ISUP

**PROMOTION 2011**

Mémoire présenté devant

**l'Institut de Statistique**
**de l'Université Pierre et Marie Curie**

Pour l'obtention du

# Diplôme de Statisticien

## Mention Actuariat

Assurance ❑ Finance ❑

Par Mlle. **Xin XU**

Sujet : **Modélisation De La Probabilité De Défaut Dans Le Cadre de Stress Scénarii Bâle II**

Lieu du stage : **CRÉDIT AGRICOLE S.A.**
**12, Place des Etats-Unis – 92120 MONTROUGE, FRANCE**

Responsable du stage : **Katy FRERE, Julien BOUDOU**

Invité(s): CONFIDENTIEL ❑

# Abstract

Basel II framework is a directive which redefined the risk management of the banking institutions. It requires banks to conduct stress tests in order to determine their needs in stockholders' equity according to the requirements defined by the directive. Thus, several risks must be assessed which the credit risk.

The risk which is brought by a portfolio of loans is the risk of default of the borrowers. Therefore, it corresponds to the fact that a person or a company cannot pay off its debt when it should be due. It is a major problem for the banks that have to control this type of risk

The aim of this study is to build a macroeconomic model of credit risk for the corporate portfolio of Groupe Crédit Agricole. The model is based on Error Correction model and the combination between the Generalized Linear Model and the ARMA model. Thanks to these models, they enable us to simulate the rate of default for two macroeconomic scenarios. The first macroeconomic scenario correspond to the Baseline scenario which consists in planning the most likely or real macroeconomic perspectives, the second scenario is the Adverse scenario which consists in planning a more severe macroeconomic situation.

Finally, in this study, we put in relation the results obtained by these models and the real data in order to present their robustness.



**KEYWORDS:** *Credit Risk, Probability of default, Basel II, Stress scenarios, Intern Model, Error Correction Model, Generalized Linear Model*

# Résumé

Bâle II est un directive qui redéfinie la gestion des risques des institutions bancaires. Elle exige des banques qu'elles effectuent des stress tests adéquats afin de déterminer leurs besoins en fonds propres selon des exigences bien définies. Ainsi plusieurs risques doivent être appréhendé dont le risque de crédit.

Le risque que porte un portefeuille de prêt est le risque de défaillances des emprunteurs. Il correspond alors au fait qu'un particulier ou qu'une entreprise ne puissent rembourser sa dette à l'échéance. C'est un problème majeur pour les banques qui se doivent de maîtriser ce type de risque.

Cette étude présente un modèle macroéconomique du risque de crédit pour le portefeuille d'entreprises du Groupe Crédit Agricole. Il est fondé sur le modèle à correction d'erreurs et sur la combinaison entre le modèle linéaire généralisé et le modèle ARMA. A partir de ces outils, le taux de défaillance est simulé pour deux scénarii macroéconomiques. Le premier correspond au scénario Baseline, qui consiste à prévoir les conditions macroéconomiques futures réelles ou les plus probables, le second est le Adverse scénario, qui consiste à prévoir le taux de défaillance en cas de situation macroéconomique plus dégradée.

Enfin, dans cette étude, nous mettons en relation les résultats obtenus à l'aide de ces modèles et les données réelles afin de présenter leurs robustesses.

**MOTS-CLÉS :** *Risque de crédit, Probabilité de défaut, Bâle II, Stress scénarii, Modèle Interne, Modèle à Correction d'erreurs, Modèle Linéaire Généralisé*

# REMERCIMENTS

*Tout d'abord, je tiens à remercier Pascal MEYER, Responsable de la gestion des risques, ainsi que Katy FRERE et Julien BOUDOU, qui ont suivi mon évolution tout au long de mon stage, qui ont su me consacrer du temps, qui ont eu la pédagogie et la patience de m'aider à avancer dans mon travail.*

*Je souhaite également exprimer ma reconnaissance à Pascal WEINSTEIN et Pierre CHAPTAL, membres de l'équipe de Gestion des Risques, pour les conseils qu'ils m'ont apportés, ainsi que pour le temps qu'ils ont consacré à répondre à mes questions.*

*Ainsi, je remercie l'ensemble des équipes de la Direction de Gestion des Risques du Crédit Agricole S.A., pour leur collaboration ainsi que pour leur aide et leurs conseils.*

*Je voudrais remercier M. Jacques CHEVALIER, le professeur de cours de mathématiques financières et Olivier LOPEZ, le professeur de cours de modèle linéaire, pour leurs conseils prolifiques sur l'étude statistique et financière dans ce mémoire.*

*Enfin, je tiens à remercier les membres du jury qui ont eu l'amabilité d'accepter de juger ce travail.*

# Sommaire

# INTRODUCTION

L'avènement des normes Bâle II modifie profondément le rapport au risque des banques au travers de l'appréhension plus profonde des risques bancaires et principalement du risque de crédit ou de contrepartie et de calcul des exigences de capitaux propres. Cette démarche mondiale de réglementation de la profession bancaire a pour l'objectif de prévenir les défaillances bancaires par une meilleure adéquation entre fonds propres et risques encourus. Pour répondre à cet objectif, les accords de Bâle fixent les règles pour une meilleure évaluation des risques. En particulier, la réforme Bâle II impose la définition de stresse scénarii destine à vérifier que les fonds propres sont suffisants pour supporter une dégradation du risque à l'occasion d'un choc économique.

Dans ce contexte, la préconisation de mise en place d'un modèle interne devient la priorité des banques en matière de gestion des risques. Dans ce mémoire, nous nous intéressons de plus près aux études pour modéliser le taux de défaillance des clients, qui est un élément clé de l'évaluation du risque de crédit. La finalité de ce mémoire est de proposer un outil fiable aux gestionnaires des risques souhaitant estimer le taux de défaillance sur le périmètre corporate du portefeuille du Groupe Crédit Agricole.

Ce mémoire est structuré autour de trois parties.

La première partie est consacrée à la présentation et à l'analyse des principes sur lesquels repose le traitement des différents risques requis par les différents référentiels. Une attention particulière est portée sur le risque de crédit et le contexte de notre étude. Enfin, une illustration des définitions des stress scénarii pour terminer la partie 1.

La deuxième partie est consacrée aux aspects théoriques de deux modèles proposés. Le modèle à correction d'erreurs consiste à étudier une véritable relation de la variable à expliquer les variables explicatives à l'aide du test de cointégartion et le modèle à court terme. T que le modèle linéaire généralisé modélise l'évolution de la variable en étudiant la loi sur la distribution de cette dernière. Chacun des modèle présente leurs avantages et défauts et. L'analyse ce ces aspects fait aussi l'objet de la deuxième partie de notre mémoire.

La troisième partie est l'étude empirique des résultats des deux méthodes dont l'objectif est d'évaluer les prévisions de taux de défaillance par les deux modèles proposés ci-dessus dans un cadre de stress scenarii. Nous présenterons, tout d'abord, les données utilisées et les variables macroéconomiques proposées par les économistes du Groupe Crédit Agricole. Nous procéderons, ensuite, à la modélisation du taux de défaillance à l'aide du modèle à correction d'erreurs dans

les scénarii. Puis, nous allons nous intéresser à l'application de la combinaison du modèle linéaire généralisé et le modèle ARMA sur le taux de défaillance. Par les comparaisons entre les deux modèles potentiels, nous allons sélectionner le modèle le plus adéquat afin de l'appliquer sur les secteurs regroupes du portefeuille corporate.

# PARTIE I. CADRE DU PROJET BÂLE II

# INTRODUCTION

Le Comité de Bâle a été créé en 1974 sous l'appellation « Comité des règles et pratiques de contrôle des opérations bancaires », par les gouverneurs des banques centrales et des organismes de réglementation et de surveillance bancaires des principaux pays industrialisés. Il est hébergé par la BRI[1] basée à Bâle, d'où son nom. Sa mission est d'œuvrer en faveur d'un dispositif plus robuste en matière de supervision et de régulation du secteur bancaire. Les normes de Bâle II devraient remplacer les normes mises en place par Bâle I en 1988 et visent notamment à la mise en place du ratio McDonough[2] destiné à remplacer le ratio Cook[3] Bâle II est fondé sur une structure de trois piliers. Le premier, plus quantitatif, prévoit essentiellement les règles pour le calcule des exigences minimales de fonds propres selon diverses approches. Celles-ci peuvent être définies par des formules standards ou être développées en interne par chaque établissement de crédit. Le deuxième pilier est plus qualitatif. Il confère au jugement des superviseurs un rôle clé dans l'évaluation du profil de risque et de la qualité de la gestion par chaque établissement ainsi que, in fine, du seuil minimum de fonds propres. Enfin, le troisième pilier concerne la transparence et la communication d'informations relatives aux deux piliers précédents.

En s'intéressant, tout particulièrement, à l'évaluation du niveau des fonds propres dans le premier pilier, un ratio est introduit pour le calcul du niveau de ces derniers. Il est sensible aux risques réellement assumés par les banques. Dans ce cadre, tous les risques importants et quantifiables auxquels un établissement bancaire est exposé doivent être pris en compte. Il y a trois principaux types de risques : risque de marché, risque de crédit et de contrepartie et risque opérationnel.

Nous allons commencer par la présentation du risque de crédit. Ensuite, nous allons exposer la probabilité de défaut, qui est un paramètre important pour exprimer le risque de crédit. Enfin, nous allons nous concentrer sur les stress scénarios du taux de défaut, qui fait l'objet final de ces études.

[1] BRI : La Banque des règlements internationaux.

[2] Ratio McDonough : Ratio prudentiel destiné à mesurer la solvabilité des banques qui fixe une limite à l'encours pondéré des prêts (et autres actifs) accordés par un établissement financier en fonction de ses capitaux propres.

[3] Ratio Cook: Ratio prudentiel destiné à mesurer la solvabilité des banques qui définit le montant des Fonds Propres minimum que doit posséder une banque en fonction de sa prise de risque.

# Chapitre 1 Le Risque De Crédit Et De Contrepartie

L'accord Bâle cherche à rendre les fonds propres cohérents avec les risques réellement encourus par les établissements financiers. Parmi ceux-ci sont pris en compte les risques opérationnels et les risques de marché, et complétés du risque de crédit et de contrepartie.

La figure **Fig.1.1** présente l'ensemble des risques exposés par les établissements financiers.

