le risque de défaut. Ce qui conduit à surestimer le risque de défaut, Dionne et al. (2007) ayant démontré définitivement que seulement 42% de ces spreads peut être imputé au risque de défaut.

Nous arrivons maintenant au Credit Portfolio View. Il est du type MM, sous sa forme générale. En réalité, c'est un type de modélisation à part, dit macroéconomique. Néanmoins, il peut être considéré comme structurel du moment qu'il explique les dégradations de crédit par l'impact des facteurs macroéconomiques explicitement intégrés dans le modèle. Les travaux de Wilson (1997a, b) servent de référence à cette modélisation. Ce modèle nous intéresse tout particulièrement. Il est le seul à permettre explicitement d'envisager des chocs macroéconomiques précis et d'évaluer leur

effet directement sur les distributions de pertes résultantes. Nous le retenons, pour le moment pour cette caractéristique très intéressante, quitte à en détailler le concept et examiner son adéquation pour notre type de portefeuille, dans la section suivante.

Notons à la fin de ce comparatif rapide, qu'en dépit de leur différence parfois radicale de point de vue méthodologie, il a été prouvé que ces modèles sont équivalents. Gordy (1998) le montre en examinant les structures mathématiques qui sont très proches de CreditMetrics et CreditRisk+, par exemple. Bucay et Rosen (2001) soulignent ce même point en proposant de vérifier la similitude des résultats pour un portefeuille de cartes de crédit. Ce papier indique que, s'il y a une différence dans les résultats obtenus, cela provient essentiellement des hypothèses de travail admises dans chacun des modèles. Or, les principales hypothèses faites concernent les distributions comme vu plus haut. Le message ici est que le choix du modèle à retenir doit nécessairement être motivé et accompagné par une utilisation raisonnable des hypothèses simplificatrices sur les distributions. Par conséquent, nous porterons un soin maximal aux choix de celles que nous prendrons.

2.4 Les approches stress testing du risque de crédit

Dans cette section, nous identifions des recommandations fondamentales quant à la modélisation du stress testing appliqué au risque de crédit. Nous nous faisons guider par Sorge (2004). Dans le document interne de Bâle #165, sur la classification des modèles de stress testing pratiqués, cet auteur souligne un fait important au sujet de certaines pratiques qui partent de trouvailles économétriques pour monter des modèles en pièces. Malgré des prédictions très performantes à court terme, ces modèles souffrent souvent d'instabilité dans le temps, par manque de fondements théoriques. Ce qui limite leur utilisation et réduit leur intérêt pour les grands horizons de temps. D'autre part, Sorge ressort le bien fondé des modèles basés sur Merton (1974) et Wil-

son (1999a, b) tout en exhibant leurs nombreuses adaptations.

Le modèle de Wilson (1997a,b) de base repose sur un système de trois équations telles que précisé par Sorge. La première modélise la relation entre le taux de défaut, $P_{i,t}$, et la notion d'indice de crédit, $Y_{i,t}$. Cet indice est donc défini comme l'inverse de la fonction logistique, *Logit*, du taux de défaut, avec changement du signe. L'indice de crédit est lié par la deuxième équation à un ensemble de facteurs macroéconomiques, $X_{k,t}$. La troisième écrit chacun des facteurs sous forme autorégressive univariée d'ordre p .

$$P_{i,t} = \text{Logit}(-Y_{i,t}) \quad (2.2)$$

$$Y_{i,t} = \beta_{i,0} + \beta_{i,1}X_{1,t} + \beta_{i,2}X_{2,t} + \dots + \beta_{i,m}X_{m,t} + \epsilon_{i,t} \quad (2.3)$$

$$X_{k,t} = \alpha_{0,k} + \alpha_{1,k}X_{k,t-1} + \alpha_{2,k}X_{k,t-2} + \dots + \alpha_{p,k}X_{k,t-p} + \nu_{k,t} \quad (2.4)$$

avec $i = 1, \dots, n$ catégories de crédit, et $k = 1, \dots, m$ facteurs économiques.

Les résidus ϵ et ν sont supposés indépendants et normalement distribués. P_{it} est le taux de défaut de la catégorie de crédit i à la période t . La variable P_{it} est transformée en Y_{it} par l'inverse de la fonction de répartition de la loi logistique avec inversion du signe. La variable Y_{it} ainsi définie représente l'indice du crédit comme appelée dans la littérature. L'équation (2.3) lie directement les facteurs de l'économie, X_{it} , aux indices de crédit Y_{it} . Classiquement, le nombre de retards p est fixé à 2. Bien que présenté sous forme d'un système d'équations, il est généralement considéré en estimant chacune des équations seule. À la fin de ces estimations individuelles une matrice $\Sigma_{(i+m)X(i+m)}$ est calculée comme matrice des variances-covariances du vecteur E construit par les résidus ϵ et ν tel que

$$E = \begin{bmatrix} \epsilon \\ \nu \end{bmatrix} \sim N(0, \Sigma) \text{ avec } \Sigma = \begin{bmatrix} \Sigma_\epsilon & \Sigma_{\epsilon\nu} \\ \Sigma_{\nu\epsilon} & \Sigma_\nu \end{bmatrix}$$

Le modèle considère que l'information sur les chocs exogènes est suffisamment représentée par la matrice Σ ainsi calculée.

Discutons maintenant de trois points primordiaux. Le premier est de rappeler que la conception initiale de cette modélisation vise à analyser le risque de crédit au niveau de l'ensemble de l'économie étudiée et non pour une institution financière en particulier. Si nous voulons l'utiliser ici, nous devons faire l'hypothèse que l'institution est de taille suffisamment importante pour que sa clientèle soit affectée par les conditions économiques quant aux aptitudes de paiement des échéances dues. Le deuxième point est le fait que ce modèle, appliqué tel quel sur les données d'une institution, donne des coefficients non significatifs voire contre-intuitifs pour la plupart des facteurs macroéconomiques choisis pourtant avec grand soin². Nous pensons à une piste pour expliquer ce problème bloquant. Il s'agit de l'existence d'une composante spécifique, trop importante pour l'ignorer par rapport à la composante systématique du risque de ce type de portefeuille. Des essais préliminaires font pencher le raisonnement vers cette voie. La présence importante de la composante spécifique des prêts aux particuliers est démontrée par de nombreuses études tel que fait par Dionne et al. (1996). À la base, ce travail traite du scoring des banques. Il démontre que des variables relatives aux prêts avec d'autres caractérisant les individus ont un pouvoir prédictif très grand pour la détermination de la probabilité de défaut des emprunteurs particuliers. Ce qui nous permet de creuser plus la question dans la section suivante.

²Mémoire de maîtrise HEC, Bouvrette J.-C. 2006, "Procédure de simulation de crise pour la détermination des fonds propres associés à un portefeuille de crédit d'une institution financière québécoise".

2.5 Les préliminaires de la modélisation à retenir

Nous connaissons mieux, maintenant, le terrain où nous avançons. Pour modéliser la probabilité de défaut, PD , nous avons dégagé les principaux points fondamentaux à satisfaire. Le modèle de Wilson (1997a,b) permet de les envisager d'une façon convenable. En effet, ce modèle intègre par construction les conditions économiques et permet une extension intéressante. L'idée est de tenter de capter simultanément la composante idiosyncratique et celle systématique dans la même équation (2.3). Nous en aurons fait un modèle hybride à partir du moment où les variables idiosyncratiques sont purement statistiques et n'ont pas toujours une explication causale pour justifier l'état de défaut.

Coleman et al. (2005), dans le document présenté à la conférence sur la supervision des banques du comité de Bâle, confirment l'idée de devoir capter la composante spécifique et proposent quelques variables idiosyncratiques à injecter dans le modèle de Wilson. Ce document concerne la modélisation du stress testing d'un portefeuille d'hypothèques en Australie. Il est intéressant de noter que ces auteurs proposent aussi de modéliser LGD de la même façon que le taux de défaut. Rodriguez et Trucharte (2006) se basent sur ce travail, et proposent de modéliser l'équation (2.3) directement au niveau de chaque individu plutôt que de travailler sur des taux de défaut agrégés.

Pour nous, sachant que chaque prêt de notre échantillon possède sa propre spécificité, nous allons travailler au niveau de chaque prêt sans faire aucune agrégation, ni au niveau des individus, ni au niveau des types de prêts. Cette manière de faire respecte aussi le principe qui consiste à contabiliser les défauts au niveau de chaque instrument de crédit directement. Analysons justement les données qui nous intéressent dans notre échantillon.

Chapitre 3

Les données

Notre base de données échantillon renferme quelque quinze mille observations, dont cinq mille environ concernent les prêts à la consommation sur lesquels nous focalisons désormais notre attention dans ce travail. Chaque prêt occupe une seule observation. Avec plus de trois cent variables résumant la vie de chaque prêt, cette base est normalement destinée aux analystes qui contrôlent et perfectionnent le système de scoring de l'institution.

D'autre part, l'échantillon contient les prêts de la période allant de septembre 1997 jusqu'à septembre 2006. En principe, ceci devrait nous donner une plage de dix ans. Cependant, pour des raisons que nous ne maîtrisons pas, nous n'avons des enregistrements de prêts en défaut qu'à partir de janvier 2003. Cela réduit considérablement la plage temporelle aux seules quatre dernières années. En fait, l'institution aurait opéré des changements importants dans ses systèmes informatiques. De ce fait, les données relatives à ces quatre dernières années s'apprêteraient mieux à notre recherche dans ce travail. Par ailleurs, l'institution les juge suffisamment représentatives, malgré leur nombre relativement réduit. Examinons maintenant les variables qui nous intéressent.

Pour chaque demande de prêt, l'institution calcule une cote de risque que nous

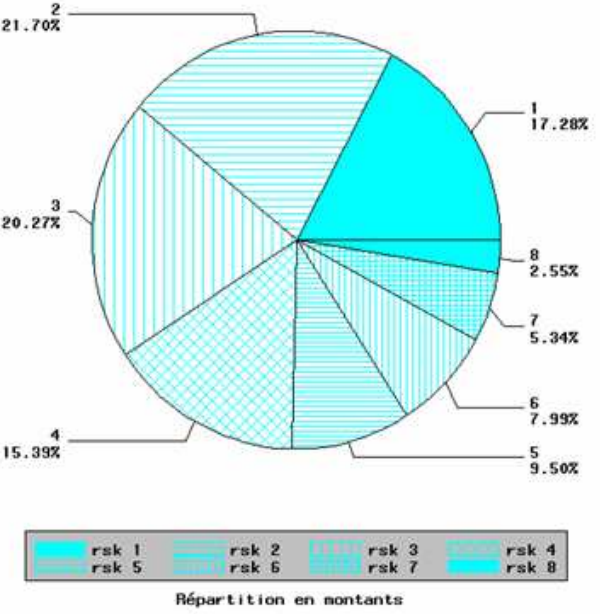
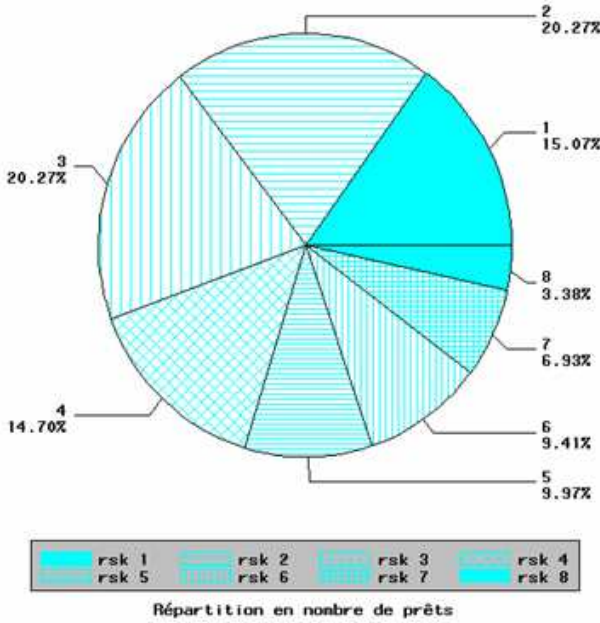


FIG. 3.1 – Répartition des prêts par cote de risque, en nombre de prêts en haut, et en montants attribués en bas.

nommons ici rsk . cette variable peut prendre une valeur de 1 à 8. La figure 3.1 donne la répartition des prêts par cote de risque. Quasiment 75% des prêts, en nombre et en montants, sont accordés aux risques 1 à 4. Les cotes 7 et 8, n'obtiennent que 8% des montants accordés, ce qui représente à peine 10% du nombre des prêts au total. La raison réside dans le fait que le taux de défaut est très fort parmi cette clientèle. La figure 3.2 représente les taux de défaut observés pour chaque cote de risque. Ce taux suit bien la cote de risque calculée au début du prêt, à l'exception de la cote 6 qui affiche une fréquence de défaut moins grande que pour les risques 4 et 5. À l'opposé, nous obtenons dans le même graphique que la cote de risque 8 affiche un taux de près de 11.5% des prêts accordés aux clients de cette cote, ce qui est très grand.

Analysons maintenant les défauts dans le temps avec la figure 3.3. L'évolution de la fréquence des défauts montre une allure très accidentée. Dans l'hypothèse que le portefeuille global reste sensiblement le même au cours du temps, les seules choses qui changent d'un trimestre à un autre sont les conditions économiques, outre le facteur temps lui-même. En effet, il est généralement admis que le comportement des emprunteurs change avec le temps quant à la possibilité de faire défaut. Par ailleurs, nous pouvons détecter grossièrement une tendance à la hausse des taux de défaut constatés dans ce graphique. Cette tendance pourrait être expliquée par une croissance que connaîtrait l'institution dans son portefeuille de prêts à la consommation. En effet, une analyse complémentaire montre qu'environ 300 nouveaux prêts sont injectés dans le portefeuille échantillon à chaque trimestre. D'autre part, à peu près 250 prêts quittent l'échantillon, par trimestres toujours. Ce qui donne une croissance trimestrielle de 50 prêts en moyenne. Pour nous dans ce travail, nous ignorerons cette croissance par simplification. Cependant, en toute rigueur, si cette tendance est confirmée par l'institution, elle devrait être explicitement prise en compte dans la modélisation finale du stress testing, puisque le taux de croissance du portefeuille de l'institution est une variable qui conditionne les prévisions futures.