Le comité de Bâle définit les risques majeurs de manière suivante :

- Risques opérationnel: *le risque de pertes provenant de processus internes inadéquats ou défaillants, de personne et systèmes ou d'événements externes.* Cette définition recouvre les erreurs humaines, les fraudes et malveillances, les défaillances des systèmes d'information, les problèmes liés du personnel…
- Risques de marché : *le risque de perte sur les positions du bilan[4] et du hors-bilan[5] à la suite de variation des prix de marché.* Cette définition recouvre les risques relatifs aux instruments liés aux taux d'intérêt et titre de propriété du portefeuille de négociation ainsi que le risque de change…

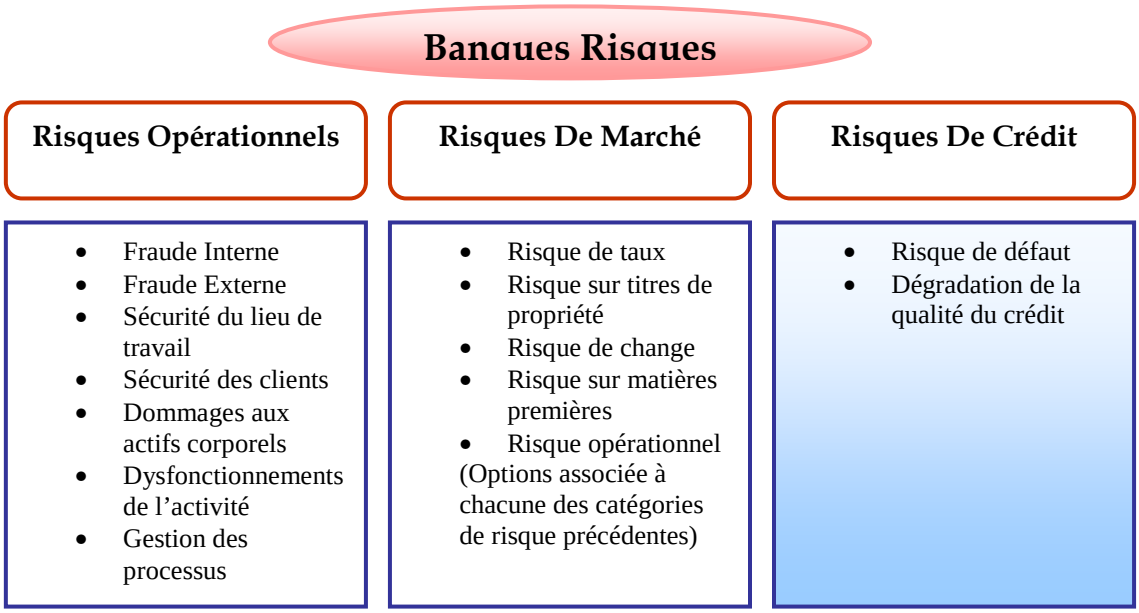



**Fig. 1.1 Présentation de l'ensemble des risques dans le cadre de Bâle II**

[4] Bilan : Bilan de la banque est constitué de l'Actif, qui représente tout ce que l'entreprise possède (Actifs immobilisés, Crédits et prêts, Titres, Trésorerie…), et du Passif, qui représente les dettes en générale (Fonds propres, Dettes, Dépôts…).

[5] Hors-bilan : Hors-bilan de la banque peut se définir comme étant un compte dans lequel sont enregistrées les sommes engagées par l'entreprise et qui ne sont pas encore disponibles ou n'ont pas encore été payées.

Dans le cadre de ce mémoire, nous allons nous intéresser principalement à la définition du risque de crédit.

## Section 1 Définitions du risque de crédit et de contrepartie

Le risque de crédit est le risque qu'un débiteur fasse défaut ou que sa situation économique se dégrade au point de dévaluer la créance que l'établissement détient sur lui. Ce risque est en effet lourd de conséquence pour toutes les banques : toute dette non remboursée est économiquement une perte sèche que supporte le créancier. En amont, le risque peut faire l'objet d'une évaluation grâce à différentes mesures. La présentation de ces mesures est citée dans le paragraphe suivant.

## Section 2 Mesure du risque de Crédit

Pour mesurer le risque de crédit, on va donc pondérer le montant total de la créance, ce qu'on appelle "l'encours pondéré" (RWA[6]), par la qualité du débiteur. La mise en place d'un système de mesure du risque de crédit permettant d'appréhender différents niveaux de risque à partir d'informations qualitatives et quantitatives.
La détermination du risque de crédit est la suivante :

$$RWA = f(PD; LGD) \times EAD$$

avec :

$f$ : Loi Normale
$PD$ : Probabilité de défaut de la contrepartie
$LGD$ (*Loss Given Default*) : Taux de perte en cas de défaut sur la ligne de crédit
$EAD$ : Encours au moment du défaut

Il existe trois méthodes pour l'évaluation du risque de crédit : méthode standard, méthode de base des notations internes et méthode avancée des notations internes.

### 2.1 Méthode standard

La méthode standard consiste à utiliser des systèmes de notation fournis par des organismes externes. Dans cette méthode, la $PD$ et le $LGD$ sont imposés par le régulateur (commission bancaire en France). A titre d'exemple, le $LGD$ est directement imposé par la Banque de France à 45% pour les créances de premier rang non assorties de sûretés reconnues[7] et 75% pour les créances subordonnées[8]. Ainsi la

[6] RWA : Risk Weighted Assets.
[7] Non assorties de sûretés reconnus : Il n'y pas de garantie intervenant dans les rapports du créancier et du débiteur.
[8] Créances subordonnées : une créance reposant sur une dette qui ne sera remboursé que lorsque les autres dettes Auront été remboursées.

*PD* à appliquer dépend des notes attribuées à la contrepartie par les agences de notation[9].

### 2.2 Méthode se base des Notations Internes (IRB[10]-Fondation)

Le dispositif de méthode des notations internes doit également satisfaire aux exigences liées au projet de réforme des ratios de solvabilité des banques. En effet, le recours aux systèmes de notation interne permet de quantifier le risque de crédit, notamment pour le calcul du ratio de solvabilité, fixant l'exigence en fonds propres réglementaire nécessaire à la couverture de ce dernier.

En méthode IRB - Fondation, la Banque utilise sa propre *PD* et le paramètre *LGD* reste imposé par la commission bancaire en France, qui est fixé à 45%.

### 2.3 Méthode avancée des notations internes (IRB-Avancée)

En méthode IRB - Avancée, la banque maitrise toutes ses composantes : La notation interne, *PD, LGD, EAD*.
En effet, la banque s'articule autour de la notation interne de la contrepartie à partir de laquelle est établie sa *PD* à un an. Ainsi l'estimation de facteurs de risques supplémentaires telles les *EL*[11] ou *UL*[12] se fait à l'aide d'estimations standard données.

Les banques sont encouragées à développer leurs propres systèmes d'évaluation du risque (notation interne, scores spécifiques, modèles internes) leur permettant d'optimiser leurs besoins de fonds propres.

Dans ce qui suit, nous allons définir la notion et l'évaluation de la probabilité de défaut. Cette dernière est un élément essentiel dans notre étude.

[9] Agences de notation : Les agences de notation (en anglais CRA, Credit Rating Agency) sont des entreprises privées dont l'activité principale consiste à évaluer la capacité des émetteurs de dette à faire face à leurs engagements financiers. Les principales agences présentes sur le marché sont Moody's, Standard & Poor's et Fitch Ratings.
[10] IRB : Internal Ratings Based.
[11] EL : Les pertes attendues.
[12] UL : Les pertes exceptionnelles.

# Chapitre 2 La Probabilité De Défaut Et Le Taux De Défaut

Une des difficultés, liée à l'élaboration d'un modèle de prévision de la défaillance, réside dans la définition du concept de défaut.

## Section 1 Définition de la probabilité de défaut

Avant de définir la probabilité de défaut, nous allons essayer d'approcher la notion de défaut dans le cadre de Bâle II.

### 1.1 Définition du défaut

La définition du débiteur en état de défaut est la suivante :
*Un emprunteur ou une contrepartie est en situation de défaut lorsque l'un ou l'autre des cas suivants est advenu :*

*- la banque estime que l'emprunteur ne sera pas en mesure d'honorer la totalité de ses obligations financières envers elle-même ou son groupe bancaire, à leur échéance, sans qu'elle recoure à une procédure particulière*

*- l'emprunteur est en retard de plus de 90 jours sur le paiement d'une échéance de crédit.*
(Bâle-Document consultatif 2003)

### 1.2 Définition de la probabilité de défaut

La probabilité de défaut modélise le taux de défaut moyen sur un horizon d'un an.
Pour calibrer cette probabilité de défaut, il faut au moins 5 ans d'historique d'après la réglementation Bâloise.
Pour la section suivante, nous allons nous intéresser, en particulier, à l'évaluation de la probabilité de défaut dans le cadre de l'approche interne.

## Section 2 Evaluation de la probabilité de défaut

### 2.1 Méthode standard

Comme nous l'avons vu précédemment, dans le contexte de l'approche standard, les indicateurs proviennent des sources externes à la banque (*i.e.* agence de notation, Banque de France…). Ainsi, la *PD* appliquée pour évaluer le risque de crédit dépend de la note calculée par l'agence de notations.
Néanmoins elle n'est pas directement déterminée par l'agence de notations. En pratique, le concept de *PD* est absent par la méthode standard, mais il est implicitement contenu dans la définition des pondérations. La formule de pondération se complète par la formule suivante :

$$Pondération = f(PD, LGD)$$

avec :

$f$ : Loi normale
$PD$ : Probabilité de défaut de la contrepartie
$LGD$ : Taux de perte en cas de défaut sur la ligne de crédit

La matrice de pondération standard est basée sur un découpage des notations externes. Ainsi elle établie en considérant un découpage des créances en trois catégories selon la nature de l'émetteur, désignant donc les créances des états, des banques et des entreprises. Chacune de ces catégories est attribuée d'une pondération en fonction de sa note. A titre d'exemple, la matrice de pondération suivante (*cf.***Tab.1.1**) nous illustre que la pondération va de 0% pour les Etats souverains (ce qui revient à dire que nous considérons les créances sur les Etats souverains comme sans risque) à 150% pour les contreparties les moins bien notées (ce qui revient à dire que les créances sur ces contreparties sont très risquées).

| Contrepartie | | Appréciations | | | | | |
|---|---|---|---|---|---|---|---|
| | | AAA à AA- | A+ à A- | BBB+ à BBB- | BB+ à B- | moins de B- | Non Noté |
| Etats | | 0% | 20% | 50% | 100% | 150% | 100% |
| Banques | | 20% | 50% | 100% | 100% | 150% | 100% |
| Soc | | | | | | | 100% |
| Opérations de détail | Autres | | | | | | 75% |

**Tab.1.1 Matrice de pondération standard définie par le Comité de Bâle**

## 2.2 Méthode de base des notations internes (IRB[13]-Fondation) du Group Crédit Agricole

Le Group Crédit Agricole utilise la méthode interne (IRB-Fondation) du dispositif de Bâle II sur l'ensemble de son périmètre. Dans le paragraphe suivant, nous allons présenter le périmètre étudié dans ce mémoire.

[13] IRB : Internal Ratings Based

### 2.2.1 Périmètre corporate

Le périmètre étudié correspond aux contreparties de type « Entreprise ». Il est à noter que le périmètre Corporate comprend toutes les entités à l'exception des entités juridiques des secteurs Banque, Assurance et Services Publics. La méthodologie de notations des entreprises du Group Crédit Agricole s'appliquera à seize secteurs d'activité dans notre étude.

### 2.2.2 Structure de la méthode interne

La méthodologie de la méthode interne est constituée d'une partie quantitative et d'une partie qualitative. Le modèle quantitatif est une combinaison de six facteurs financiers qui donne une note financière (NF). La partie qualitative contient 12 questions qui résultent d'une note qualitative (NQ). L'assemblage du modèle quantitatif et des réponses aux questions qualitatives donne la note système (NS) du modèle interne. La note finale (NCF) est automatiquement calculée par le système corrigé après confrontation avec des critères supplémentaires. Au moment de la décision finale (*cf.* **Fig1.2**), il convient de consulter l'analyste pour validation. En cas de forte divergence, il faut avoir une justification

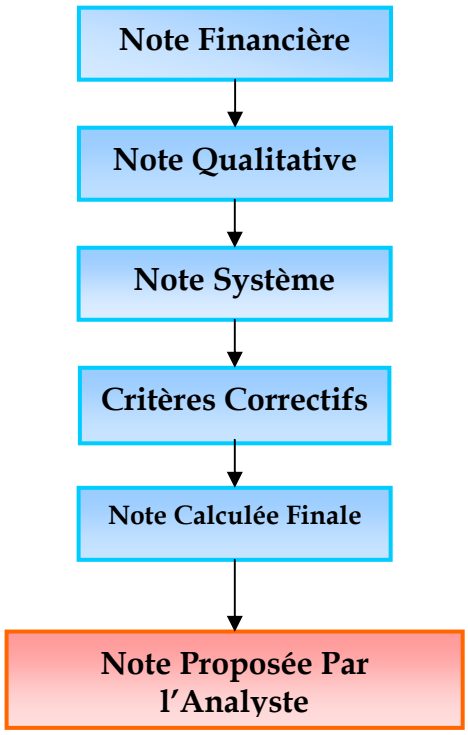



**Fig. 1.2 Procédure d'évaluation de la note finale interne**

### 2.2.3 La probabilité de défaut interne

Nous allons obtenir la probabilité de défaut à l'aide d'une fonction logistique qui est associé avec le score final issu d'un modèle interne précédent :

$$PD = \frac{1}{1 + e^{-score}}$$

avec :

*PD* : probabilité de défaut de la contrepartie
*SCORE* : score calculé avec le modèle interne

Dans le chapitre suivant, nous allons nous intéresser à la définition d'un test de stress scénarios, qui fait le sujet principal de nos études.

# Chapitre 3 Les Stress Scénarios De La Probabilité De Défaillance

Le pilier 2 du texte consultatif Bâle II incite les banques à mettre en place des modèles de stress scénario sur tous les segments du portefeuille.
L'objectif du test de stress scénario est double. D'abord, il s'agit d'effectuer de la prévention en s'assurant que les banques peuvent faire face à de mauvaises situations financières en estimant un montant suffisamment important de fonds propres. Ainsi, du point de vue des banques, les stress scénarios est un outil pour compléter leur stratégie de bonne gestion des risques. Dans ce chapitre, nous aborderons en section 1 la définition des tests des stress scénarios.

## Section 1 Définitions des tests des stress scénarii

Les stress tests visent à simuler les conditions macro ou micro économiques dégradées pour en mesurer l'impact sur les revenus, le coût du risque, les emplois pondérés et le ratio de solvabilité. L'objectif est de s'assurer que les banques disposent de fonds propres suffisants pour absorber des chocs sévères.

Dans cette optique, Bâle 2 prévoit que les stress tests conduits par la banque doivent intégrer les effets d'une forte augmentation des risques de crédit et de marché ainsi que ceux d'un accroissement du risque de liquidité. En section 2, nous nous concentrons principalement sur les stress tests du risque de crédit, qui se traduit essentiellement par le risque de défaillance.

## Section 2 Les stress tests du taux de défaillance

Dans le cadre de la simulation du taux de défaillance, nous mélangerons des dévaluations pour constituer le stress test : des hausses du taux de chômage, des baisses du PIB, des chutes de cours de CAC 40. Car la défaillance d'une entreprise peut être engendrée par une mauvaise gestion impactant sa rentabilité à long terme. Mais indépendamment de cela, l'entreprise peut faire faillite s'il y a une dégradation de l'environnement économique et financier.

L'impact de la dégradation de l'environnement économique et financier sur le taux de défaillance du portefeuille des entreprises de la banque. Il y a deux points à souligner dans ce test. D'un coté, il est important de trouver le lien causal qu'il existe entre le taux de défaillance de la banque et la situation économique. Ceci revient à trouver des variables macro-économiques qui permettent d'expliquer constamment la variation du taux de défaillance. D'autre coté, les variations de variables

explicatives sont déterminée par estimation. Or, cette estimation est faite par les experts économiques.
En ce sens, le modèle des stress tests mis en place pour le taux de défaillance de l'entreprise, est modélisé en fonction d'un ensemble restreint de variables macroéconomiques.
Ce modèle nous permet d'évaluer les relations passées entre le contexte macroéconomique et les taux de défaillance, et de mener une série de simulations selon scénarios.
La partie suivante vise à présenter des aspects théoriques des méthodes : la première approche est menée par le modèle à correction d'erreurs, alors que la méthode alternative est construite à l'aide du modèle linéaire généralisé.

# PARTIE II. METHODOLOGIES DE MODELE DU TAUX DE DÉFAILLANCE

# INTRODUCTION

Suite au passage aux exigences de solvabilité dans le cadre de Bâle II visant à bien quantifier les fonds propres couvrant le risque de crédit via la méthode standard ou la méthode interne, les autorités de contrôle s'intéressent de plus en plus à l'utilisation d'un modèle interne.

Dans le même temps, les banques ont mis au point des modèles sophistiqués visant à améliorer le niveau de compréhension des conséquences potentielles des risques qui seraient générés par leurs propres portefeuilles.

Le présent chapitre fournit une présentation des études et modèles sur la probabilité de défaillance, qui est un élément primordial pour évaluer le risque de crédit.
Dans notre étude, nous retenons deux modèles, d'une part, le modèle à correction d'erreurs, d'autre parte, le modèle linéaire généralisé.

Tout d'abord, nous allons présenter la procédure d'estimation en deux étapes proposées par Engle et Granger en matière théorique. Son principe inconvénient est que les estimations des paramètres de l'équation de cointégration peuvent être biaisées dans le cas de petits échantillons. En revanche, son avantage est d'être facile à mettre en ouvre et à interpréter.

Ensuite, nous allons nous intéresser au plan théorique du modèle linéaire généralisé. Ce dernier prend compte de la nature de la variable étudié : le nombre d'entreprises faisant faillite. Or, le résultat de la prévision de la variable ciblé dépend des variables explicatives de manière considérable. En conséquence, ceci pourrait engendre une dégradation de qualité de la prévision.

# Chapitre 1 Première Méthode : Modèle A Correction d'erreurs (MCE)

Le risque inhérent au portefeuille des prêts du secteur bancaire aux entreprises est la défaillance possible d'emprunteurs. Du point de vue de la stabilité financière du portefeuille, ce qui nous intéresse, c'est le genre de circonstances dans lesquelles un grand nombre d'entreprises pourraient se trouver en situation de défaillance. Plusieurs études ont été faites pour étudier les impacts de l'environnement économique su la défaillance de l'entreprise. En effet, les faibles performances en termes de prévision des modèles macroéconomiques à équations simultanées ont conduit à de nouvelles recherches sur l'économétrie des séries temporelles. L'analyse de la cointégration, présentée par Granger (1983) et Engle (1984) est considérée par beaucoup d'économistes comme un des nouveaux concepts qui constitue un des plus grands progrès dans le domaine de l'économétrie et de l'analyse de séries temporelles.

Une des difficultés de la modélisation d'un tel modèle est que beaucoup de variables macroéconomiques ne sont pas stationnaires. Ce chapitre a pour objectif d'introduire les problèmes induits par la non stationnarité des séries, le concept de cointégration et les tests de cointégration ainsi que le modèle à correction d'erreurs.

## Section 1 Stationnarité et cointégration

### 1.1 Stationnarité

Dans de très nombreux cas, pour que le modèle déduit à partir d'une suite d'observations ait un sens, il faut que chaque portion de la trajectoire observée de la suite fournisse des informations sur la loi de la série et que les portions différentes, mais de même longueur, aient les mêmes indications. C'est ce qui nous amène à définir la notion de stationnarité.

- **Définition de la stationnarité**

Soit $X=(X_t, t\in T)$ processus du second ordre, $X$ est faiblement stationnaire[14] si les conditions suivantes sont satisfaites :

| |
|---|
| $E(X_t)$ : est indépendant de $t$. |
| $V(X_t)$ : est une constante finie indépendante de $t$. |
| $Cov(X_t, X_{t-k})$ : est une fonction finie de $k$ et ne dépend pas de |

[14] Faible stationnaire : Voire la définition ci-dessus. Il y a une autre définition de la stationnarité : fortement stationnaire. Si un processus $X$ est dite fortement stationnaire si $P(x_{t_1,\dots}x_{t_k}) = P(x_{t_1+h,\dots}x_{t_k+h}), \forall k \geq 1, \forall(t_1,\dots,t_k)\in T, \forall h$ tel que $(t_{1+h},\dots,t_{k+h})\in T$

En pratique, L'importance de la faible stationnarité tient surtout aux problèmes de prédiction ou de régression. En effet, pour l'estimation des paramètres dans un modèle de prédiction, nous utilisons souvent des critères de moindres carrés. Cela signifie alors d'utiliser des prédicateurs linéaires optimaux dont le calcul ne fait pas intervenir, dans la totalité, la structure probabiliste du processus mais seulement les angles et longueurs de la suite. Or ces angles et longueurs ne dépendent que des moments d'ordre 2 du processus. La notion naturelle de stationnarité est l'invariance de ces moments d'ordre 2 par translation dans le temps.

### 1.2 Non stationnarité

La définition de la stationnarité nous conduit à définir deux types de non stationnarité selon que c'est plutôt la condition portant sur le moment d'ordre 1 qui n'est pas vérifiée (non stationnarité déterministe) ou les conditions portant sur les moments du second ordre qui ne sont pas vérifiées (non stationnarité stochastique).

- **Non Stationnarité Déterministe**

Le processus $X_t$ est caractérisé par une non stationnarité déterministe, ou encore que le processus $X_t$ est TS (*Trend stationary*), s'il peut s'écrire :

$$X_t = \mu + \beta t + \varepsilon_t$$

avec :

$\varepsilon_t$ : est un bruit blanc faible[15].

La série $X_t$ n'est pas stationnaire car $E(X_t) = \mu + \beta t$ dépend du temps. En revanche, sa variance est égale à celle du bruit blanc supposée constante. Pour stationnariser le processus, il faut enlever la partie déterministe $\mu + \beta t$. Car une série TS est caractérisée par un trend déterministe.

- **Non Stationnarité Stochastique**

Comme nous l'avons précédemment mentionné, il existe une autre forme de non stationnarité, provenant non pas de la présence d'une composante déterministe tendancielle, mais d'une source stochastique. Nous allons à présent introduire la définition des processus DS.

[15] Bruit blanc faible: un bruit faible est une suite $(\varepsilon_n)_{n\in Z}$ de variables aléatoires orthogonales, centrées et telle que $E(\varepsilon) = 0, Var(\varepsilon_n^2) = \sigma_\varepsilon^2 < \infty, \text{cov}(\varepsilon_n, \varepsilon_m) = 0, n \neq m.$ Ainsi, il y a une notion de bruit blanc fort, il faut remplacer la notion « orthogonales » par « indépendantes ». Un bruit blanc faible est faiblement stationnaire tant dis que un bruit blanc fort est strictement stationnaire.

Le processus $X_t$ est caractérisé par une non stationnarité stochastique, ou encore que le processus $X_t$ est DS (*Difference stationnary*) si le processus différencié une fois $X_t - X_{t-1}$ est stationnaire. Nous parlons aussi de processus intégré d'ordre 1, nous notons $X_t \rightarrow I(1)$.

De manière générale, le processus $X_t$ est un processus intégré d'ordre *d*, avec *d* le degré d'intégration. Nous pouvons rendre $X_t$ stationnaire en le différenciant *d* fois, nous notons
$X_t \rightarrow I(d)$ :

$$(1-L)^d X_t = Z_t$$

avec :

*L :* est une opération de retard telle que $(1-L)X_t = X_t - X_{t-1}$
$Z_t$ : est stationnaire.