Dans cette base, l'état de défaut lui-même est identifié par une variable précise. En préparation au chapitre 4 où nous construirons notre modèle, nous définissons plutôt la variable *Survie* qui représente l'état de non défaut. Cette variable est binaire et est égale à 1 tant que le prêt en question n'est pas en défaut, et passe à 0 si défaut. Historiquement, seuls 77 prêts sont tombés en défaut, dans tout l'échantillon qui renferme 5184 prêts, ce qui donne 1.49% de cas de défaut. Notons que ce pourcentage est faible relativement à ce qui est observable dans les institutions financières canadiennes. Cependant, nous comprenons que nous sommes dans le domaine du confidentiel, et que notre institution semble nous avoir transmis un échantillon représentatif mais avec des proportions connues par elle seule. Par conséquent, l'échantillon dont nous disposons sera considéré comme données d'une institution financière fictive, pour laquelle nous développerons notre méthodologie, la testerons empiriquement et interpréterons les conclusions.

Intéressons-nous maintenant à *LGD*. L'une des variables de la base approxime correctement le montant qui était exposé au risque de perte au moment du défaut. Nous l'appellerons *Expos* par la suite. Une autre variable, *Imp*, donne le montant final laissé impayé après les relances et les opérations de recouvrement. À partir de ce moment, nous sommes en mesure d'obtenir notre distribution *LGD* historique étant le rapport $Imp/Expos$. Le tableau 3.4 donne quelques chiffres clés décrivant ces différentes variables.

Nous arrivons maintenant à la distribution *EAD*. Pour les prêts aux consommateurs, il est généralement admis que *EAD* est prédéterminée. C'est tout simplement le solde du prêt en cours au moment du calcul (Gouriéroux et al., 2006).

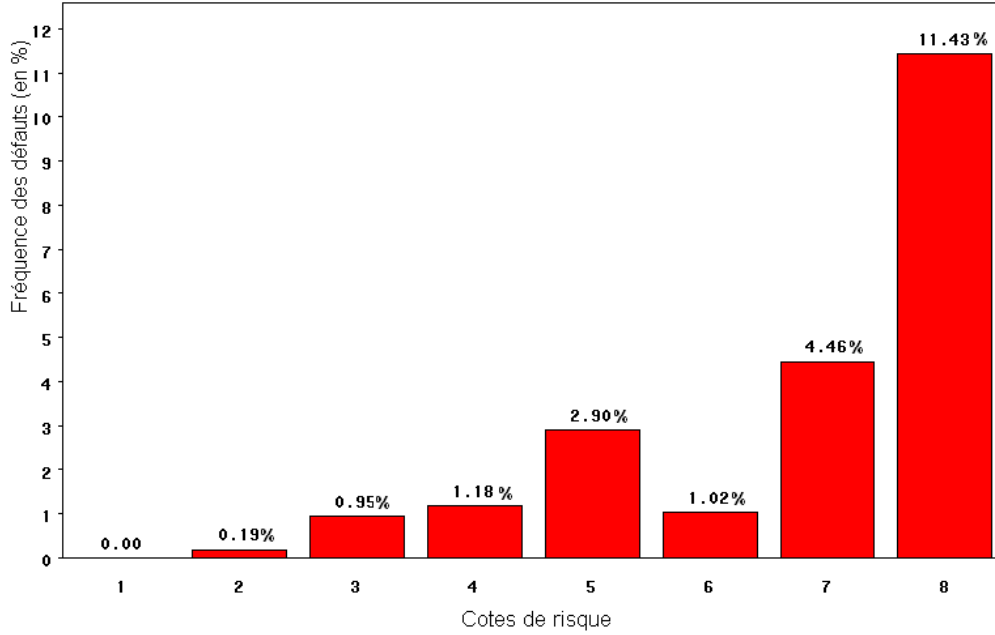


FIG. 3.2 – Répartition des défauts par cote de risque (en %)

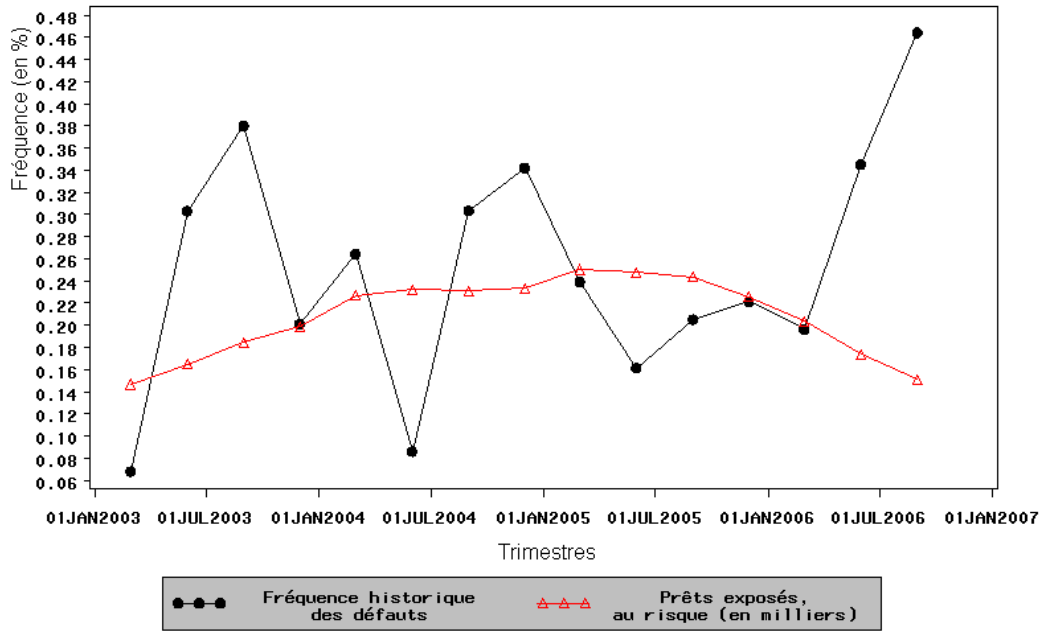


FIG. 3.3 – Répartition des défauts par trimestre

Précisons ici que Bâle II demande explicitement que l'institution modélise ses distributions *LGD* et *EAD*. En effet, *LGD* est propre à chaque institution découlant à la fois du comportement des clients sélectionnés par l'institution, d'une part, et de la performance de ses services de recouvrement, d'autre part. *EAD* est aussi en grande partie spécifique à l'institution puisqu'il est lié aux performances du système de sélection des clients. Les paragraphes 468-470 de Bâle II incitent les institutions à étudier leurs distributions *LGD* et à en maîtriser les interdépendances avec les variables spécifiques des prêts, ainsi que par rapport aux cycles économiques. Concernant *EAD*, ce sont les paragraphes 474-476 qui précisent les exigences en matière de connaissance de cette distribution avec les facteurs qui en influencent l'évolution.

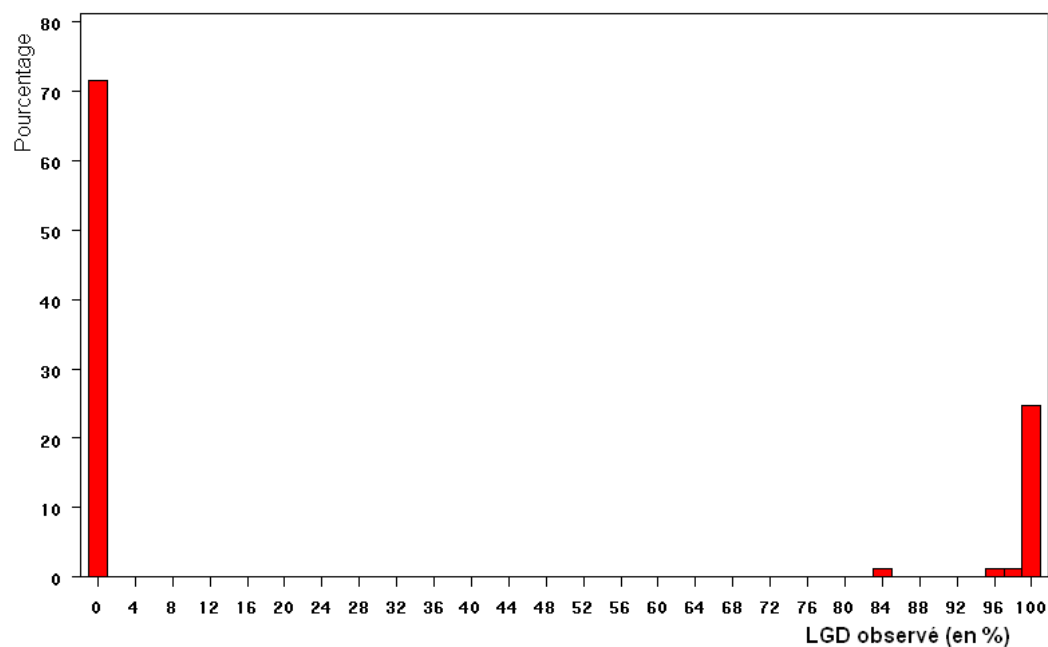
Pour revenir à la distribution *LGD*, examinons la figure 3.4. Nous avons 71% des

TAB. 3.1 – Statistiques descriptives de *LGD* et *EAD*

	#	Montants (en \$)		Écart- type	Valeurs (en \$)		Somme
		Moyenne	Médiane		minimum	maximum	
<i>Expos</i>	77	5628	2949	5814	0	18878	433339
<i>Imp</i>	77	2119	0	4621	0	18878	163173
<i>LGD</i> en %	77	28.2%	0	45%	0%	100%	-
<i>EAD</i>	1374	8121	6036	9666	180	201539	11159262

cas où le recouvrement arrive à récupérer 100% des soldes au moment des défauts. Dans les 29% des cas restant, l'institution affiche une perte de montants allant de quelques centaines de dollars à 20 mille dollars environ. La moyenne de cette perte est de 2100\$, l'écart-type est de 4600\$ avec une perte totale de 163000\$ durant les quatre années étudiées dans l'échantillon (2003 à 2006).

Retenons maintenant que l'investissement total en cours de l'échantillon s'élève à

FIG. 3.4 – Distribution historique de *LGD*

TAB. 3.2 – Évolution de la fréquence des défauts et des pertes

Trimestres	Nombres		Montants		LGD en %
	Prêts	Défauts	Exposés au moment du défaut en\$	Pertes finales en \$	
Mar-03	1469	1	15890	0	0.0
Jun-03	1649	5	43667	574	1.3
Sep-03	1842	7	15616	9292	59.5
Dec-03	1987	4	30648	0	0.0
Mar-04	2269	6	35892	0	0.0
Jun-04	2323	2	8383	0	0.0
Sep-04	2308	7	43162	41448	96.0
Dec-04	2338	8	54129	1606	3.0
Mar-05	2507	6	13815	0	0.0
Jun-05	2479	4	13745	0	0.0
Sep-05	2436	5	9495	0	0.0
Dec-05	2259	5	30613	13533	44.2
Mar-06	2039	4	33718	25803	76.5
Jun-06	1738	6	38264	33720	88.1
Sep-06	1508	7	46302	37198	80.3

11 millions de dollars environ. Quand nous calculerons les distributions de pertes, ce chiffre donnera la base de l'exposition totale au risque, pour déterminer les pertes en pourcentage.

Il nous reste à présenter les variables macroéconomiques dont nous comptons explorer les relations avec les défauts des emprunteurs. Nous retenons six candidates (Tableau 3.3) : le taux de chômage, l'épargne, l'indice des prix des biens semi-durables, le PIB réel, le taux de change par rapport au dollar américain, et le taux directeur de la Banque du Canada. Nous justifierons leur choix dans le chapitre suivant. Nous y verrons aussi que nous nous intéressons plus particulièrement aux variations de ces facteurs économiques plutôt qu'aux valeurs en niveau. Ces variations seront désormais nommées respectivement : $dChom$, $dEparn$, $dIPsem$, $dPIBr$, $dChng$ et $dBanqCan$. Le tableau 3.4 en résume les caractéristiques principales. Enfin, rappelons que pour les calculs du stress testing, toutes les observations de ces séries nous intéressent. Ceci est dans le sens où nous ne les filtrerons pas puisque nous voulons, justement, capter les événements exceptionnels de leur historique dans notre modélisation.

TAB. 3.3 – Description des variables macroéconomiques retenues

nom	Designation	Source	Code	Fréquence
Chom	Canada, Taux de chômage 15 ANS ET PLUS ; désaisonnalisées de janvier 1990 à septembre 2006	CANSIM II	V2062815	Mensuelle
Eparn	Canada, Epargne des particuliers et entreprises individuelles, désaisonnalisées de 1er trimestre 1990 au 3ème trimestre 2006	CANSIM II	V498490	Trimestre
IPsem	Canada, indice des prix à la consommation, contenu du panier de 2001, biens semi-durables de janvier 1990 à septembre 2006	CANSIM II	V735591	Mensuelle
PIB	Canada, Produit Intérieur Brut au prix de base de 2001 de 1er trimestre 1990 au 3ème trimestre 2006	CANSIM II	V1992226	Trimestre
Chng	Canada, Taux de change du dollar des Etats-Unis, cours au comptant à midi, moyenne de janvier 1990 à septembre 2006	CANSIM II	V37426	Mensuelle
BnqCan	Canada, Taux officiel d'escompte de janvier 1990 à septembre 2006	CANSIM II	V122530	Mensuelle

TAB. 3.4 – Statistiques descriptives des variables macroéconomiques retenues

Variables en variation	Trimestres #	Valeurs		Écart- type	Valeurs	
		Moyenne	Médiane		minimum	maximum
dChom (en %)	68	0.09	-0.06	0.98	-0.63	7.63
dEparn (en 10 milliers \$)	68	0.02	-0.08	0.90	-1.59	5.98
dIPsem (en \$)	68	1.52	0.10	10.93	-1.23	89.93
dPIBb (en 10 milliers \$)	68	0.50	0.32	1.94	-1.09	15.03
dChng (en %)	68	0.02	0.00	0.15	-0.11	1.18
dBnqCan (en %)	68	0.07	-0.04	1.77	-1.95	12.97

Chapitre 4

Le modèle

Rappelons que nous cherchons à modéliser la distribution des pertes non anticipées, L , qui est elle-même le produit des trois distributions PD , LGD et EAD (équation 2.1). D'autre part, nous voulons considérer la probabilité de défaut par instrument de crédit. En d'autres termes, nous comptabiliserons les situations de défaut et de non défaut au niveau de chaque prêt. Par ailleurs, à la fin du chapitre 2, nous avons retenu le modèle de Wilson comme base théorique de notre développement. Nous allons l'adapter ici à notre problématique de prêts aux particuliers. Ensuite, nous passerons à LGD et EAD , pour accéder aux simulations Monte Carlo.