Les exemples les plus connus de processus $I(1)$ sont :

Marche aléatoire pure : $X_t = X_{t-1} + \varepsilon_t$
Marche aléatoire avec dérive *c* :
$X_t = c + X_{t-1} + \varepsilon_t$
Où $\varepsilon_t$ = est un bruit blanc faible

La marche est aléatoire lorsque nous procédons par récurrence, par exemple pour $X_t$ :

$X_1 = X_0 + \varepsilon_1$
$X_2 = X_1 + \varepsilon_1 = X_0 + \varepsilon_1 + \varepsilon_2$
…
$X_t = X_0 + \sum_{i=1}^{t} \varepsilon_1$ Où $\varepsilon_t$ = est un bruit blanc faible
Ce processus est non stationnaire car nous avons :

$$Var(X_t) = Var(\sum_{i=1}^{t} \varepsilon_i) = \sum_{i=1}^{t} Var(\varepsilon_i) = \sum_{i=1}^{t} \sigma_\varepsilon^2 = t\sigma_\varepsilon^2$$

Nous constatons que la variance du processus $X_t$ dépend du temps t, plus $t \to \infty$ et plus $Var(X_t) \to \infty$.

Le fait de savoir si la série statistique est une réalisation d'un processus stationnaire, non stationnaire DS ou non stationnaire TS nous permet de déterminer l'ordre d'intégration.

### 1.3 Cointégration

La notion de cointégration permet de mettre en évidence des relations long termes stables entre des séries stationnaires. Avant de décrire la technique du modèle à correction d'erreurs, nous allons présenter quelques rappels sur la définition de la cointégration, et sur l'objectif du test de la cointégration.

- **Définition de la cointégration**

Soient deux séries temporelles $X_t$ et $Y_t$ non stationnaires ( $X_t \to I(d)$ et $Y_t \to I(d)$ ) sont dites **cointégrées**, si on a :

$$\varepsilon_t = Y_t - aX_t - b \to I(d-1)$$

Alors les séries $X_t$ et $Y_t$ sont notées :

$$X_t, Y_t \to CI(d,d)$$

Autrement dit, il est nécessaire que les séries soient intégrées d'ordre d et que les combinaisons linéaires de ces deux séries permettent de se ramener à une série d'ordre d'intégration inférieur.
Le concept de la cointégration reproduit l'existence d'un équilibre à long terme et le bruit blanc $\varepsilon_t$ peut s'interpréter comme une distance à l'instant t par rapport à cet équilibre. Car $\varepsilon_t$ représente les déviations qui vont fluctuer autour de zéro et donc cet équilibre strict se produira un certain nombre de fois tout au long de la période.
L'objectif d'introduire la notion de la cointégration est d'éviter une régression fallacieuse désignant des résultats trop optimistes dans une régression linéaire avec utilisation des séries temporelles non stationnaires, et qui font apparaître une relation pertinente entre les variables alors que ce n'est pas le cas.

- **Régression fallacieuse**

Nous allons introduire le problème de la régression fallacieuse sur une série de long terme en comparant deux situations différentes qui sont présentés dans les figures Fig.2.1 et Fig.2.2

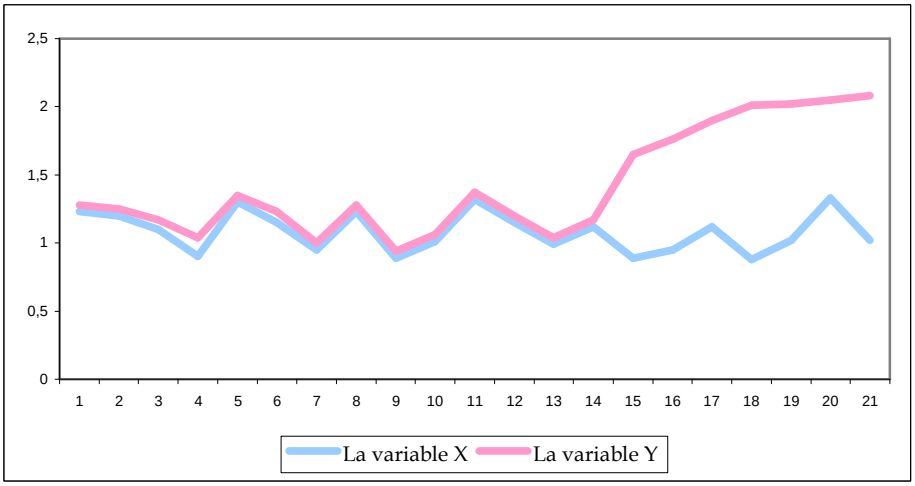



**Fig.2.1 Les variables Xt et Yt ne sont pas cointégrées**

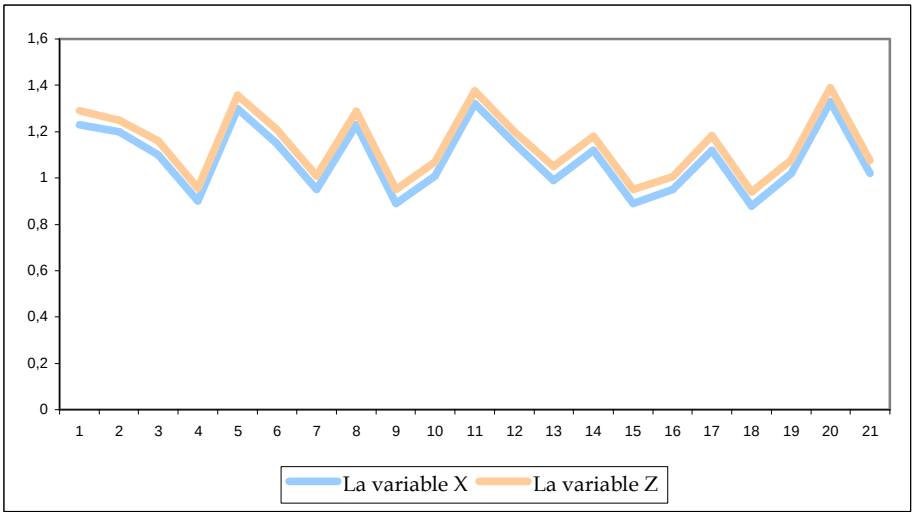



**Fig.2.2 Les variables Xt et Zt sont cointégrées**

- Dans un premier cas (*cf.***Fig.2.1**), les deux séries non cointégrées ont une tendance d'évolution constante sur une première période. En revanche, une tendance divergente s'installe dans une deuxième période.
- Dans un deuxième cas (*cf.***Fig.2.2**), les deux séries cointégrées ont une évolution constante sur toute la période.

En pratique, il faut toujours stationnariser des séries non stationnaires. Dans le cas contraire, il existe un risque de régression fallacieuse.
La situation de la régression fallacieuse qui se présente est que, dans ces opérations, le $R^2$ soit très élevé malgré l'absence de relation évidente entre les variables.
Reprenons l'exemple de la marche aléatoire pure :

Marche aléatoire pure 1 : $X_t = X_{t-1} + \varepsilon_t$
Marche aléatoire pure 2 : $Y_t = Y_{t-1} + \nu_t$
Où $\varepsilon_t, \nu_t$ = sont des bruits blancs faibles indépendants

Comme nous l'avons démontré précédemment, les deux variables ne sont pas stationnaires, puisqu'elles sont $I(1)$ (*cf.* 1.2 non stationnarité). Le $R^2$ de la régression entre la variable $X_t$ et la variable $Y_t$ devrait en principe être assez faible et tend vers zéro. Toutefois, il se peut qu'il n'en soit pas ainsi. De manière générale, la régression est dite fallacieuse, s'il existe une relation statistique significative entre deux variables, alors qu'en réalité elles n'ont aucun lien.

Pour éviter ce problème, nous pouvons effectuer la régression sur des variables stationnaires en les différenciant (*eg.* si $X_t \to I(1)$ et $Y_t \to I(1)$ , alors $(1-L)X_t \to I(0), (1-L)Y_t \to I(0)$ ).
La relation de cointégration est une relation d'équilibre entre les variables en régime de croissance équilibré mais des chocs peuvent affecter cette relation à court terme, c'est-à-dire avoir des effets temporaires. La section suivante consiste donc à déterminer les séries cointégrées puis à estimer la relation à long terme et de court terme entre deux variables.

## Section 2 L'approche de Engle Et Granger entre deux variables

Un éventuel test de cointégration est imposé sur les variables afin de mieux traiter les séries chronologiques longues dans l'approche de Engle et Granger. En effet, les risques de la régression fallacieuse et d'interprétation erronée sont très élevés dans une régression linéaire. Pour nos études, nous allons nous intéresser à l'exam du cas à deux variables : test de cointégration (i.e. relation à long terme) et estimation du modèle à correction d'erreurs (i.e. relation à court terme).

### 2.1 Test de coitégration

Le test de cointégration consiste en deux étapes. Dans un premier temps, nous allons déterminer l'ordre d'intégration des variables. Dans un second temps, nous nous concentrons sur l'estimation de la relation de long terme.

#### 2.2.1 Test de l'ordre d'intégration des variables

Comme nous l'avons précisé précédemment, la définition de la cointégration implique que toutes les variables intégrées soient de même ordre. Il convient donc de déterminer le type de tendance déterministe ou stochastique de chacune des variables. Car dans un cas de non stationnarité déterministe, nous n'avons pas besoin de différencier la variable pour la rendre stationnaire. Ensuite pour les variables de type de tendance stochastique, nous allons déterminer l'ordre d'intégration d à l'aide des tests de Dickey-Fuller et Dickey-Fuller augmenté (*cf.***Fig.2.3**).

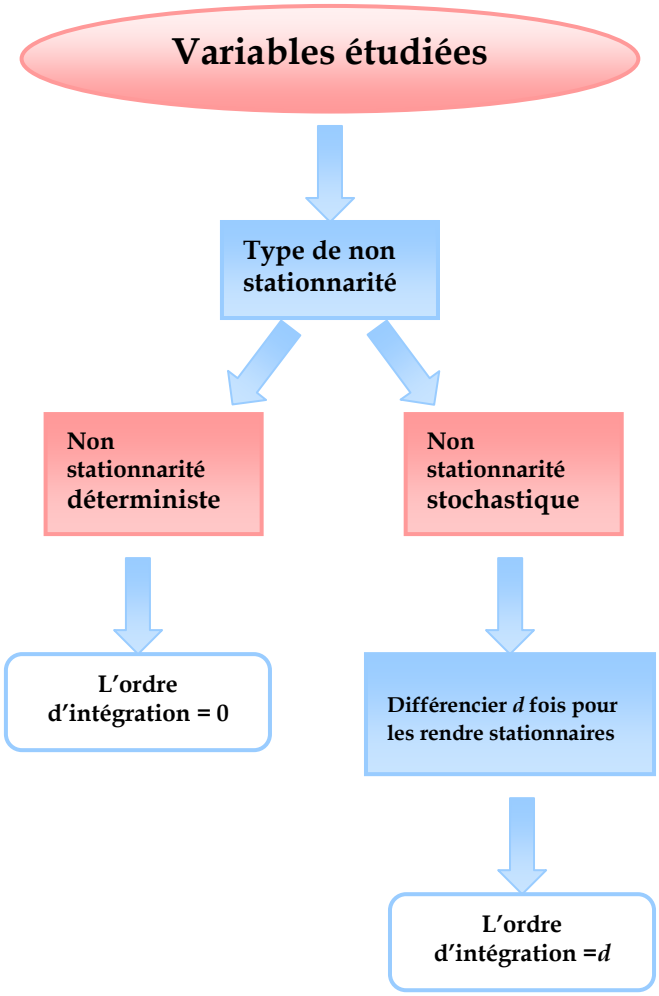



**Fig.2.3 Le processus du test de cointégration**

Tout ceci montre l'importance de bien choisir la méthode de stationnarisation des séries en fonction de l'origine de la non stationnarité. Il convient donc de présenter un test fondamental qui nous permet, tout d'abord de vérifier que les séries sont non stationnaires et d'autre part de discriminer les processus DS et TS afin de déterminer leurs ordres d'intégration : c'est le test de Dick-Fuller ou le test de racine unitaire.

- **Test de Dickey-Fuller**

Le test de Dickey-Fuller est un test de racine unitaire dont l'hypothèse nulle est la non stationnarité d'un autorégressif d'ordre un *AR (1)*[16].

Soit un processus $X_t$ est *AR(1),* il peut s'écrire :

$$X_t = \rho X_{t-1} + \varepsilon_t$$

avec :

$\varepsilon_t$ : est un bruit blanc faible.

Le principe général du test de Dickey-Fuller consiste à tester l'hypothèse nulle de la présence d'une racine unitaire :

$$\begin{array}{ll} H_0 : & \rho = 1 \\ H_1 : & |\rho| < 1 \end{array}$$

Sous l'hypothèse nulle $H_0$, la série se ramène à une marche aléatoire pure qui n'est pas stationnaire. Par conséquent l'hypothèse nulle testée correspond à une hypothèse de non stationnaire. Remarquons que le test de Dickey-Fuller est réalisé par une statistique de Student associée à l'hypothèse *H0* de non stationnarité. En revanche, la distribution asymptotique de l'estimateur du paramètre $\rho$ par la méthode des moindre carrée n'est pas standard. Par conséquent, nous ne sommes pas non plus dans une distribution asymptotique normale de base. De la même façon, la statistique de Student associé au test $\rho = 1$ n'a plus une distribution asymptotique standard. La preuve de la conclusion d'une distribution asymptotique non standard dans une série stationnaire fait l'objet de la section suivante.

- **Estimation d'un AR(1) par la méthode des moindres carrés**

Comme nous avons vu précédemment, un processus autorégressif *AR(P)* est exprimé par la forme suivante :

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + ... + \rho_p X_{t-p} + \varepsilon_t$$

De manière générale, l'étude des propriétés de l'estimateur des moindres carrés dans ce modèle sont plus difficile car la partie des répresseurs (i.e. $\rho_1 X_{t-1} + \rho_2 X_{t-2} + ... + \rho_p X_{t-p}$ ) est stochastique. Pour garantir une certaine

[16] Processus autorégressif *AR(p)* : Observation $X_t$ présente est générée par une moyenne pondérée des observations passées jusqu'à la *p*-ème période sous la forme suivante : $AR(p) : X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + ... + \rho_p X_{t-p} + \varepsilon_t$ où $\rho_1, \rho_2, ..., \rho_p$ sont des paramètres non nuls à estimer et $\varepsilon_t$ est un bruit blanc faible. Donc $AR(1) : X_t = \rho_1 X_1 + \varepsilon_t$ où $\rho_1 \neq 0$

indépendance entres les termes d'erreurs et les répresseurs à cause du caractère dynamique du modèle, nous faisons les deux hypothèses suivantes :

- Hypothèse 1: les termes d'erreurs $\varepsilon_t$ sont iid[17] avec $E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma_\varepsilon^2$.
- Hypothèse 2 : l'équation caractéristique :

$$1-\sum_{t=1}^{T}\rho_t z^t = 0$$

avec :

$|z|>1$, toutes les racines de $\rho$ sont en dehors du cercle unité.

Il convient d'adopter une écriture matricielle pour introduire le théorème suivant :

$$x = X\rho + \varepsilon$$

avec :

$x$ : est le vecteur de dimension $T$ représente les $T$ observations de $X_t$.
$X$ : est la matrice de dimension $T \times p$.

Théorème : Soit le modèle *AR(P)*, supposons que ce modèle satisfasse les hypothèses précédentes (hypothèse 1 et hypothèse 2), alors :

- La matrice $\frac{X'X}{T}$ converge en probabilité vers une matrice finie positive définie symétrique $Q$. (1)
- L'estimateur des moindres carrés $\hat{\rho}$ est consistant[18]. (2)
- $\sqrt{\frac{1}{T}}X'\varepsilon$ a une distribution limite Normale de moyenne nulle et de matrice de variance covariance $\sigma^2 Q$. (3)
- $\sqrt{T}(\hat{\rho}-\rho)$ a une distribution limite Normale de moyenne nulle et de matrice de variance covariance $\sigma^2 Q^{-1}$. (4)

Le lecteur intéressé peut se référer à Theil (1971) pour la démonstration de ce théorème.

[17] IID : indépendants et identiquement distribués.

[18] La consistance d'un estimateur : Un estimateur $\hat{\rho}_T$ est consistant $\lim \hat{\rho}_T = \rho$ (ie convergence en probabilité). Une condition suffisante pour un estimateur soit consistant est qu'il soit sans biais, $E(\hat{\rho}_T)=\rho$, et que sa variance tende vers zéro quand $T \to \infty$.

Les résultats du théorème sont très fragiles. Car ils disparaissent dès qu'il y a présence d'autocorrélation dans les résidus. De la même façon, ils ne sont pas valides lorsque les séries ne sont pas stationnaires.
Nous allons voir la preuve de cette suggestion avec un cas particulier *AR (1)* :

$$X_t = \rho X_{t-1} + \varepsilon_t \text{ , avec } |\rho| < 1 \text{et } \varepsilon_t \text{ est iid } (0, \sigma_\varepsilon^2)$$

La méthode la plus recommandée pour rechercher une relation affine entre la variable exogène et la variable endogène est la méthode des moindres carrés (MCO). Cette méthode consiste à rechercher une droite qui s'ajuste le mieux possible à un nuage de points qui représente les données dans un plan. Parmi toutes les droites possibles, nous retenons celle qui jouit d'une propriété remarquable : c'est celle qui rend minimale la somme des carrés des écarts des valeurs observées $X_t$ à la droite $\hat{X}_t$. Si $e_t$ représente cet écart, appelé aussi résidu, le principe des MCO est de choisir les paramètres $\rho$ qui minimisent :

$$S = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (X_t - \rho X_{t-1})^2$$

D'après les conditions du premier ordre, $\frac{\partial S}{\partial \rho} = 0$, la solution est donnée par :

$$\sum_{t=1}^{T} (-2 X_t X_{t-1} + 2\hat{\rho} X_{t-1}^2) = 0$$

La résolution de l'équation est donnée :

$$\hat{\rho} = \frac{\sum_{t=1}^{T} X_t X_{t-1}}{\sum_{t=1}^{T} X_{t-1}^2}$$

En remplaçant $X_t$ par la valeur donnée par le modèle $X_t = \rho X_{t-1} + \varepsilon_t$, nous obtenons :

$$\hat{\rho} = \rho + \frac{\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t X_{t-1}}{\frac{1}{T} \sum_{t=1}^{T} X_{t-1}^2}$$

D'après les résultats du théorème, nous constatons que :

Si $|\rho|<1$, alors $\lim \frac{1}{T}\sum_{t=1}^{T} X_{t-1}^2 = E(X_{t-1}^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$ qui tend vers zéro d'après (1)

Ensuite, Si les termes d'erreurs sont *iid*, alors $\lim \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t X_{t-1} = 0$ d'après (3)

Considérons maintenant la transformation :

$$\sqrt{T}(\rho - \hat{\rho}) = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_t X_{t-1}}{\frac{1}{T}\sum_{t=1}^{T} X_{t-1}^2}$$

D'après le théorème de transformation de Cramer (1941), il vient que :

$$\sqrt{T}(\rho - \hat{\rho}) \xrightarrow{Loi} N(0,1-\rho^2) \text{ (Convergence en loi[19])}$$

Nous pouvons donc constater qu'une hypothèse implicite de stationnarité est faite dans l'hypothèse 2. La raison pour laquelle nous faisons cette hypothèse est qu'il n'y pas de convergence $\lim_{T\to\infty} \frac{X'X}{T} = Q$ avec $Q$ une matrice finie positive définie symétrique d'après le théorème de Slutsky. Il suffit que $X$ comporte un trend (ie non stationnarité déterministe) pour que $\frac{X'X}{T}$ ne converge plus vers une matrice finie $Q$. Donc si nous sortons la condition de la stationnarité, nous perdons le résultat de Normalité asymptotique. Par conséquent, pour mener le test de Dickey-Fuller, nous ne pouvons plus utiliser la statistique de Student standard car les propriétés asymptotiques ne sont plus standards quand la série est non stationnaire. Ainsi, les valeurs critiques à utiliser pour conclure le test ne sont plus les mêmes. Dickey et Fuller ont calculé les nouvelles valeurs critiques (*cf.* **Annexe.1**). Les 1%, 5% et 10% quantiles sont respectivement -3.43,-2.86 et -2.57. Le test de D.F rejette donc H0 (ie l'hypothèse de non stationnarité) au niveau 5% si le t-ratio20 est plus petit que -2.86. Remarquons que cette valeur -2.86 est beaucoup plus petite que celle correspondant à une approximation normale (i.e -1.645), nous rejetterons donc moins facilement l'existence d'une racine unitaire en utilisant la loi asymptotique standard.

[19] Convergence en loi : La suite de variables aléatoires $X_T$ converge en loi vers une variable aléatoire $X$ de distribution $F_\infty(x)$ si en tout point de continuité $x$ de $F_\infty$, nous avons $\lim_{T\to\infty} F_T(x) = F_\infty(x)$.

[20] T-ratio : T ratio est la statistique de test qui suit une loi de Student à *n-1* avec n est le nombre d'observations.

A l'issue du test de Dickey et Fuller, nous pouvons constater l'ordre d'intégration des variables concernées. En s'appuyant sur ces résultats retenus, nous continuons à estimer la relation de long terme du modèle à correction d'erreurs dans le paragraphe suivant.

### 2.2.2 Estimation de la relation de long terme

Soient les séries $X_1, X_2, ..., X_t \to I(1)$ et $Y_t \to I(1)$ , la procédure de Engle et Granger suggère l'étape suivante : estimation avec la méthode des Moindres carrés la relation de long terme :

$$Y_t = a_1 X_1 + a_2 X_2 + ... + a_t X_t + b + \varepsilon_t$$

La méthode d'estimation de base des paramètres de la relation de long terme est la méthode des Moindres Carrés Ordinaires. Dans le paragraphe suivant, nous allons faire un appel sur les principes et les priorités du modèle MCO.

- **Méthode des moindres carrés**

Dans le cas d'une régression linéaire, nous cherchons à expliquer une variable quantitative $Y_t$ en fonction de plusieurs variables explicatives $X_1, X_{2,...,} X_t$ où $t$ est le nombre de variables.
Le modèle linéaire multiple s'écrit de manière suivante :

$$Y_t = a_1 X_{1t} + a_2 X_{2t} + ...a_k X_{kt} + b + + \varepsilon_t$$

Le modèle utilisé implique plusieurs hypothèses :

- Le modèle est linéaire
- $\varepsilon_t$ est iid, $E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2$
- $\varepsilon_t$ est indépendante de la variable explicative

L'analyse des erreurs est importante, puisqu'elle permet de diagnostiquer un modèle non approprié (i.e. non linéaire, colinéaire, etc). La méthode des moindres carrés que nous utilisons pour estimer les paramètres $a_1, a_2, ..., a_{k,} b$ consiste à minimiser la distance au carré entre chaque individu et la droite de régression d'où l'appellation, une minimisation au sens des moindres carrés ordinaires (MCO). Une fois nous avons estimé des paramètres $\hat{a}_1, \hat{a}_2, ..., \hat{a}_{k,} \hat{b}$ , nous pouvons entamer la phase de prédiction de la variable réponse $\hat{Y}_t$ . Par déduction, nous obtenons l'erreur du modèle :

$$\varepsilon_t = Y_t - \hat{Y}_t$$

Ainsi, nous avons besoin d'un bon indicateur de la qualité de notre modèle qui est le coefficient de détermination : $R^2$. En effet, la variable à expliquer $Y_t$ possède une variance totale qui peut être divisée en deux parties, une partie de cette variance est expliquée par l'équation du modèle. Une autre partie est expliquée par le terme d'erreur. Le coefficient $R^2$ est le rapport entre la variance due au modèle et la variance totale. Il est compris entre 0 et 1. Plus $R^2$ se rapproche de la valeur 1, meilleure est l'adéquation du modèle aux données. Un $R^2$ faible signifie que le modèle à un faible pouvoir explicatif.