4.1 Distribution PD

Nous nous intéressons au système d'équations du modèle de Wilson. Nous suggérons de le traiter en deux sous-systèmes, au lieu de le considérer comme ensemble d'équations indépendantes les unes des autres comme fait classiquement. L'intérêt est de tenter d'en préserver le sens systémique afin de bénéficier des interactions mutuelles qu'il contient. Pour ce faire, nous faisons l'hypothèse que les défauts constatés au niveau de notre institution financière ne sont pas de taille à influencer les variables macroéconomiques du pays. Ceci nous permet de regrouper ensemble les équations

2.2 et 2.3 d'un côté, et les équations 2.4 - une équation par facteur- de l'autre. Les équations 2.4 formeront alors un système VAR(p) plutôt qu'un ensemble d'équations AR(p) univariées indépendantes. L'intérêt d'un VAR(p) est double. D'abord, il modélise explicitement les interactions entre les facteurs macroéconomiques. Le deuxième gain est d'obtenir une matrice des variances-covariances d'une façon directe après l'estimation du VAR(p), ce qui est plus précis que ce que propose le modèle initial de Wilson¹. Nous allons prendre un VAR(2), comme détaillé dans l'annexe B, dont le nombre de paramètres à estimer reste acceptable par rapport au nombre de trimestres dont nous disposons. Rappelons ici que les informations sur les chocs exogènes sont contenues dans cette matrice variances-covariances. Par conséquent, la précision apportée par le VAR(2) est une première amélioration par rapport au modèle initial de Wilson.

D'autre part, nous réécrivons les équations 2.2 et 2.3 en une seule équation sous la forme

$$PD_{it} = \text{Logit}(\alpha' Age_{it} + \beta_1' Indiv_i + \beta_2' Macro_t) \quad (4.1)$$

PD_{it} est notre variable dépendante dichotomique, portant la valeur 1 si défaut et 0 sinon. Les variables du bloc Age_{it} capteront l'effet de l'âge du prêt sur la probabilité de défaut. Le vecteur $Indiv_i$ contient les variables caractérisant le risque spécifique du prêt i . Le bloc $Macro_t$ contiendra le vecteur des variables macroéconomiques pertinentes pour expliquer l'état de défaut. α , β_1 et β_2 sont les vecteurs de paramètres de la modélisation à estimer pour chaque bloc.

En fait, au moment de la demande de prêt, l'institution calcule la cote de risque tenant compte de toutes les variables personnelles et celles du prêt demandé qu'elle pense pertinentes, dans son système scoring. Nous mettons cette variable, rsk , seule pour représenter le bloc $Indiv_i$, et nous justifions ce choix par le fait que l'institution

¹Voir la discussion de la section 2.4

calcule $risk$ d'une façon optimale, afin de capter le risque spécifique, puisqu'elle l'utilise pour statuer sur l'octroi des prêts. Quant aux variables macroéconomiques pertinentes, une fois identifiées, nous supposons qu'elles représenteront suffisamment le risque systématique pour dire que les probabilités jointes de défaut sont déjà incorporées dans la formulation de PD ainsi obtenue. Il nous restera l'effet de l'âge des prêts sur la probabilité de défaut des prêts. Ne connaissant pas, a priori, sa fonction exacte, il est généralement conseillé de la modéliser en polynôme d'ordre 2, 4 ou 6 (Glennon et Nigro, 2005). Nous nous contenterons ici d'une modélisation en polynôme d'ordre 4, dont les coefficients seront déterminés et analysés lors de l'estimation de l'ensemble.

Nous devons discrétiser la vie de chaque prêt en segments de temps égaux. Intuitivement, ce découpage est nécessaire afin d'isoler le moment de défaut, avec toutes les caractéristiques qui changent avec le temps. Vu que la plupart des variables macroéconomiques ont une fréquence trimestrielle, nous procéderons à un découpage par trimestre. Par ailleurs, les durées à l'intérieur de chaque trimestre seront mesurées en nombre de mois complets. Ainsi, un prêt qui démarre en mars 2001 aura un âge de 4 mois au bout du deuxième trimestre de l'année 2001. Ce prêt aura autant d'observations dans la base de données discrétisée qu'il aura vécu de trimestres ou parties de trimestres. La dernière observation sera relative au trimestre où le prêt arrive à échéance, tombe en défaut, ou alors c'est le dernier trimestre de la période l'échantillonnée.

C'est ici que nous commençons à nous éloigner des modèles Logit ou Probit monopériodiques simples très populaires pour modéliser les probabilités de défaut. Notons que les observations ainsi démultipliées ne seront pas indépendantes les unes des autres, ce qui pourrait caractériser l'effet panel aussi bien individuel que temporel dans nos données (Amemiya 1985). En effet, le même prêt va apparaître autant de fois que de trimestres vécus. D'autre part, chaque trimestre sera représenté plusieurs

fois dans cette décomposition. La conséquence est que nos estimations pourraient ne pas être efficaces, ce qui conduirait à des conclusions erronées. Shumway (2001) est le premier à résoudre cette situation d'une façon simple. En effet, dans son article, il montre l'équivalence mathématique entre le Logit multi-périodique du genre de ce que nous avons, avec l'état de défaut comme variable dépendante dichotomique, et la fonction de survie de Cox (1975) en modélisation discrète. Il démontre que les fonctions de vraisemblance sont les mêmes, et donc les estimations doivent être les mêmes.

Shumway propose ensuite une correction des statistiques χ^2 des coefficients estimés. Cette correction est la seule contribution systématiquement et explicitement rejetée par tous les chercheurs qui ont utilisé les résultats de Shumway, dont Chava et Jarrow (2004). L'argument principal que ces auteurs avancent est que le Logit ainsi estimé est encore pénalisé par l'effet panel et qu'en toute rigueur il est possible de le corriger. Cependant, la modélisation de Cox (1975) ne suppose aucune indépendance entre les observations étudiées. De ce fait, tous les résultats de l'estimation Logit sont considérés corrects y compris le calcul des χ^2 des coefficients estimés. Sachant ceci, nous allons admettre à notre tour ce raisonnement pour nos estimations Logit que nous utiliserons sans correction.

Cette discussion mise à part, Shumway réduit considérablement et définitivement la difficulté théorique de la modélisation de Cox, et offre ses grands avantages, au prix d'une démultiplication des observations comme la nôtre, de l'injection d'une fonction de durée dans le modèle (polynôme de l'âge des prêts pour nous), et d'une simple régression Logit de l'ensemble. Au passage, dans ce même article rempli de résultats importants, Shumway montre que la modélisation Logit en mono-période simple est biaisée malgré sa grande popularité, et calcule mathématiquement ce biais. En effet, en n'étudiant que les prêts terminés (clôturés normalement ou par un défaut), le modèle Logit simple ignore totalement les prêts exposés aux mêmes risques, pendant la

même période d'étude, mais qui courent toujours à la fin de la période échantillonnée. Shumway souligne, enfin, la supériorité conceptuelle de la modélisation discrète de Cox, et surtout la stabilité de ses estimations, puisque toute l'information disponible est prise en compte.

Le papier de Chumway devient une référence en la matière. De plus en plus de travaux confirment et prolongent les résultats de Shumway dans la modélisation du crédit. Nous pourrions citer Du(2003) qui modélise les migrations des cotes de crédit des entreprises. Chava et Jarrow (2004) confirment les résultats de Shumway et montrent, entre autres, qu'une discrétisation du temps dans une base mensuelle est nettement meilleure qu'un pas annuel par exemple. Citons encore Drehmann et al. (2005) qui utilisent explicitement une modélisation très proche à la nôtre, destinée au stress testing appliqué aux entreprises publiques, avec des variables macroéconomiques. Leur idée est de proposer des développements non linéaires très intéressants. Ils argumentent que les modèles linéaires supposent des variations petites, alors que les chocs à étudier en stress testing sont, justement, grands par définition. Bien que très convaincant, ce travail dépasse largement le cadre de notre étude.

Désormais, nous ferons une lecture des résultats en termes de fonction de survie. Nous parlerons de la probabilité de défaut trimestrielle conditionnelle à la survie du prêt jusqu'au trimestre en question. Elle se définit par

$$\Lambda(t) = \text{Logit}(g(\text{Age}_{it}) + \beta'_1 \text{Indiv}_i + \beta'_2 \text{Macro}_i) \quad (4.2)$$

où $g(\cdot)$ est la fonction qui modélise l'effet de l'âge des prêts, le polynôme de degré 4. Nous disposons alors des égalités de Cox en modélisation discrète

$$S(t) = \prod_{j < t} (1 - \Lambda(j)) \quad (4.3)$$

où $S(t)$ est la fonction de survie, qui est la probabilité de survie jusqu'à la période t .

Afin de fixer les idées, calculer la probabilité de survie au bout d'un an revient à calculer la probabilité de survie jusqu'au début du trimestre 5 :

$$S(5) = (1 - \Lambda(1))(1 - \Lambda(2))(1 - \Lambda(3))(1 - \Lambda(4)).$$

En fait, les calculs probabilistiques reviennent à déterminer la fonction Λ pour tous les trimestres consécutifs qui constituent la période souhaitée, en tenant compte des prédictions des variables macroéconomiques qui changent d'un trimestre à un autre. Pour simplifier, nous nous contenterons d'effectuer les calculs pour un seul trimestre, dans le cadre du présent travail.

Pour finir cette sous-section, nous nous proposons de présenter la liste des variables explicatives candidates dont nous tenterons de mesurer l'effet sur PD , dans le chapitre suivant. Le tableau 4.1 les récapitule avec les prédictions des signes attendus. La cote de risque, rsk , devrait avoir un signe positif. Plus la cote de risque est élevée plus il y a de chances que le prêt tombe en défaut. Nous pensons également que cette variable devrait être très significative pour expliquer PD , puisque c'est l'évaluation de la qualité de crédit du prêt au moment de son octroi. L'âge du prêt à la période t , age , devrait avoir un signe positif. Le défaut ne survient que rarement aux premières périodes des prêts. Concernant les variables age^2 et age^3 , a priori, aucune prédiction intuitive possible. Cependant, nous pourrions attendre un signe négatif pour age^4 qui signifierait que les défauts sont rarement observés aux dernières périodes des prêts. En effet, si la valeur de age est suffisamment grande, les trois premiers termes du polynôme de l'âge du prêt, age , age^2 et age^3 deviennent négligeables devant age^4 car élevé à la puissance 4. Donc, le signe négatif pour cette variable devrait permettre que la probabilité de défaut diminue à partir d'un âge donné pour tendre vers zéro à la fin de la vie des prêts.

La variation du taux de chômage, $dChom$, est attendue avec un signe positif. Si le taux de chômage augmente, il y a plus de chance que cela impacte directement ou indirectement la population des emprunteurs. Encore un signe positif à prédire pour $dIPsem$. La hausse des prix des biens semi-durables qui est, en effet, négative pour le pouvoir d'achat des emprunteurs, et donc devrait affaiblir leur capacité à payer les mensualités de leurs prêts. Pour l'épargne, le raisonnement concernerait le surplus de la consommation par rapport au revenu. Ce surplus traduirait une disponibilité des fonds qui, quand elle augmente, impacte positivement la survie des prêts et vice-versa. D'où nous attendons un signe négatif pour la variable $dEparn$. Une mauvaise conjoncture économique, donc baisse du PIB réel $dPIB$, devrait faire accroître les défauts, d'où un signe négatif.

Une variation positive du taux de change par rapport au dollar américain, $dChng$, signifierait que le dollar canadien baisse. Si nous prenions ceci comme signe de faiblesse relative de l'économie canadienne par rapport aux É.-U., le signe prédit pourrait être négatif. Cependant, nous savons aussi que si le dollar canadien se déprécie, l'exportation aux É.-U. en bénéficie, ce qui devrait à son tour améliorer l'économie canadienne. De ce fait, le signe prédit est ambigu pour la variable $dChng$.

Nous arrivons enfin au taux d'escompte de la Banque du Canada. Si ce taux augmente, l'argent devient plus cher à avoir, d'où un pouvoir moindre pour rembourser les mensualités. Ceci nous guide pour prédire alors un signe positif pour la variable $dBnqCan$.

Précisons ici que nous utilisons les variations des variables macroéconomiques plutôt que leurs valeurs en niveau. L'argument en est que, ces variables ne sont pas stationnaires et possèdent un trend tangible avec le temps. Ceci rend ces variables fortement corrélées entre elles, et donc difficilement utilisables directement comme variables explicatives. La même raison existe pour ses variables avec la fonction de durée, le polynôme d'ordre 4, que nous modélisons explicitement parmi les variables

TAB. 4.1 – Signes prédits pour modéliser la probabilité de défaut PD

Variable	Signes	
	prédits	Commentaires
rsk	+	cote de risque du prêt
<i>age</i>	+	âge du prêt à la période t
age^2	?	carré de la variable <i>age</i>
age^3	?	<i>age</i> à la puissance 3
age^4	-	<i>age</i> à la puissance 4
dChom	+	variation du taux de chômage
dIPsem	+	variation de l'indice des prix semi-durables
dEparn	-	variation de l'épargne
dPIBb	-	variation du PIB réel
dChng	-	variation du taux de change (USD)
dBnqCan	+	variation du taux d'escompte

explicatives. Nous poursuivons maintenant avec l'étude des deux autres distributions LGD et EAD .

4.2 Distributions LGD et EAD

Commençons par la distribution EAD . La majorité des concepteurs la considèrent comme prédéterminée pour les portefeuilles de prêts à la consommation des particuliers (Gouriéroux et al, 2006). EAD est le solde du compte de l'emprunteur au moment du calcul. Il existe une exception à cela pour les marges de crédit et les prêts sur cartes de crédit que nous ne traitons pas dans ce travail².

La figure 3.4, du chapitre précédent, trace la distribution historique de LGD . Une série d'essais effectués sur cette distribution tentent de confirmer qu'elle ne serait pas

²Dans les cas des marges de crédit et des prêts sur cartes de crédit, EAD à prendre est le plafond maximal autorisé. En effet, les emprunteurs peuvent à tout moment prendre la totalité du montant que leur fixe l'institution financière

facilement modélisable. En effet, $LG D$ ne dépendrait d'aucune de nos variables macroéconomiques. Elle ne serait pas non plus sensible aux paramètres spécifiques des prêts, dont la présence ou non d'un collatéral.

Une autre piste serait d'explorer directement la distribution composite $LG D * EAD$ comme fait dans certaines modélisations (Moody's pour les entreprises publiques). En fait, $LG D * EAD$ est le solde impayé des emprunteurs en défaut. La figure 4.1 dessine cette distribution pour le cas de notre base échantillon. Elle montre un point de masse à zéro de 70 % environ, suivi d'une plage de montants que nous pouvons considérer aléatoires et uniformément distribués de 0 à 20000\$. Il y a deux inconvénients majeurs à cette approche. D'abord, cela ne permettrait pas d'assurer que EAD soit égale au solde des prêts tel que précisé au début de cette sous-section. Le deuxième problème est que des tentatives de modélisation économétrique n'ont pas été plus chanceuse que celles de $LG D$. L'unique possibilité qui s'offre serait de modéliser $LG D * EAD$ "par pièce". Nous pourrions faire cela en deux étapes. Générer une première variable aléatoire qui suit une distribution binomiale avec 70% des tirages de 0. Les 30% des tirages à 1 seraient multipliés par les valeurs d'une deuxième variable aléatoire uniformément distribuée entre 0 et 20000. Nous voyons qu'en procédant ainsi, nous perdons toute la finesse de pouvoir relier le comportement de $LG D$ aux paramètres idiosyncratiques des emprunteurs, et surtout que des simulations par les conditions économiques n'auraient pas été possibles.

C'est maintenant que nous faisons appel à l'intuition de Gouriéroux et al (2006). Ces auteurs présentent leur idée de modélisation conjointe de PD et $LG D$ dans deux cas. Le premier reflète typiquement le nôtre puisqu'il s'adresse aux modèles Logit-Probit. Le deuxième concerne une extension du modèle de Merton, s'adressant généralement aux entreprises, et que nous n'utilisons pas ici. La contribution majeure de ces auteurs est de dire que les institutions, quand elles enregistrent leurs taux de pertes des prêts,

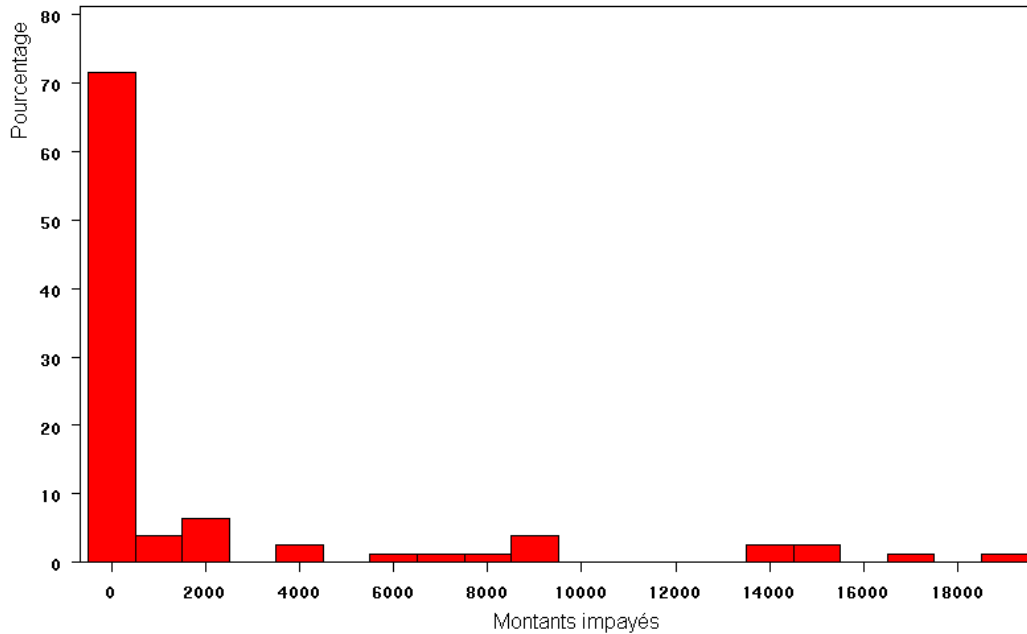


FIG. 4.1 – Distribution des pertes historiques : montants laissés impayés

ceci n'est que la partie observable d'une distribution LGD^* qui serait censurée par l'état de défaut. En d'autres mots, cela revient à réécrire le problème de LGD sous forme d'un Tobit généralisé

$$LGD_i = \begin{cases} LGD_i^* & \text{si } PD_i = 1 \quad (\text{si défaut}) \\ 0 & \text{sinon} \end{cases} \quad (4.4)$$

où LGD_i^* est la distribution totale dont nous observons les LGD_i en cas de défaut.

Avant de passer au chapitre suivant, soulignons que cette manière de voir LGD devrait normalement aider à élucider la difficulté de modélisation de cette distribution comme mentionné plus haut. Dans l'hypothèse que les services de recouvrements de l'institution font un effort maximal pour tenter de récupérer l'argent impayé, la distribution LGD devient caractérisée uniquement par l'emprunteur et les conditions économiques. Ceci étant réglé, nous passons aux signes prédits. Commençons par les variables macroéconomiques pour dire que c'est exactement la même intuition que

pour modéliser PD . En effet, c'est le même individu qui fait défaut qui ne pourra pas payer LGD . Par contre, il est très possible que les variables qui seront significatives pour expliquer la probabilité de défaut PD ne soient pas les mêmes pour le taux de perte LGD . La cote de risque rsk devra avoir un signe positif. Intuitivement, plus l'emprunteur est du type mauvais risque plus grande sera la proportion du montant impayé qu'il laissera. La nouvelle variable ici est Gar qui est binaire (1 pour un prêt avec présence de garantie, 0 sinon). Dans ce cas, nous pourrions envisager que la présence de la garantie devrait alléger LGD . D'où un signe attendu négatif. L'âge du prêt pourrait impacter LGD , cependant, il n'est pas facile d'en prévoir le sens.

Il nous reste à donner une précision concernant ce modèle de Gouriéroux et al. Afin de garantir que les valeurs issues des calculs sur LGD restent positives, les auteurs modélisent directement la variable $Log(LGD)$. Ceci assure en effet la positivité des valeurs de LGD par l'utilisation de la fonction exponentielle pour inverser la fonction logarithmique. Gouriéroux et al. suggèrent aussi de ne pas limiter à 100% la borne supérieure de l'intervalle de LGD . L'argument est que dans la réalité des taux de pertes supérieurs à 100% existent dans le cas de frais conséquents que dépense l'institution après le défaut de l'emprunteur.

TAB. 4.2 – Signes prédits pour modéliser *LGD*

Variable	Signes	
	prédits	Commentaires
rsk	+	
Gar	-	binaire : 1 si présence d'une garantie pour le prêt, 0 sinon
age	?	âge du prêt à la période t
dChom	+	variation du taux de chômage
dIPsem	+	variation de l'indice des prix semi-durables
dEparn	-	variation de l'épargne
dPIBb	-	variation du PIB réel
dChng	-	variation du taux de change (USD)
dBnqCan	+	variation du taux d'escompte

Chapitre 5

Résultats et analyse

5.1 La distribution PD à la Shumway

Le tableau 5.1 montre les résultats des régressions Logit multi-périodes. La cote de risque, rsk , est très significative, ainsi que l'âge du prêt, age , et age^2 qui est le carré de l'âge. Les signes des coefficients sont conformes à l'intuition décrite dans le chapitre précédent. En effet, la coté de risque est positivement très significative. L'augmentation du taux de chômage affecte PD en l'aggravant, $dChom$ étant significative quasiment à 5%. L'augmentation des indices des prix des biens semi-durables est négative pour le pouvoir d'achat des emprunteurs comme déjà vu. Ce qui confirme le signe positif obtenu. Pour l'épargne, le surplus du revenu par rapport à la consommation, impacte positivement la survie des prêts, d'où le signe négatif du coefficient correspondant. Notons ici que ni le taux d'escompte, ni le taux de change ne sont pas significatifs même à 30%. La variable $dPIB$, bien que très peu corrélée avec les autres variables (table 5.2), nous ne la mettons pas dans la régression Logit, et elle jouera son rôle comme expliqué dans le paragraphe suivant. Une dernière remarque concernant les variables macroéconomiques choisies. Lors de cette étude, nous n'avons pas la prétention d'être exhaustifs dans la sélection de ces variables. Cependant, nous avons mis dans le modèle celles que nous pensons les plus pertinentes tout en privilégiant

le fait de garder leur nombre le plus petit possible. Ce choix est justifié par application du principe de parcimonie. Une deuxième raison est d'ordre économétrique. En fait, si nous voulions étudier un système VAR(2) pour 10 variables, par exemple, cela nécessiterait 265¹ paramètres à estimer pour un tel système. Ce qui serait aberrant comme nombre de degrés de liberté par rapport au nombre d'observations à étudier.

TAB. 5.1 – Estimation du Logit multi-périodique

Variable	Signe attendu	Estimation pour 2003-2006			Estimation	
		Coefficients	Chi-Carré	Pr > Chi2	2003à2004	2003à2005
Intercept		-11.42370	120.348	<.0001	-11.80460	-12.57190
rsk	+	0.58880	89.0078	<.0001	0.61520	0.64260
age	+	0.62800	5.8965	0.015200	0.73270	0.87930
age²	?	-0.04430	3.8236	0.050500	-0.06040	-0.07140
age³	?	0.00131	2.8382	0.092000	0.00210	0.00242
age⁴	-	-0.00001	2.3543	0.124900	-0.00002	-0.00003
dChom	+	1.21600	3.6506	0.056000	0.82440	1.05640
dIPsem	+	0.29430	2.2435	0.134200	0.45000	0.32850
dEparn	-	-0.25040	1.8173	0.177600	-1.01430	-0.49380
# d'observations		31151			13847	23607
# de prêts		5184			3620	4823
-2 Log L		948.314			387.131	666.370

Nous détaillons la modélisation du système d'équation 2.4 en VAR(2) dans l'annexe B. Notons ici que nous y étudions un système de quatre séries. Il s'agit des trois variables que nous avons déjà décrites précédemment. La quatrième variation est le PIB réel. Nous nous intéressons à cette dernière pour les simulations de récession en

¹ Si n est le nombre des variables, un VAR(p) nécessite (n*p+1)*n coefficients pour les n équations du système, plus n*(n+1) pour estimer les éléments de la matrice symétrique des variances-covariances des résidus.

TAB. 5.2 – Corrélations entre les variables explicatives

	rsk	age	dChom	dIPsem	dEparn	dPIB
rsk	1	-0.0074	0.0038	-0.0020	0.0015	0.0032
age	-0.0074	1	-0.0197	-0.0331	-0.0046	0.0314
dChom	0.0038	-0.0197	1	-0.1315	0.0092	0.1804
dIPsem	-0.0020	-0.0331	-0.1315	1	-0.0794	-0.0279
dEparn	0.0015	-0.0046	0.0092	-0.0794	1	-0.1238
dPIB	0.0032	0.0314	0.1804	-0.0279	-0.1238	1

stress testing. Nous laisserons cette variable en dehors du Logit, sachant qu'elle jouera son rôle en interaction avec les trois autres variables dans le VAR(2).

Le tableau 5.1 des estimations de notre modèle Logit donne aussi les coefficients pour trois périodes. Outre la période de l'étude, 2003 à 2006, nous avons également les coefficients de 2003 à 2004 et de 2003 à 2005. Nous pouvons remarquer que ces coefficients sont passablement stables dans le temps. C'est ce que tentent de confirmer les tests économétriques réalisés sur ces estimations, sauf pour age^3 et age^4 . Ces dernières variables mises à part, l'ensemble des résultats illustrerait la supériorité conceptuelle de la modélisation de Cox telle que démontrée par Shumway(2001). Cette stabilité du modèle est primordiale pour l'adéquation des simulations en hors échantillon que nous comptons faire.