Le test de stationnarité à lui seul ne permet pas de savoir si les variables sont en relation d'équilibre de long terme et si la corrélation qui existe entre elles n'est pas une fausse corrélation, c'est-à-dire une régression fallacieuse. Pour le savoir, il est nécessaire de réaliser le test de cointégration.

Pour que la relation de cointégration soit acceptée, le résidu $e_t$ issu de cette régression doit être stationnaire :

$$e_t = Yt - \hat{a}_1 X_{1t} + \hat{a}_2 X_{2t} + ...\hat{a}_k X_{kt} \rightarrow I(0)$$

La stationnarité de l'erreur est testée à l'aide des tests Dickey-Fuller .
Des tests de Dickey-Fuller de stationnarité sur le résidu doivent s'effectuer à partir des valeurs critiques tabulées par MacKinnon (1991) (*cf.***Annexe.2**) en fonction du nombre de variables total du modèle. La table de Mckinnon est calculée par la méthode de Monte Carlo. (MacKinnon 1991).
Nous pouvons, dès lors, estimer le modèle à correction d'erreurs qui intègre les variables en variation et en niveau.

### 2.2.3 Estimation du modèle à correction d'erreurs

Lorsqu'une relation est cointégrée (ie. $e_t = Yt - \hat{a}_1 X_{1t} + \hat{a}_2 X_{2t} + ...\hat{a}_k X_{kt} \rightarrow I(0)$), le modèle à correction d'erreurs est sa meilleur représentation de cour terme (Engle et Granger).

L'équation du modèle à correction d'erreurs peut s'écrire de la façon suivante :

$$\Delta Yt = \gamma_1 \Delta X_{1t} + \gamma_2 \Delta X_{2t} + ... + \gamma_k \Delta X_{kt} + \delta e_{t-1} + \nu_t \text{ où } \delta < 0$$

avec :

$e_{t-1} : Y_{t-1} - \hat{a}_1 X_{1t-1} + \hat{a}_2 X_{2t-1} + ...\hat{a}_k X_{kt-1}$
$\nu_t$ : est un bruit blanc faible

Le modèle à correction d'erreurs a pour objectif de reproduire la dynamique d'ajustement vers l'équilibre de long terme. D'une part, nous devons retirer la relation commune de cointégration (i.e. la tendance commune), d'autre part, rechercher la liaison réelle entre les variables. La représentation du modèle à correction d'erreurs est à la fois un modèle statique (i.e. $\gamma_1 \Delta X_{1t} + \gamma_2 \Delta X_{2t} + ... + \gamma_k \Delta X_{kt}$) et un modèle dynamique (i.e. $\delta e_{t-1}$).

Ainsi ce modèle décrit la variation de $Y_t$ autour de sa tendance de long terme. Remarquons que les variables $e_{t-1}$ , $\Delta Yt$ , $\Delta X_{1t}$ ,…, $\Delta X_{kt}$ sont *I(0)*, c'est-à-dire stationnaires. Le fait qu'elles soient stationnaires évite une estimation très optimiste dans la régression linéaire (i.e. régression fallacieuse). Par conséquent, les estimateurs des MCO sont plus efficaces.

Autour de la relation de long terme, le modèle à correction d'erreurs permet d'intégrer les fluctuations de court terme. Le coefficient $\delta$ négatif donne une force de rappel vers l'équilibre de long terme.

La dynamique de court terme s'écrit :

$$Yt = a_0 + a_1 Y_{t-1} + a_2 X_{1t} + a_3 \Delta X_{1t-1} + ... + \gamma_{2k} X_{kt} + \gamma_{2k+1} X_{kt-1} + \nu_t$$

La littérature sur le modèle à correction d'erreurs est très volumineuse, notamment sur le plan de la théorie statistique. Nous avons tenté de donner un fil conducteur sur la partie statistique. D'abord, nous nous sommes attachés à présenter le test de stationnarité et celui de cointégration. Car ces tests sont des instruments de modélisation utiles pour spécifier un modèle à correction d'erreurs. Puis nous avons essayé de montrer l'hypothèse et le principe de la relation long terme et du modèle à correction d'erreurs. Compte tenu du fait de la cointégration entre plusieurs variables, c'est-à-dire qu'une série macro économique stationnaire peut être le résultat d'une combinaison de variables non stationnaires, l'approche de Engle et Granger ne permet pas de distinguer plusieurs relations de cointégration.

En effet, si nous étudions simultanément *t* variables, nous pouvons avoir jusqu'à *t-1* relations de cointégration. Le VECM (*Vector Error Correction Model*) est le meilleur modèle pour résoudre cette difficulté. L'introduction de la possibilité de cointégration entre les variables complique grandement les choses. Il devient alors difficile de construire un modèle structurel qui ait valeur de généralité. Pour une revue détaillée sur le modèle VECM, le lecteur peut se référer à Johanson (1988).

Il existe des nombreux modèles pour étudier le lien entre les variables à expliquer et les variables explicatives. En effet, dans tout relevé d'expérience, les données présentent un effet aléatoire. Nous allons donc passer maintenant par le choix d'une méthode statistique adaptée à la nature de cet effet aléatoire. Les modèles linéaires classiques ne contiennent que des effets fixes. En revanche, les effets aléatoires ont été introduits dans les modèles linéaires généralisés, qui sont des modèles de régression ordinaire, où la variable à expliquer peut suivre une distribution autre que la normale. De ce fait, le modèle permet de mieux comprendre comment les variables à expliquer sont liées aux variables explicatives en tenant compte d'une part un effet d'aléatoire et d'autre part une représentation simple. Dans ce cadre, le modèle linéaire généralisé présente ces deux qualités : il respect bien la réalité et sa linéarité

lui confère une simplicité très attractive. Le chapitre fait l'objet de l'introduction du modèle linéaire généralisé.

# Chapitre 2 Deuxième Méthode : Modèles Linéaires Généralisés – Régression De Poisson

Le modèle à correction d'erreurs est utilisé pour analyser la causalité entre les variables, au sens de Wiener-Granger. Ces analyses de causalité sont fondées sur une spécification autorégressive du processus générateur des données. Nous avons décrit la version la plus simple de ce modèle dans le chapitre précédent. Cette version suppose qu'il n'y pas de relation de cointégration entre les variables explicatives. Toutefois, cette hypothèse pourrait polluer le résultat attendu. Cela suggère la nécessité de recourir à une autre modélisation afin d'avoir plusieurs pistes.
Nelder & Wedderburn (1972) ont introduit le modèle linéaire généralisé, qui permet d'analyser la relation de causalité entre une variable à expliquer de la famille exponentielle et des variables explicatives par des techniques de régression.
En se référant à McCullagh et Nelder[21] (1989), dans ce chapitre, nous allons parler dans les deux premières parties de la structure du GLM, ainsi que le principe de l'estimation. Une troisième partie sera entièrement consacrée à la construction d'un GLM. Nous nous concentrons sur un cas particulier du GLM-régression de Poisson dans la dernière partie.

## Section 1 Structure du GLM

Le modèle linéaire généralisé permet d'étudier la liaison entre une variable réponse $Y$ et un ensemble de variables explicatives $X_1, X_2, ..., X_k$ selon le schéma suivant.

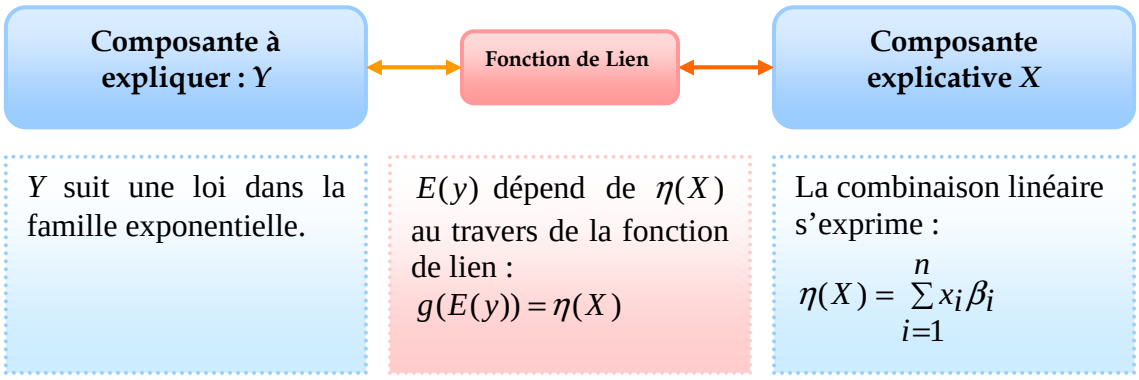



**Fig.2.4 La structure du GLM**

[21]Mc Cullagh et Nelder : Les modèles linéaires généralisés sont présentés pour la première fois par Nelder et Wedderburn (1972), et exposé de façon complète par Mc Cullagh et Nelder (1989).

Globalement, il y a trois techniques de modélisation : le modèle log-linéaire, la régression logistique et la régression de poisson. Ainsi, le GLM se caractérise par les trois composantes suivantes :

- **Composante aléatoire**

La loi de probabilité de la composante aléatoire appartient à la famille des lois exponentielles. Notons $Y_1, Y_2, ..., Y_k$ un échantillon aléatoire de taille n de la variable à expliquer $Y$, les variables $Y_1, Y_2, ..., Y_k$ étant supposées indépendantes. Parmi la famille des lois exponentielles, il y a trois possibilités de lois en fonction de la caractéristique de la variable (*cf.***Fig.2.5**).

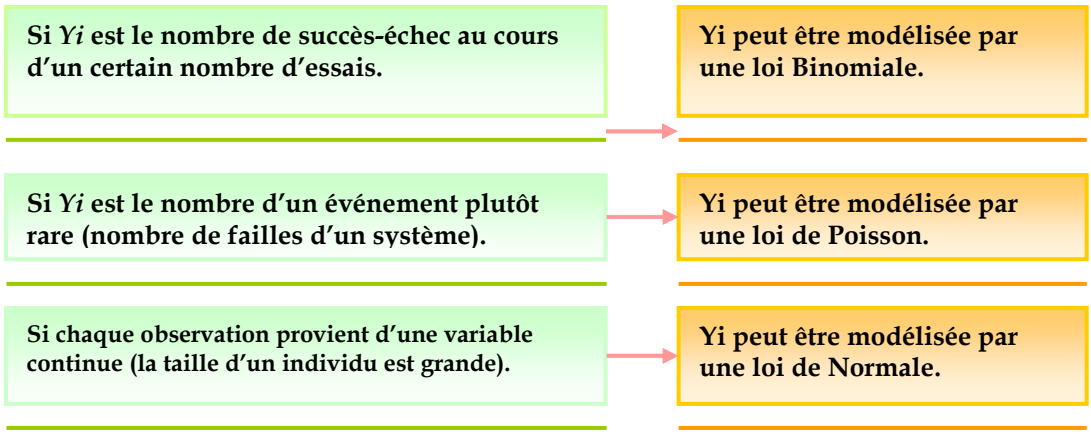



**Fig.2.5 La famille des lois exponentielles**

De manière générale, la famille exponentielle a la forme suivante :

$$f(y_i, \theta_i, \phi, \omega_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi, \omega_i)\right)$$

avec :

*a, b, c* : sont les fonctions spécifiées en fonction du type de famille exponentielle.
$\theta_i$ : est le paramètre canonique inconnu et dépend de la fonction de lien
$\theta_i$ : : $g(\mu_i)$ où $\mu_i$ est l'espérance mathématique de la variable aléatoire $Y$.
$\varphi$ : est le paramètre de dispersion[22] supposé connu
$\omega_i$ : est un poids. Pour des données non groupées $(i = 1,...,n)$, $\omega_i = 1$

- **Moyenne et Variance**

Pour une variable $Y$ dont la densité peut se mettre sous la forme précédente, nous pouvons exprimer la moyenne et la variance de manière suivante :

[22] Paramètre de dispersion : Le paramètre de dispersion contrôle la variance qui est considéré comme un paramètres de nuisance

$$E(Y_i) = b'(\theta_i) \text{ et } V(Y_i) = b''(\theta_i)\frac{\phi_i}{\varpi_i}$$

avec :

' et '' : désignent les dérivées premières et secondes par rapport à $\theta_i$.