Nous nous intéressons, maintenant, à la fonction polynôme de l'âge des prêts que nous avons obtenue. Avec tous les coefficients du polynôme significatifs, il est plus aisé de tracer un graphique pour l'analyser. La figure 5.1 représente l'évolution de la probabilité de défaut pour chaque cote de risque, de 1 à 8. Nous voyons qu'il existe une sorte de palier pour lequel cette probabilité est à son maximum. Ce palier est fonction croissante de la cote de risque. La probabilité de défaut atteint un niveau

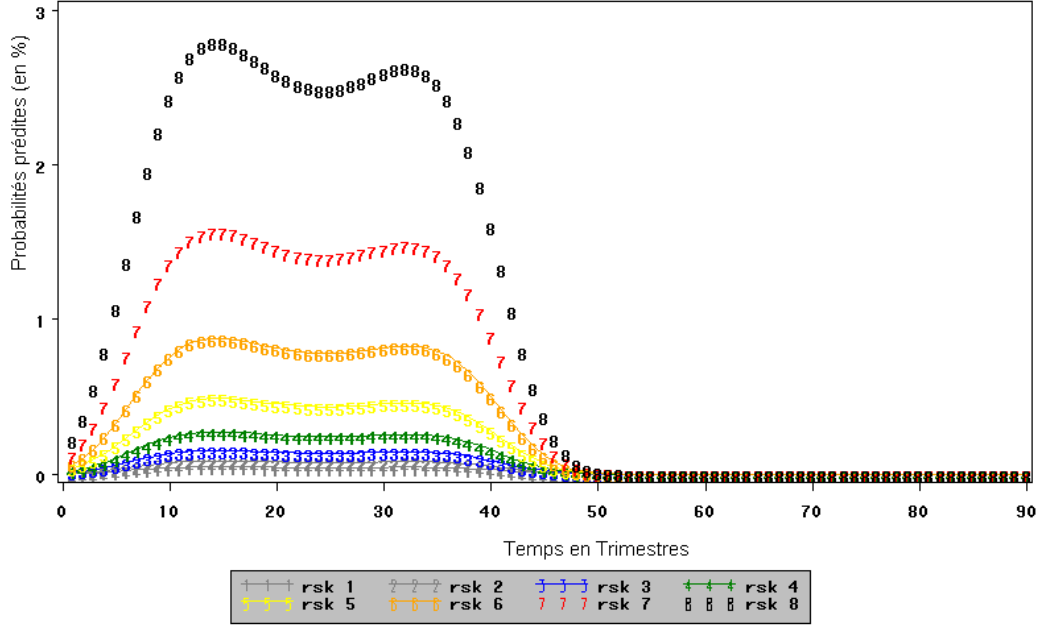


FIG. 5.1 – Influence de l’âge des prêts sur les probabilités de défaut

important de plus de 2,50% sur le palier du risque 8. Ce n’est pas étonnant puisque nous avons vu, au chapitre précédent, que le taux de défaut atteignait plus de 11% des prêts de ce niveau de risque. Ce qui est aussi intéressant est qu’aux premiers mois, la probabilité de défaut est généralement petite, d’après ce même graphique. En supposant la bonne fois des emprunteurs, ces derniers assument leur décision de financement et arrivent généralement à payer les premières échéances.

Enfin, nous allons tenter ici de justifier le fait de garder les trois variables $dChom$, $dIPsem$ et $dEparn$ dans notre modélisation de la probabilité de défaut. En fait, la p-value de $dEparn$ est relativement grande par rapport à celles des autres variables. Le tableau 5.3 resume les paramètres estimés pour trois modèles. Le modèle 1 utilise uniquement rsk et le polynôme de l’âge des prêts. L’ajout de la variable $dChom$ caractérise le modèle 2. Enfin, nous mettons nos trois variables $dChom$, $dIPsem$ et $dEparn$ dans le modèle 3. Le paramètre χ^2 de Hosmer et Lemeshow semble affirmer que la meilleure adéquation des prédictions par rapport aux réalisations est représen-

tée par le modèle 3. Par ailleurs, le critère d'information d'Akaike est à son minimum pour le même modèle 3. Nous nous aidons également par la figure 5.2. Celle-ci trace les prédictions des trois modèles en parallèle avec les fréquences historiques des défauts observées durant la période 2003 à 2006. Le modèle 1 qui ne comporte que la variable $risk$ et l'âge des prêts ne semble suivre que grossièrement l'évolution historique des défauts. Ceci prouve graphiquement son insuffisance pour simuler les défauts. Par contre, le modèle 2, qui intègre la variable $dChom$ simule déjà très bien les défauts historiques. Cependant, c'est le modèle 3, avec les trois variables macroéconomiques simultanément, qui semble mieux correspondre à ces fréquences. Ceci est confirmé par l'analyse des coefficients U de Theil évaluant les prédictions des trois modèles comme séries chronologiques. Pour finir, examinons les résultats des tests de Wald affichés à la fin de ce même tableau 5.3. L'hypothèse nulle est que les coefficients des variables macroéconomiques sont simultanément tous nuls pour le modèle 3. Nous voyons alors qu'à 10% de degré de confiance nous pouvons rejeter cette hypothèse. Ce qui justifie l'importance de l'ajout des variables macroéconomiques et nous conduit à les garder toutes les trois dans notre modèle final.

5.2 La distribution LGD à la Gouriéroux et al.

Voici maintenant les résultats de la dernière distribution qui nous intéresse (Tableau 5.5). Nous utilisons le modèle décrit par l'équation 4.4. Dans un premier temps, signalons que $risk$ est ici très significative avec un signe positif comme attendu. Cependant, la variable Gar (présence de garantie) n'est pas significative. Ceci est surprenant. Malgré tout, on devait avoir un résultat qui confirme plutôt l'intuition. La variable age n'est pas significative non plus. L'âge des prêts n'aurait aucune incidence sur les taux de pertes en cas de défaut.

TAB. 5.3 – Comparaison des modèles Logit multi-périodiques

Variable	Signe	Modèle1		Modèle2		Modèle3	
	attendu	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Intercept		-11.50430	<.0001	-11.41510	<.0001	-11.42370	<.0001
rsk	+	0.58780	<.0001	0.58710	<.0001	0.58880	<.0001
age	+	0.61450	0.0167	0.61260	0.0173	0.62800	0.0152
age²	?	-0.04300	0.0551	-0.04290	0.0571	-0.04430	0.0505
age³	?	0.00126	0.1001	0.00126	0.1024	0.00131	0.0920
age⁴	-	-0.00001	0.1352	-0.00001	0.1366	-0.00001	0.1249
dChom	+	-	-	1.12630	0.0851	1.21600	0.0560
dIPsem	+	-	-	-	-	0.29430	0.1342
dEparn	-	-	-	-	-	-0.25040	0.1776
Paramètres Hosmer & Lemeshow							
χ^2		5.6047	0.6914	9.6977	0.2869	5.0716	0.7499
AIC		967.16		966.382		966.314	
SBC		1017.24		1024.808		1041.434	
-2 Log L		955.16		952.382		948.314	
Test χ^2 de Wald				χ^2 de Wald		degrés de liberté	
(relatif au modèle 3)							
Hypothèse nulle :							
les coefficients des variables							
dChom et dIPsem							
sont tous nuls							
				5.4216		2 0.0665	
dChom, dIPsem et dEparn							
sont tous nuls							
				6.7847		3 0.0791	

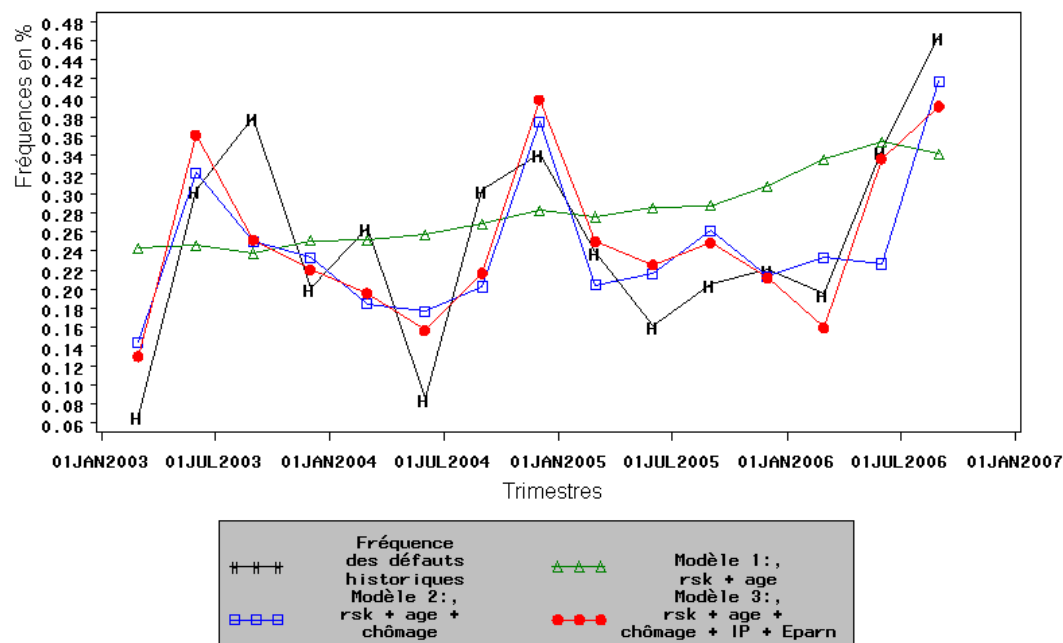


FIG. 5.2 – Comparaison des prédictions des probabilités de défaut

TAB. 5.4 – Fréquence historique et prédictions des défauts des modèles 1,2 et 3

Trimestre	Fréquences des défauts	Prédiction des défauts			Perte historique	# défauts	dChom	dIPsem	dEparn
		Modèle 1	Modèle 2	Modèle 3					
Mar-03	0.068	0.243	0.207	0.130	0.00	1	0.367	-0.600	-0.475
Jun-03	0.303	0.246	0.383	0.361	573.76	5	0.067	-0.100	-0.024
Sep-03	0.380	0.238	0.258	0.252	9291.97	7	-0.333	1.133	0.236
Dec-03	0.201	0.251	0.168	0.220	0.00	4	-0.133	-0.500	-0.236
Mar-04	0.264	0.251	0.214	0.196	0.00	6	-0.100	-0.867	0.477
Jun-04	0.086	0.257	0.228	0.157	0.00	2	-0.167	-0.267	-0.270
Sep-04	0.303	0.268	0.219	0.217	41448.05	7	0.067	0.700	-0.245
Dec-04	0.342	0.282	0.306	0.398	1606.02	8	-0.200	-0.200	-0.811
Mar-05	0.239	0.276	0.217	0.250	0.00	6	-0.067	-0.667	-0.166
Jun-05	0.161	0.285	0.263	0.226	0.00	4	-0.167	0.367	0.207
Sep-05	0.205	0.288	0.235	0.249	0.00	5	-0.200	-0.433	0.022
Dec-05	0.221	0.308	0.242	0.212	13532.97	5	-0.100	-0.833	1.520
Mar-06	0.196	0.336	0.297	0.160	25802.70	4	-0.200	-0.700	-1.590
Jun-06	0.345	0.354	0.278	0.336	33720.39	6	0.200	-0.133	0.263
Sep-06	0.464	0.341	0.434	0.391	37197.77	7	0.200	-0.133	0.263

TAB. 5.5 – Estimation conjointe Tobit LGD/PD

Paramètre	Estimation	écart-type	p value
LGD			
<i>rsk</i>	0.51104	0.1782	0.0042
<i>dIPsem</i>	1.71088	0.8559	0.0456
<i>_sigma LGD</i>	4.22950	0.3288	<.0001
PD			
intercept	-4.60308	0.3327	<.0001
<i>rsk</i>	0.19602	0.0211	<.0001
<i>age</i>	0.21038	0.0829	0.018
<i>age</i> ²	-0.014882	0.0073	0.0597
<i>age</i> ³	0.000438	0.0002	0.1035
<i>age</i> ⁴	-0.000004561	<.0001	0.1344
<i>dChom</i>	0.398681	0.2265	0.0784
<i>dIPsem</i>	0.092229	0.0686	0.1787
<i>dEparn</i>	-0.076061	0.0626	0.2243
<i>_Rho</i>	0.32827	0.0859	0.0001

Un autre résultat surprenant est que seule la variable *dIPsem* est significative et avec le bon signe (postif). Ainsi, les variables qui affectent la probabilité de défaut s'avèrent être différentes de celles qui affectent le taux de perte en cas de défaut.

Un mot sur le coefficient de corrélation obtenu. *Rho* est très significatif, ce qui tente de confirmer l'existence d'une relation très étroite entre les distributions *PD* et *LGD* dans notre cas. Nous rajoutons enfin qu'en parallèle des coefficients de *LGD*, nous avons également ceux relatifs à la distribution *PD*. Ils diffèrent bien entendu des coefficients obtenus lors de la première estimation Logit, puisque les calculs du Tobit généralisé supposent plutôt une distribution normale pour les résidus relatifs à *PD*. Cela dit, nous savons que les deux estimations peuvent être considérées équivalentes. C'est ce que nous allons utiliser pour nos simulations.

5.3 Simulations Monte Carlo

Nous avons maintenant les ingrédients nécessaires pour mettre en oeuvre la technique de simulation Monte Carlo. Pour commencer, nous avons besoin de faire une hypothèse importante. Nous supposons que l'échantillon de la base de données des prêts de l'institution est représentatif en termes de composition du portefeuille réel, d'une part, et que la composition du portefeuille de l'institution ne change pas significativement dans le temps. Au lieu de modéliser les proportions des types de prêts et les corrélations entre prêts, nous pouvons travailler directement avec les prêts de l'échantillon.

Notre stratégie s'effectuera en deux étapes. La première étape a pour objectif de générer, pour chaque prêt, n valeurs possibles des pertes probables. Pour ce faire nous calculons n valeurs possibles de la probabilité de défaut PD en parallèle de n autres valeurs pour LGD . Nous devons générer et ajouter un bruit blanc binormal pour chacun des prêts. Nous utiliserons le coefficient de corrélation Rho et l'écart-type $sigma$ issus de l'estimation Tobit pour générer cette distribution binormale. EAD est fixé au solde du prêt au moment du calcul. Le résultat est un ensemble de n vecteurs du produit $PD * LGD * EAD$ pour chaque prêt. La deuxième étape consiste à calculer m pertes probables du portefeuille. Dans ce calcul, chaque prêt participera avec l'une des n pertes possibles prises au hasard. Ceci nous donnera une distribution des pertes possibles au niveau global du portefeuille. Concrètement, nous fixons n et m respectivement à 8000 et 18000 itérations pour lesquels l'écart-type des valeurs calculées semble devenir suffisamment petit.

En déterminant les VaR à différents degrés de confiance, nous obtenons les pertes maximales dans les conditions économiques précisées par les valeurs de nos variables macroéconomiques. Comme déjà dit, l'annexe B détaille la méthodologie pour calculer les prédictions des variables aussi bien dans le cas des prévisions normales hors

échantillon du modèle VAR que dans le cas des réponses aux chocs exogènes au système VAR(2). Il est important de souligner que certaines de nos variables macroéconomiques contiennent de l'effet GARCH, ce qui prouve que les corrélations entre les quatre variables peuvent changer dans le temps. Toutefois, nous ignorerons ce fait dans ce travail, par souci de simplification toujours. Cela pourrait constituer une amélioration importante à apporter à notre méthodologie puisque les corrélations entre les facteurs changent pendant les crises.