Toutes les lois de probabilité dont la densité peut se mettre sous la forme précédente. Le tableau ci-dessous (*cf.***Tab.1.2**) indique les composantes de la famille exponentielle pour des lois de probabilité usuelles, l'espérance et la variance.

| **Distribution** | $\theta(\mu)$ | $b(\theta)$ | $a(\varphi_0)$ | $E(Y)$ | $V(Y)$ |
|---|---|---|---|---|---|
| Normale $N(\mu,\sigma^2)$ | $\mu$ | $\frac{\theta^2}{2}$ | $\sigma^2$ | $\theta$ | $\sigma^2$ |
| Bernoulli $B(1,\mu)$ | $\log\frac{\mu}{1-\mu}$ | $\log(1+e^{\theta})$ | 1 | $frac(e^{\theta}1+e^{\theta})$ | $\mu(1-\mu)$ |
| Poisson $P(\mu)$ | $\log\mu$ | $_e\theta$ | 1 | $_e\theta$ | $\mu$ |

**Tab.1.2 Les composantes de la famille exponentielle**

- **Composante déterministe**

Dans un modèle linéaire généralisé, la composante déterministe est exprimée sous forme d'une combinaison linéaire $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$

- **Fonction de lien**

La troisième composante d'un modèle linéaire généralisé est le lien entre la composante aléatoire et la composante déterministe. Le paramètre de fonction de lien est l'espérance mathématique de la variable aléatoire $Y$, noté $\mu$. Nous pouvons modéliser la fonction de lien par une fonction monotone $g(\mu)$ de l'espérance. Nous avons alors :

$$g(\mu_i) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i}$$

avec :

$\mu_i : E(Y_i)$

Pour la suite, il convient d'adopter une écriture matricielle pour le modèle :

$$g(\mu_i) = X_i^{'}\beta$$

avec :

$\beta$ : le vecteur des paramètres inconnus, $\beta = (\beta_0, \beta_1, ..., \beta_k)$

La fonction de lien se résume en trois formes suivantes :
- Fonction de lien canonique : $g(\mu) = \mu$
- Fonction de lien log-linéaire: $g(\mu) = \log(\mu)$
- Fonction de lien logit : $g(\mu) = \log \frac{\mu}{1-\mu}$

A toute loi de probabilité de la composante aléatoire est associée une fonction spécifique de l'espérance appelée paramètre canonique. La figure suivante (*cf.***Fig.2.6**) montre les lois exponentielles et leur fonction de lien canonique correspondante :

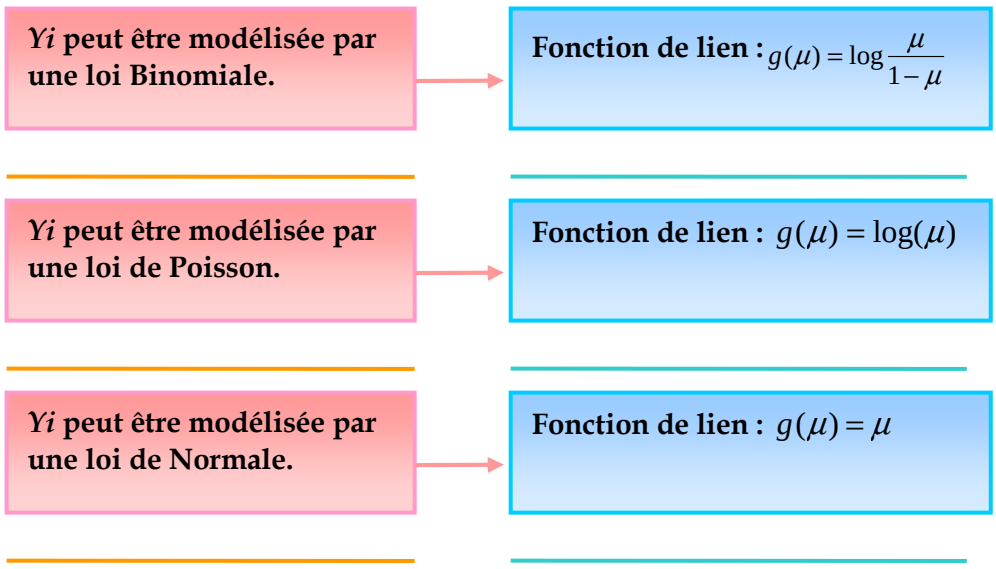



**Fig.2.6 Les fonctions de lien correspondantes**

## Section 2 Principe d'Estimation d'un GLM

En pratique, les coefficients de régression $\beta_0, \beta_1, ..., \beta_k$ sont inconnus et doivent être estimés sur base de données. Dans cette section, nous nous concentrons sur l'estimation des coefficients de régression $\beta$ par la méthode du maximum de vraisemblance

La procédure de la méthode du maximum de vraisemblance est basée sur l'algorithme d'optimisation. Cet algorithme utilisé est de type Newton-Raphson. Avant de décrire l'algorithme de Newton-Raphson, nous allons faire un bref rappel sur la maximisation de la vraisemblance dans sous section.

- **Méthode du maximum de vraisemblance**

Soit $Y$ une variable aléatoire de densité de probabilité $f(y,\theta)$ dont l'un des paramètres $\theta$ est inconnu. La méthode du maximum de vraisemblance (*MMV*) consiste à trouver la valeur numérique la plus vraisemblable pour le paramètre $\theta$. Supposons que les variables aléatoires sont indépendantes mais non identiquement distribuées $Y_1, Y_2, ..., Y_n$ dont la densité est de la forme précédente. Supposons que les $\theta_i$ sont en fonction de $\beta$, la fonction de vraisemblance (*likelihood*) est définie comme le produit des probabilités :

$$Vraisemblance = L = \prod_i f(y_i, \theta_i(\beta)), i = 1,2,...,n$$

Cette définition a l'avantage de donner à $L$ une interprétation naturelle. En raison de l'indépendance des observations individuelles. La valeur de la vraisemblance de la distribution est considérée comme une unique observation engendrée par la distribution multivariée.
Le principe de la MMV est de rechercher la valeur de $\beta$ qui rend la fonction de vraisemblance maximale. La valeur $\hat{\beta}$ est obtenue en résolvant l'équation :

$$\frac{\partial InL}{\partial \beta} = 0 \Rightarrow \hat{\beta} : \frac{\partial^2 InL}{\partial \hat{\beta}^2} < 0$$

L'emploi du logarithme sur la fonction $L$ permet de passer de la maximisation d'un produit à celle d'une somme. Le résultat reste le même car la fonction logarithme est monotone strictement croissante. Pour trouver la valeur $\hat{\beta}$ dans telle fonction, la méthode de Newton - Raphson est fortement recommandée (*cf.* **Annexe.3** : le principe de la méthode de Newton - Raphson).

- **Algorithme de Newton - Raphson**

D'après le principe de l'algorithme de Newton - Raphson, sur le i+1 ème itération, le paramètre s'écrit :

$$\beta_{i+1} = \beta_i - H^{-1} s$$

avec :

$H$ : est la matrice de Hessian, $H = \left[ \frac{\partial^2 L}{\partial \beta_i \partial \beta_j} \right]$

$s$ : est le gradient $s = \left[ \frac{\partial L}{\partial \beta_j} \right]$

Après avoir présenté l'aspect théorique du modèle linéaire généralisé, nous allons décrire les conditions essentielles pour choisir un GLM plus adapté à nos études des stress scénarios.

## Section 3 Construction d'Un GLM

Le plus souvent le choix d'un modèle linéaire généralisé est autour de deux questions : d'une part le choix de la loi de probabilité de la variable à expliquer et d'autre part l'adéquation de la fonction de lien.

- **Choix de la loi de probabilité et de la fonction de lien**

La modélisation GLM prend les principes exposés dans la section précédente. Les principales caractéristiques peuvent être résumées :

-Une transformation de l'espérance (i.e. plus précisément : espérance conditionnelle) par la fonction de lien $g$ : $g(E(Y)) = \eta(x)$

-Cette transformation de l'espérance est modélisée par une combinaison linéaire des variables explicatives : $\eta(x) = \sum_{i=0}^{n} \beta_i x_i$ avec $x_0 = 0$

-La variable aléatoire $y$ suit une loi de la famille exponentielle

Pour choisir un modèle GLM, il faut d'abord choisir la loi de la variable à expliquer $y$ dans la famille exponentielle, ce qui fixe la moyenne, la variance. Ensuite il convient de déterminer une fonction de lien. Comme nous avons vu dans la section précédente, il y a toujours une fonction de lien qui s'associe à la loi de probabilité (e.g. loi Binomiale avec la fonction de lien canonique…).
Cependant, il est toujours possible d'utiliser d'autres fonctions de lien. Ainsi, nous pouvons parfois essayer différentes lois de probabilité associées à la variable aléatoire, et retenir celle qui au meilleur indicateur statistique. Nous allons décrire des indicateurs statistiques importants du GLM dans le paragraphe suivant.

- **Adéquation du modèle**

Après avoir choisi la loi de probabilité et la fonction de lien, estimé les paramètres, il convient de juger de l'adéquation du modèle. Il y a deux statistiques qui sont utilisé principalement : La déviance normalisée (*scaled deviance*), la statistique du khi-deux de Pearson.

- **La déviance normalisée**

Nous définissons la qualité d'un modèle en prenant comme référence le modèle saturé qui est caractérisé par $\hat{\mu}_i = y_i$ avec $i = 1,2,...,n$. Car un modèle saturé contient autant de paramètres que d'observations, ceci suggère que le modèle donne

une description parfaite des données. La fonction de vraisemblance associée à ce modèle sera notée $l(y,y)$. Le principe de la statistique proposée est de comparer la différence entre la fonction de vraisemblance du modèle saturé et celle du modèle initial $l(y,\hat{\mu})$. Le modèle décrira bien les données lorsque $l(y,\hat{\mu}) \approx l(y,y)$ et mal lorsque $l(y,\hat{\mu}) \prec\prec l(y,y)$. Ceci introduit la statistique pour mesure la qualité de l'ajustement des données par un modèle. Définissons la statistique :

$$D = 2In\frac{l(y,y)}{l(\hat{\mu},y)}$$

Celle-ci est appelée la déviance réduite dans le cadre des modèles linéaires généralisés. Remarquons qu'une petite valeur de la déviance indique un ajustement de bonne qualité, puisque la vraisemblance du modèle est proche de celle du modèle saturé. Au contraire, une grande valeur de la déviance traduira un mauvais ajustement.