À titre de vérification ex-post, prenons le troisième trimestre de 2004 où nous enregistrons la perte la plus forte avec 41448\$ (Tableau 5.4). La simulation dans les conditions économiques définies par nos trois variables macroéconomiques de ce trimestre donnent successivement, 135900\$ à 99.9% et 77950\$ à 99% comme capital requis pour cette période. Ce qui assure la couverture de la perte réelle. Nous poursuivons nos tests avec le dernier trimestre de 2005. La perte s'y élève à 13532\$. La simulation fournit un montant de 11700\$ de perte à 99.9% et 63900\$ à 99% comme degré de confiance de la VaR. Un dernier cas concernant le troisième trimestre de 2006. Là encore, nous obtenons 164650\$ à 99.9% et 93400\$ à 99%, alors que la perte réelle enregistrée est de 33720\$.

Nous pouvons maintenant envisager les simulations de crises de stress testing. Les figures A.1, A.2 et A.3 tracent la distribution des pertes trimestrielles calculées pour trois situations. La première est relative au cas des conditions ordinaires de l'économie, basées sur les prévisions standard hors échantillon des facteurs pour le premier trimestre 2007. La deuxième simule une baisse de 5 écarts-types en variation relative du PIB réel, ce qui correspond à une baisse de 5.45% du PIB trimestriel ($-5 \times 1.09 = -5.45\%$, Tableau). Nous considérerons cette simulation comme récession, bien que cela ne concerne qu'un seul trimestre plutôt que deux trimestres consécutifs comme le veut la définition macroéconomique de ce terme. Quant à la troisième figure, elle visualise

TAB. 5.6 – Pertes trimestrielles prévisionnelles du portefeuille

	Moyenne(\$)	Écart-type	VaR 99.9%	VaR 99%	VaR 97.5%	VaR 95%
Conditions normales	-34154	10877	-128529	-65196	-57436	-52073
			1.17%	0.59%	0.52%	0.47%
Récession	-50414	15414	-180526	-94181	-83418	-75745
(-5 écarts-types)			1.64%	0.86%	0.76%	0.69%
Crise de chômage :	-59488	17248	-200967	-108732	-96506	-87859
(+3 écarts-types)			1.83%	0.99%	0.88%	0.80%

l'impact d'un hypothétique choc exogène de 3 écarts-types en tant que hausse du taux de chômage, de 0.69% dans notre cas. Toutes les trois distributions présentent l'asymétrie caractéristique des distributions des pertes de crédit. Nous voyons bien que la distribution des pertes s'aplatit et s'élargit lors des conditions économiques sévères de la récession, et surtout pour la crise de chômage. Pour chacun de ces graphiques, la superposition de la distribution normale la plus proche permet de juger visuellement l'importance des caractères asymétriques et leptocurtiques des distributions des pertes obtenues. En pratique, nous déterminons les pertes maximales relativement aux VaR 99.9%, 99%, 97.5% et 95%, en non paramétrique à l'aide du logiciel SAS. Le tableau 5.6 présente un récapitulatif des chiffres.

Relativement à 11 millions de dollars comme exposition totale du portefeuille, nous obtenons pour une VaR de 99.9% un capital économique requis de 1.17% contre 1.64% en cas de récession et 1.83% pour la crise de chômage. Rappelons que ces chiffres sont relatifs à une période d'un trimestre. Afin d'avoir un ordre de grandeur des pertes dans le cas d'une année, nous pourrions multiplier ces chiffres par 4. Dans ce cas, toujours pour la même VaR de 99,9%, nous obtenons environ 4.67% (514000\$) en capital requis à garder par l'institution, 6.56% (722000\$) de perte en cas de la récession, et 7.31% (804000\$) comme perte suite à la crise de chômage simulée dans ce

travail. Notons déjà que les montants de pertes dégagés par les simulations ne sont pas excessivement grands par rapport aux pertes historiques, même pour une VaR à 99.9%. En fait, nous pouvions nous attendre à des chiffres conséquents puisque nous parlons de crises et de désastres économiques et financiers. Ceci est un constat souvent souligné par les recherches sur le stress testing. La réponse résiderait dans le fait que c'est dû à un bon niveau de sélection et de diversification des risques que les institutions financières assurent dans leurs choix de portefeuilles, par leurs propres outils internes. L'utilité de la procédure du stress testing est, justement, de pouvoir détecter les vulnérabilités d'une façon anticipée afin que les gestionnaires prennent les décisions adéquates en couverture, diversification, ou changement de position dans leurs portefeuilles.

Chapitre 6

Conclusion

Au bout de ce tour dans le monde des concepts des distributions, nous nous rendons compte combien la modélisation de la gestion des risques du crédit est aujourd'hui loin des premières hypothèses sur la normalité des distributions des pertes. Nous touchant également à la complexité de la problématique du risque de crédit qui nécessite de plus en plus de travaux et d'outils afin d'améliorer la précision des anticipations.

Du côté des résultats de cette étude, le premier est la validité du modèle de Wilson pour notre portefeuille, sous trois conditions. D'abord, il est nécessaire de tenir compte des spécificités des prêts et des individus. La deuxième condition est de modéliser les effets économiques en termes de variations de facteurs macroéconomiques plutôt qu'en niveau. La troisième condition est d'injecter une fonction de durée dans le modèle et de lire les résultats en tant que probabilités conditionnelles à la manière de Cox(1975). D'autre part, avec les simulations réalisées sur l'échantillon de notre institution, et selon le degré de représentativité de cet échantillon, nous savons que la perte maximale en cas de crise importante ne dépasse pas 8% de l'investissement. Par rapport à la stratégie interne de l'institution, ce chiffre devrait refléter aux gestionnaires une mesure sur la qualité de crédit du portefeuille global, en termes de degré

de sélection des prêts, de diversification et de couverture à prévoir.

La contribution de ce travail réside dans l'interconnexion des différents concepts pour aboutir au modèle final destiné à un portefeuille de prêts à la consommation pour les particuliers d'une institution financière. En effet, classiquement, les études prennent l'ensemble de la problématique des défauts des particuliers mais à l'échelle nationale. De plus, la quasi totalité des recherches s'adresse aux prêts hypothécaires uniquement. Pour nous, dans cette étude, l'intuition conductrice est de supposer que notre institution est suffisamment grande et importante à l'échelle nationale pour que sa clientèle soit impactée par les variations des conditions économiques. Nous avons, ensuite travaillé en VAR(2) plutôt qu'en plusieurs AR(2) pour modéliser les facteurs macroéconomiques. Enfin, nous avons fait appel aux fonctions de survie de Cox, pour résoudre notre modèle, et en utiliser les résultats. Concernant la distribution *LGD*, l'apport de Gouriéroux et al (2006) est de taille pour redonner cette distribution une place dans les efforts déployés. Comme nous avons expérimenté dans ce travail, l'intuition est solide, malgré quelques questions pratiques à résoudre pour une mise en oeuvre avantageuse. L'enjeu est important puisque cela remplit une exigence de Bâle II pour cette distribution. Toute défaillance dans la modélisation de *LGD* se retrouve forcément dans la distribution finale des pertes, exactement au même titre que pour *PD*.

Dans le chapitre des améliorations, comme nous l'avons déjà soulevé, il y aurait une piste pour explorer les possibilités des modèles GARCH(p,q) en parallèle de VAR(p). Par ailleurs, et avec un échantillon plus riche, une meilleure appréhension des distributions *LGD* et *EAD* devrait affranchir l'étude des hypothèses qui sont pour le moment nécessaires pour modéliser ces deux distributions importantes d'une façon plus adéquate. Enfin une piste qui semble de plus en plus explorée dans les recherches. Les fonctions de survie permettent de faire des prédictions non seulement sur la vraisem-

blance qu'un prêt tombe en défaut, mais aussi quand cela est plus à même d'arriver. Ceci pourrait ouvrir des perspectives nouvelles dans la manière d'octroyer les prêts pour les institutions, disposant d'une connaissance meilleure quant à l'évolution du comportement des emprunteurs dans le temps.

Bibliographie

- ALLEN, L., ET A. SAUNDERS (2004) : “Incorporating Systemic Influences into Risk Measurements : A Survey of the Literature,” *Journal of Financial Services Research*, Vol 26, numéro 2 / Octobre, 2004, 161–191.
- ALTMAN, E., ET V. KISHORE (1996) : “Almost Everything You Wanted to Know about Recoveries on Defaulted Bonds,” *Financial Analysts Journal*, pp. Novembre/Décembre pp. 57–64.
- (1998) : “Defaults and Returns on High Yield Bonds : Analysis through 1997,” Document de travail S-98-1. Salomon Center, New York University.
- ALTMAN, E. I. (1968) : “Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy,” *Journal of Finance*, 23 :, 589–609.
- ANDRADE, F.-M., ET L. THOMAS (2004) : “Structural models in consumer credit,” Document de travail, AF 4-19, School of Management, University of Southampton.
- ASARNOW, E., ET D. EDWARDS (1995) : “Measuring Loss on Defaulted Bank Loans : a 24-Year Study,” *The Journal of Commercial Lending*, pp. 11–23.
- BAKER, M., ET J. WURGLER (2000) : “The Equity Share in New Issues and Aggregate Stock Returns,” *Journal of Finance*, vol 55, issue 5, pp 2219–58.
- BRANNAN, S., C. SMITHSON, ET D. M. M. ZMIEWSKI (2002) : “2002 Survey of Credit Portfolio Management Practices,” Rutter Associates, International Association of Credit Portfolio Managers (IACPM), International Swaps and Derivatives Association (ISDA), Risk Management Association (RMA).

- BUCAV, N., ET D. ROSEN (2001) : “Applying Portfolio Credit Risk Models to Retail Portfolios,” *The Journal of Risk Finance*, Vol. 2, No. 3, Spring 2001, pp 230–241.
- CAREY, M. (2001) : “Dimensions of Credit Risk and Their Relationship to Economic Capital Requirements,” in Frederic S. Mishkin, ed., *Prudential Supervision : Why Is It Important and What Are the Issues*. Chicago : University of Chicago Press and NBER.
- CHAVA, S., ET A. JARROW (2004) : “Bankruptcy Prediction with Industry Effects,” *Review of Finance*, 8, 537–569.
- COLEMAN, A., N. ESHO, I. SELLATHURAI, ET N. THAVABALAN (2005) : “Stress Testing Housing Loan Portfolios : a Regulatory Case Study,” Document de travail, Australian Prudential Regulation Authority, Basel Committee on Banking Supervision Conference 2005-01.
- COLLIN-DUFRESNE, P., ET R. GOLDSTEIN (2001) : “Do Credit Spreads Reflect Stationary Leverage Ratios?,” *Journal of Finance*, vol LVI, (5), pp 1929–57.
- COX, D. R. (1972) : “Regression Models and Life Tables (with discussion),” *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 34, No 2, 187–220.
- (1975) : “Partial Likelihood,” *Biometrika*, Vol. 62, 267–276.
- CREDIT SUISSE (1997) : “A Credit Risk Management Framework,” Credit Suisse Financial Products.
- CROUHY, M., D. GALAI, ET R. MARK (2000) : “A Comparative Analysis of Current Credit Risk Models,” *Journal of Banking and Finance*, vol 24, pp 59–117.
- DIONNE, G., M. ARTIS, ET M. GUILLÉN (1996) : “Count data models for credit scoring system,” *Journal of Empirical Finance* 3, pp. 303–325.
- DIONNE, G., G. GAUTHIER, K. HAMMAMI, M. MAURICE, ET J.-G. SIMONATO (2007) : “Default Risk, Default Risk Premium and Corporate Yield Spreads,” Document de travail, HEC Montreal.

- DREHMAN, M., A. J. PATTON, ET S. SRENSEN (2005) : “Corporate Defaults and Large Macroeconomic Shocks,” Document de travail, Bank of England, Monetary Policy Committee.
- DU, Y. (2003) : “Predicting Credit Rating and Credit Rating Changes : A New Approach,” Document de travail, Queen’s School of Business Queen’s University.
- DUFFIE, D., ET K. SINGLETON (1999) : “Modeling Term Structures of Defaultable Bonds,” *Review of Financial Studies*, 12(4), pp. 197–226.
- FRYE, J. (2000a) : “Collateral Damage,” *Risk*, pp. 91–4.
- (2000b) : “Depressing Recoveries,” *Risk*, pp. 108–11.
- GLENNON, D., ET P. NIGRO (2005) : “An Analysis of SBA Loan Defaults by Maturity Structure,” *Journal of Financial Services Research*, 28 :1/2/3, pp 77–111.
- GORDY, M. (1998) : “A Comparative Anatomy of Credit Risk Models,” Board of Governors of the Federal Reserve System.
- GOURIÉROUX, C., ET A. MONFORT (2006) : “(Non) Consistency of the Beta Kernel Estimator for Recovery Rate Distribution,” Document de travail, CREST, CNAM, CEPREMAP & University of Toronto.
- GOURIÉROUX, C., A. MONFORT, ET A. TIOMO (2006) : “Joint Analysis of Default Occurrence and Recovery Rate,” Document de travail, CREST, CNAM, CEPREMAP, University of Toronto, DEXIA, CREF, ERUDITE & University Paris XII.
- HU, Y.-T., ET W. PERRAUDIN (2002) : “The dependence of Recovery Rates and Defaults,” Document de travail, Birkbeck College.
- JORION, P. (2005) : *Financial Risk Manager Handbook*. Wiley Finance, 3rd edn.
- JP MORGAN (1997) : “CreditMetrics .Technical Document,” JP Morgan New York.
- LOWN, C. S., ET D. P. MORGAN (2001) : “The Credit Cycle and the Business Cycle : New Findings Using the Survey of Senior Loan Officers,” Document de travail, Federal Reserve Board of New York.

- MEI, J., ET A. SAUNDERS (1997) : “Have US Financial Institutions Real Estate Investments Exhibited Trend Chasing Behaviour?,” *The Review of Economics and Statistics*, pp. 248–58.
- MERTON, R. (1974) : “On the Pricing of Corporate Debt : the Risk Structure of Interest Rates,” *Journal of Finance*, vol 29, pp 449–70.
- MUSTO, D., ET N. SOULELES (2005) : “A Portfolio View of Consumer Credit,” Document de travail 05-25, Federal Reserve bank of Philadelphia.
- PARIS, F.-M. (2005) : “Selecting an Optimal Portfolio of Consumer Loans by Applying the State Preference Approach,” *European Journal of Operational Research* 163, pp. 230–241.
- PYE, G. (1974) : “Gauging the Default Process,” *Financial Analysts Journal*, pp. Jan–March, pp 49–52.
- RENAULT, O., ET O. SCAILLET (2003) : “On the Way to Recovery : Nonparametric Bias Free Estimation of Recovery Densities,” *Journal of Banking and Finance*, vol. 28(12), pp 2915–2931.
- RODRIGUEZ, A., ET C. TRUCHARTE (2006) : “Loss Coverage and Stress Testing Mortgage Portfolios : A Non-Parametric Approach,” Document de travail, Banco de España (Bank of Spain).
- SEGOVIANO-BASURTO, M.-A., ET P. PADILLA (2006) : “Portfolio Credit Risk and Macroeconomic Shocks : Applications to Stress Testing Under Data-Restricted Environments,” Document de travail, Internal Monetary Fund.
- SHUMWAY, T. (2001) : “Forecasting Bankruptcy more accurately : A Simple Hazard Model,” *Journal of Business*, 47, 1343–1366.
- SORGE, M. (2004) : “Stress-testing Financial Systems : An Overview of Current Methodologies,” Document de travail, No 165, Basel Monetary and Economic Department, BIS.

- VASICEK, O. (1991) : "Limiting Loan loss Probability Distribution," KMV Corporation.
- (2002) : "Loan Portfolio Value," *Risk*, pp. December, 160–162.
- WILSON, T. C. (1997a) : "Portfolio Credit Risk (I)," *Risk*, vol 10, issue 9, pp. 111-17.
- (1997b) : "Portfolio Credit Risk (II)," *Risk*, vol 10, issue 10, pp. 56-61.
- ZHOU, C. (1997) : "A Jump-Diffusion Approach to Modeling Credit Risk and Valuing Defaultable Securities," Document de travail 1997-15. Federal Reserve Board.

Annexe A

Graphiques

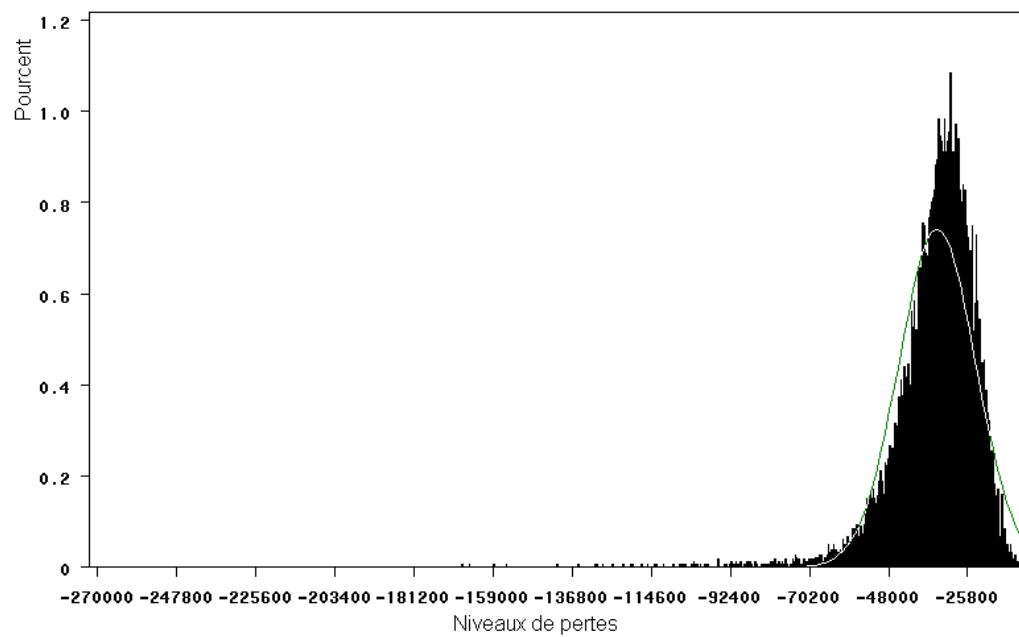


FIG. A.1 – Distribution des pertes prévisionnelles : Conditions économiques ordinaires

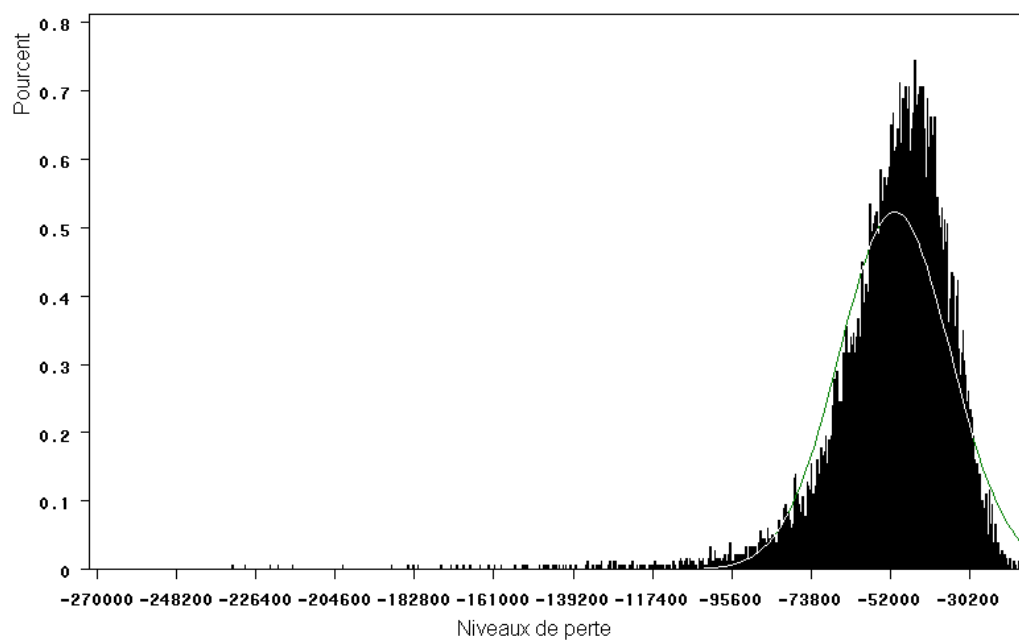


FIG. A.2 – Distribution des pertes prévisionnelles, Récession économique (Choc négatif sur le PIB réel de 5 écarts-types : -5.45%)

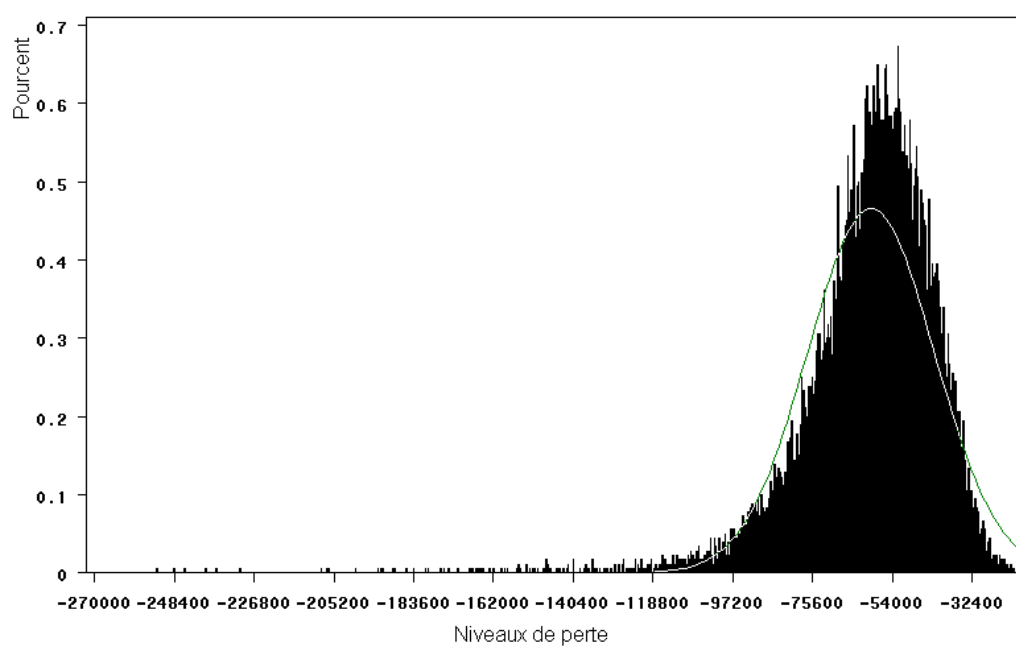


FIG. A.3 – Distribution des pertes prévisionnelles, Hausse du taux de chômage (Choc de 3 écarts-types : +0.69%)

Annexe B

Étude du système VAR(p)

Cette partie de l'étude est réalisée avec le logiciel RATS spécialisé dans les techniques économétriques des séries chronologiques. Nous étudions ici l'interaction de quatre séries : le PIB réel (PIBb), le taux de chômage (Chom), l'indice des prix de biens semi-durables (IPsem), et l'épargne (Eparn). Ces variables sont prises sur une base trimestrielle de mars 1990 à septembre 2006. Comme le montre le graphique de leur évolution, ces variables sont clairement non stationnaires, ce qui devrait nous amener à travailler en première différence des variables que nous notons dPIB, dChom, dIPsem et dEparn. Dans l'objectif d'utiliser et d'interpréter directement les variables, nous les modélisons en première différence simple (sans logarithme). Enfin, nous divisons par 10000 les valeurs de dPIB et dEparn afin d'assurer le même ordre de grandeur pour les quatre variables. La stationnarité de nos 4 variables ainsi obtenues semble graphiquement acceptable, ce que nous vérifions techniquement à l'aide de tests de Dickey-Fuller. L'objectif ici est une modélisation en VAR(p) afin de calculer les prédictions en hors échantillon, d'une part, et de simuler l'impact des chocs sur certaines variables sur l'ensemble, d'autre part.

Nous commençons par analyser les distributions de ces variables. Les moyennes de dChom, IPsem et Eparn sont statistiquement nulles à 5% (p-value= 0.5642, 0.1116

et 0.2759). Ce qui simplifie l'interprétation de l'analyse des chocs que nous verrons plus loin. La modélisation univariée de dPIB montre qu'un retard $p=5$ est optimal et permet d'avoir des résidus non autocorrélés sur 12 périodes à 5%. Ceci se fait avec la méthode du retard significatif maximal en contrôlant les critères AIC et BSC d'Akaike et Schwarz. De la même façon, le retard optimal est de 1 pour dChom, de 4 pour dIPsem, et de 1 pour la série dEparn en univarié. Par ailleurs, à l'aide des commandes RATS, le test Dickey-Fuller confirme la stationnarité des quatre variables à 5%. En effet, les t -statistiques sont respectivement de -4.2105, -4.0352, -6.0447 et -7.8289, ce qui est convenablement inférieur au seuil de Dickey-Fuller de la racine unitaire de -3.478 à 5%.

Pour passer au modèle multivarié, rappelons-nous que nous n'avons que 68 trimestres dans nos observations et que la modélisation en multivarié consomme beaucoup de degrés de liberté. En effet, un VAR(p) pour n variables nécessite l'estimation de $(p * n + 1) * n$ coefficients outre $n * (n + 1)/2$ éléments de la matrice symétrique des variances covariances des résidus. Dans notre cas, avec 4 variables, si nous voulions un VAR(3) déjà il nous faudrait 52 coefficients et 10 variances-covariances, donc 62 degrés de liberté. Pour VAR(2), le calcul donne un total de $36+10=46$ ddl, ce que nous tenterons d'accepter dans notre cas, par rapport aux 68 observations dont nous disposons. Nous examinons maintenant si un VAR(2) peut nous satisfaire, quitte à tester les résidus qui doivent être considérés bruits blancs et de vérifier la performance des prédictions hors échantillon de ce système.

Les résidus des quatre séries doivent être testés comme bruits blancs à 5%. Nous utilisons le test de Ljung-Box qui calcule la significativité des autocorrélations sur une plage de retards choisis. En fixant une fenêtre de 6 retards (donc 6 trimestres), les résidus de dPIB réussissent ce test à 5% de justesse (p -value = 0.07498 > 5%). Les résidus de dChom, dIPsem et dEparn peuvent être considérés également comme

bruit blanc 5% d'une façon plus confortable (p-value = 0.43110, 0.18691 et 0.5884). Au passage, nous en profitons pour tester s'il existe de l'effet GARCH dans ces résidus. Pour ce faire, nous faisons appel au test de McLeod qui est le même que Ljung-Box mais appliqué au carré des résidus. Seule la série dChom présente un effet GARCH dans ses résidus à 5% et pour une fenêtre d'autocorrélation de 12 périodes. Ceci nous donne une piste d'amélioration intéressante pour une modélisation plus adéquate de la volatilité du système de nos quatre variables. Cependant, nous maintenons dans notre travail une modélisation VAR(2) pour privilégier des concepts simples. Avançons maintenant dans le sens des prédictions en hors-échantillon. Nous reculons de 10 trimestres et calculerons les prédictions du VAR(2) que nous comparerons aux réalisations réelles des séries (environ 80% des observations). Les résultats sont comparés aux réalisations en utilisant le coefficient U de Theil. Celui-ci revient à comparer la distance entre les prédictions et les réalisations par rapport à une modélisation " naïve " par la moyenne de la série (constante). Un coefficient supérieur à 1 montrerait que nos prédictions ne sont pas très performantes, étant moins bonnes que la moyenne de la série, et vice-versa. Dans notre cas, nous obtenons respectivement 0.3199, 0.7339, 0.5764 et 0.9531 pour nos quatre séries. Ce qui est très bon pour dPIB, reste très intéressant pour dChom et dIPsem, et est médiocre pour dEparn mais que nous gardons.

Nous calculons maintenant trois séries. La première détermine les prévisions futures de nos 4 variables en différences. Cela concerne les quatre trimestres de 2007, la date des dernières observations étant fin décembre 2006 (Table B.1). Les deux autres séries à calculer concernent les chocs exogènes pour les simulations de crise. Nous simulerons deux crises. D'abord un choc exogène sur le PIB afin de mesurer l'impact d'une récession. Ensuite, nous verrons ce qui se passe lors d'un choc exogène sur le chômage et mesurerons son impact sur notre système. Rappelons-nous qu'une modélisation VAR est en fait une sorte de forme réduite d'un système structurel que nous ne modélisons pas, a priori. Cependant, nous savons que pour passer des chocs structurels à ceux

en forme réduite, nous pouvons utiliser la technique de la décomposition de Cholesky de la matrice Ω des résidus de la forme réduite. Si nous notons par u_t et v_t les résidus respectifs du système structurel et ceux de la forme réduite, nous pouvons écrire $v_t = L.u_t$ avec L : matrice triangulaire inférieure de Cholesky des résidus de notre VAR(2). Cette décomposition étant sensible à l'ordre des variables dans la matrice Ω , il est fondamental que nous nous prononcions sur une structure de causalité entre les quatre séries. Nous nous basons sur les tests de causalité au sens de Granger et suggérons l'ordonnancement $d\text{PIB} > d\text{Chom} > d\text{IPsem} > d\text{Eparn}$ pour faire les calculs. En gros, nous supposons qu'un choc exogène sur le PIB déclenche une cascade de réactions successives sur le chômage, l'inflation et l'épargne de l'économie, et ce dès la première période. Par contre, un choc exogène sur le chômage n'affectera qu'à la deuxième période le PIB, mais devra impacter à son tour, l'inflation et l'épargne à partir de la première période. Nous allons donc simuler deux chocs macroéconomiques exogènes à notre VAR(2). Le premier est subit en premier par le PIB, le deuxième impacte le chômage à $t=0$.

Les graphiques de la figure B2, traçant l'impact sur une longue période montrent bien que les séries sont stationnaires avec des oscillations qui finissent par s'estomper à terme. Nous représentons également ces prévisions sur 12 trimestres pour visualiser l'impact des chocs jusqu'à 3 ans (figure B3). Nous devons faire attention ici car le choc simulé sur le PIB est positif par défaut. Pour une simulation de récession, nous devons inverser les signes de tous les chiffres avant de les interpréter. Dans notre cas, suite à un choc négatif sur le PIB (un écart-type =1.44%), nous obtenons une réaction positive, donc augmentation du chômage d'une amplitude de 0.062%. Cet impact persiste à +0.027% et +0.022% à la deuxième et troisième période. L'impact immédiat de cette récession serait une augmentation de l'inflation de 0.14%, suivi d'une autre augmentation de 0.08% à la deuxième période. Ce résultat est à première vue contre-intuitif. En principe, c'est en période de croissance que l'inflation augmente.

À ce niveau, nous aurions conclu que nous avons un problème à résoudre, sauf que si nous examinons attentivement nos séries elles-mêmes, nous y verrons souvent cette opposition des signes. L'interprétation est qu'il y a un décalage d'une période pour que les variables réagissent. Nous avons exactement le même phénomène d'oscillation en opposition de phase entre les variations relatives du PIB et avec celles de l'épargne. N'oublions pas que nous sommes en régime transitoire des variables. Dans ce cas, à l'équilibre et à terme, les variables retrouvent leurs caractéristiques macroéconomiques habituelles. Nous passons au deuxième choc qui s'effectue directement sur le chômage. L'impact sur le PIB est nul à la première période, par construction du modèle, et est négatif sur le PIB à la deuxième. Nous retrouvons ici encore un comportement en opposition de phase de dIPsem. Quant à dEparn, après une réaction immédiate positive, l'épargne ne tarde pas à diminuer en deuxième période pour revenir à peu près au niveau de sa moyenne nulle dès la troisième période.

Maintenant que nous avons nos deux chocs, nous devons leur donner une amplitude pour qu'ils soient " rares et plausibles ". Dans ce sens, la littérature affecte une amplitude variant de 3 à 5 écarts-types de la variable subissant le choc exogène la première. Nous utiliserons alors 5 écarts-types, en négatif, pour les prédictions calculées de la récession, ce qui correspond à une amplitude de -5,47% comme variation relative sur le PIB (table B.2). D'autre part, nous retenons 3 écarts-types pour la crise de chômage, ce qui est l'équivalent de +0.64% comme variation du taux de chômage (table B.3). Ajoutons ici qu'à titre d'une dernière simplification des calculs, l'impact des deux chocs a été généré en supposant que toutes les variables macroéconomiques sont nulles au moment de l'application des chocs exogènes (période $t=0$).

TAB. B.1 – Prévisions des valeurs macroéconomiques

Trimestre	PIB	Chômage	Indice	
			Prix semidurables	Épargne
1-Mar-07	-0.0032	-0.0435	0.1395	-0.2381
1-Jun-07	0.0271	-0.0875	0.0612	0.0158
1-Sep-07	0.0276	-0.0501	-0.0336	-0.1389
1-Dec-07	-0.0014	-0.0581	0.1989	-0.0334

TAB. B.2 – Prévisions suite à une récession (un écart-type négatif)

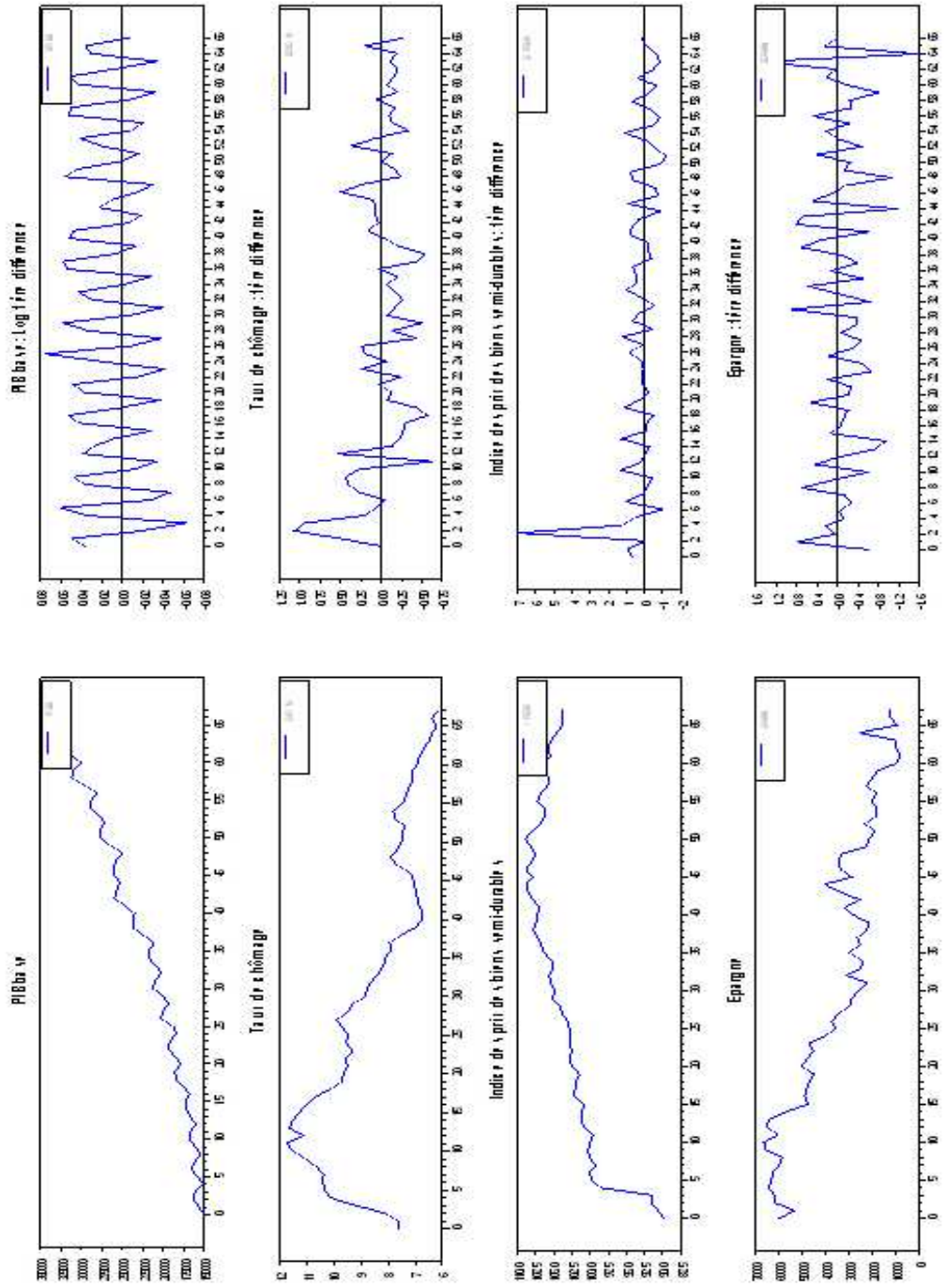
Trimestre	PIB	Chômage	Indice	
			Prix semidurables	Épargne
Mar-07	-0.0109	0.0344	-0.0283	0.0125
Jun-07	-0.0015	0.0382	-0.0502	-0.0755
Sep-07	0.0047	0.0161	0.0876	-0.0006
Dec-07	0.0008	0.0180	0.0376	-0.0406

TAB. B.3 – Prévisions suite à un choc sur le chômage (un écart-type positif)

Trimestre	PIB	Chômage	Indice	
			Prix semidurables	Épargne
Mar-07	0.0000	0.2130	-0.0704	0.1439
Jun-07	-0.0042	0.0295	-0.0557	-0.0353
Sep-07	0.0003	0.0583	0.0206	-0.1148
Dec-07	-0.0011	0.0156	0.1590	0.0909

A.

Figure B1: Évolution des séries



A

Figure B2: Effet des chocs sur les variables

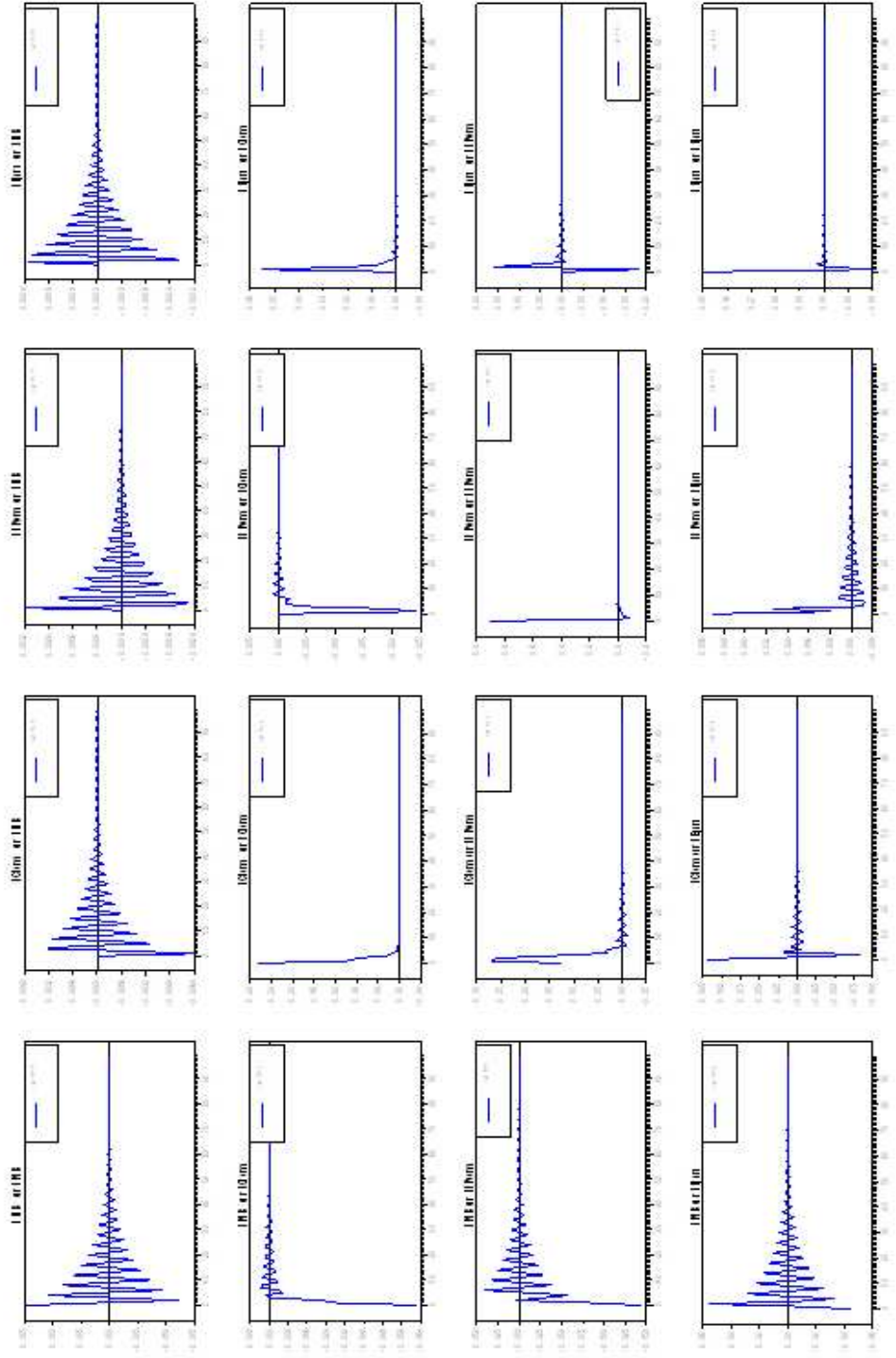


Figure B3: Effet des chocs sur les variables 12 trimestres