- **La statistique du khi-deux de Pearson**

Le deuxième indicateur statistique pour juger la qualité du modèle est la statistique du khi-deux de Pearson :

$$X^2 = \sum_{i=1}^{n} w_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

La statistique est distribué exactement comme une loi khi-carrée dans le cas particulier du modèle linéaire gaussien puisque la déviance réduite d'une loi normale a la forme suivante : $D_{normale} = \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2$ en supposant les poids $\omega_i = 1$

De manière générale, la statistique de $X^2$ est distribuée asymptotiquement comme une loi khi-carrée dans les autres cas du modèle linéaire généralisé.

## Section 4 Régression de Poisson

Dans notre contexte, une entreprise se comporte comme une expérience aléatoire qui a deux issues possibles : saine ou défaillante. On notera *p* la probabilité de défaillance de l'entreprise et q la probabilité qu'elle ait une absence de défaut. Une telle expérience peut être modélisée par la loi de Bernoulli de paramètre *p*. Si nous supposons que l'ensemble des entreprises tombent en défaillance de manière indépendante, nous pouvons modéliser la probabilité de défaillance sur un portefeuille contenant *n* entreprises par une loi binomiale de probabilité *p* et *q* respectivement.
Toutefois, nous observons que la défaillance des entreprises reste un événement rare dans les conditions générales. Nous pouvons donc avoir recours à une loi de Poisson

pour modéliser le nombre d'entreprises défaillantes en moyenne sur une période donnée. La variable aléatoire suit une loi de probabilité définie par :

$$f(y,\lambda) = e^{-\lambda}\frac{\lambda^{y}}{y!} = \exp(yIn\lambda - \lambda - Iny!)$$

avec :

$y$ : variable de contage, indépendante.
$\lambda$ : est un nombre réel strictement positif.

Une fois que nous avons déterminé la loi de probabilité de la variable de réponse, la fonction de lien est généralement associée à la loi de probabilité. Comme nous l'avons déjà vu dans la section précédente, la fonction de lien de la loi de Poisson est la fonction de logarithme ( $g(\mu) = \log(\mu)$ ). Maintenant, nous allons déterminer la fonction de vraisemblance de la loi de Poisson :
Supposons que nous disposons des réalisations de défaillance $y_1, y_2, ..., y_k$ de variables aléatoires et donc qu'ici

$$l_n(y_1, y_{2,...,}y_n, \lambda) = Ln(y_1, y_{2,...,}y_n, \lambda) = \prod_i In(e^{-\lambda}\frac{\lambda^{y_i}}{y_i!}) = -n\lambda + \sum_{i=1}^{n} y_{iIn(\lambda)-}\sum_{i=1}^{n} In(y_i!)$$

avec :

$i = 1,2,...,n$

Si nous estimons le paramètre $\lambda$, alors nous devons dériver la fonction :
Cette fonction est concave et son extremum $\hat{\lambda}$ est donc le zéro de la dérivée

$$\frac{\partial}{\partial\lambda} l_n(y_1, y_2,...,y_n, \lambda) = -n + \frac{\sum_{i=1}^{n} y_i}{\lambda}$$

c'est-à-dire : $\hat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{n}$

Pour estimer les paramètres de régressions $\beta_0, \beta_1, ..., \beta_k$, nous pouvons par exemple utiliser la méthode de Newton - Raphson, que nous rappelons brièvement dans l'annexe.
En raison de sa complexité, nous n'allons pas donner sa démonstration ici. Le lecteur intéressé peut consulter GOURIEROUX, MONFORT & TRONGNON (1984).

- **Sur-dispersion**

Dans cette partie, nous abordons le problème de l'estimation des paramètres dans une modèle régression de Poisson.

Le modèle de Poisson impose des contraintes assez fortes sur la dépendance entre la variable à expliquer et les variables explicatives. Ceci peut se traduire par l'égalité entre la moyenne et la variance. Il revient donc à dire que le nombre d'entreprises défaillantes est égal à la variabilité de ce nombre pour une période donnée. En pratique, afin de vérifier cette contrainte, nous pouvons calculer la moyenne et la variance empirique $(((\hat{m}_i, \hat{\sigma}_i), i = 1,2,...)$ des nombres de défaillances. Ceci permet de voir comment la variance évolue en fonction de la moyenne. Si la variance est plus grande que la moyenne : $\hat{\sigma}_i > \hat{m}_i$, ce phénomène est dit de surdispersion. Lorsque le modèle est adopté en présence de la surdispersion, il peut y avoir perte d'efficacité pour les études statistiques.
Il y a une méthode assez pertinente pour pallier le problème de surdispersion de la régression de Poisson. Elle consiste à introduire le modèle Binomiale Négative.

- **Une extension du modèle de Poisson : Modèle Binomiale Négative (BN)**

Le modèle binomiale négative est une généralisation du modèle de Poisson pour la sur dispersion dans le sens où, au lieu de modéliser une succession d'événements indépendants ayant une expérience constante, elle suppose qu'ils se produisent d'une manière contagieuse.
La présentation générale du modèle BN est expliquée en annexe (*cf.* **Annexe.4**).

Nous allons présenter par la suite comment les modèles décrits ci-dessus sont appliqués pour estimer le taux de défaillance sur le périmètre de Corporate. Dans un premier temps, nous nous intéressons à l'approche de Engle et Granger qui nous permet d'estimer le taux de défaillance directement. Nous nous intéressons, dans un second temps, à estimer le numérateur du taux de défaillance à l'aide du GLM et ainsi à estimer le dénominateur du taux avec ARIMA.

# PARTIE III. APPLICATIONS DES MODÈLES

# INTRODUCTION

Le cadre technique imposé par « Bâle II » a pris en considération la mise en place des exercices de *stress testing* sur les segments du portefeuille bancaire. Les exercices de *stress testing* consistent à évaluer l'impact sur les taux de défaut et par conséquent les fonds propres, d'une dégradation générale de la qualité de crédit d'un portefeuille. S'agissant du portefeuille de l'entreprise, il convient d'apprécier l'impact d'une dégradation de la situation des entreprises, cette défaillance résultant par exemple d'un contexte macroéconomique difficile.

Pour la construction d'un modèle de stress scénario, les établissements bancaires se sont vus proposés deux types de simulation de stress : un scénario de choc mono-facteur et un scénario des chocs macroéconomiques.

Le scénario de choc mono-facteur se base sur une dégradation générale des taux de défaut associés à chaque entreprise du portefeuille. Elle est assez imprécise mais présente l'avantage d'être simple à réaliser. Cependant, cette méthode ne permet pas de faire une comparaison directement entre plusieurs banques à moins que celles-ci s'obligent à dresser une table de correspondance entre notation interne et rating externe (i.e. des agences de notation). Car les banques ne calibrent à priori pas les échelles de notation interne existantes de la même manière. Nous préférons donc procéder par le second type de simulations de stress dans notre étude.

Le scénario des chocs macroéconomiques consiste à étudier la variation du taux de défaut dans les scénarios de crise. Les scénarios de crise sont précisés par les commissions bancaires en tenant compte des combinaisons de différents facteurs macroéconomiques qui, par le biais de modèle économique, sont traduits sous la forme de taux.
La présente partie est organisée de la manière suivante : nous allons présenter dans un premier chapitre les données qui seront utilisées tout au long du notre étude. Le deuxième chapitre est consacré à la mise en œuvre du modèle à correction d'erreurs sur notre portefeuille. Dans le troisième chapitre, nous allons appliquer le modèle alternatif : le modèle linéaire généralisé afin d'évaluer le taux de défaillance du portefeuille étudié dans un contexte de stress scénario.

# Chapitre 1 Génération Des Données

La probabilité de défaillance est un élément indispensable de tout modèle d'évaluation du risque de crédit. L'estimation de la relation entre les variables macroéconomiques et le taux de défaillance du portefeuille Crédit Agricole SA nécessite une longue série de données relatives à ces taux. Comme le Groupe Crédit Agricole ne dispose d'un historique en interne que de deux années, la profondeur historique des données est faible. Nous présentons ici une méthode pour construire un autre ensemble de données, en abordant tour à tour les divers problèmes qu'elle soulève.

Tout d'abord, nous recourons au taux de faillite (i.e. rapport du nombre des entreprises faisant faillite au nombre total des entreprises) pour représenter la probabilité de défaillance. Les données proviennent de l'INSEE.

Cette approche soulève deux problèmes : premièrement, les faillites ne constituent pas une approximation des incidents qui se répercutent sur les banques et le capital économique de ces dernières. La faillite représente le stade ultime de la dégradation de la situation d'une entreprise. Avant d'en arriver là, l'entreprise franchit généralement une autre étape (restructuration de la dette des entreprises en situation de détresse financière[23]), qui entraînera des pertes pour le préteur. Pour tenir compte de tous ces incidents de crédit, le service de notation interne au sein de la banque a adopté une définition large de la défaillance, qui va du retard dans les paiements à la faillite. L'utilisation du nombre des faillites entraînait une sous-estimation du nombre des incidents de crédit qui influent sur le risque de crédit des banques.

Deuxièmement, l'inclusion du nombre total des entreprises nationales reflète mal les pratiques de prêt des banques, pour qui seuls comptent les établissements emprunteurs. En ayant recours au total des établissements, nous nous retrouverions encore une fois à sous-estimer le nombre des incidents influant sur le risque de crédit des banques.

Pour combler ces lacunes, nous construisons, à partir des données relatives aux taux de faillite, de nouvelles mesures rendant mieux compte des incidents de crédit qui touchent la banque.

Les corrections effectuées s'appuient sur les renseignements suivants :

- Le périmètre étudié comprend les PME françaises réalisant un chiffre d'affaires minimum de 5 millions d'euros ayant une relation de crédit avec le Groupe Crédit Agricole.
- Les entreprises cotées ne font pas parties du périmètre étudié.
- Les institutions financières ne sont pas incluses dans le portefeuille concerné.

Les données de l'INSEE sont donc corrigées en deux étapes :

- Dans les données de l'INSEE, nous ne disposons pas des informations sur le chiffre d'affaires des établissements. Or, nous pouvons nous renseigner sur le

[23] Situation où l'émetteur offre aux porteurs d'obligations un nouveau titre ou un nouvel ensemble de titres correspondant à un engagement financier moindre, en vue d'aider l'emprunteur à éviter la défaillance.

nombre d'effectifs des entreprises du secteur donné. En faisant intervenir une matrice de correspondance entre le nombre d'employés et le chiffre d'affaires construite à partir des données internes du Crédit Agricole, nous extrayons les entreprises issues des données de l'INSEE en fonction de leurs nombre d'employés supposées avoir un chiffre d'affaires minimum de 5 millions d'euros.

- Ensuite, nous employons les données de l'INSEE pour avoir les taux de faillite.
- Nous essayons comparer les taux de faillite de l'INSEE avec les taux de défaillance du Crédit Agricole pour la période de 2008 à 2009 avec les graphiques suivants :

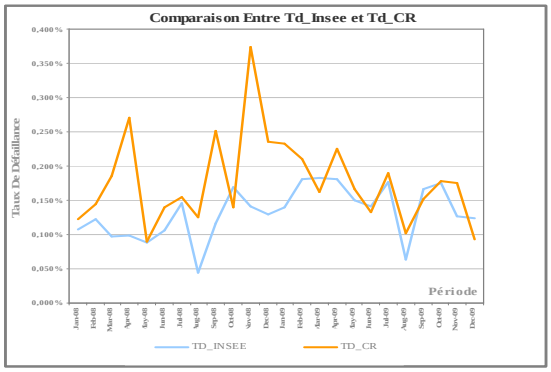



**Fig.3.1 Comparaison entre le taux de défaillance d'INSEE et le taux de défaillance du Crédit Agricole**

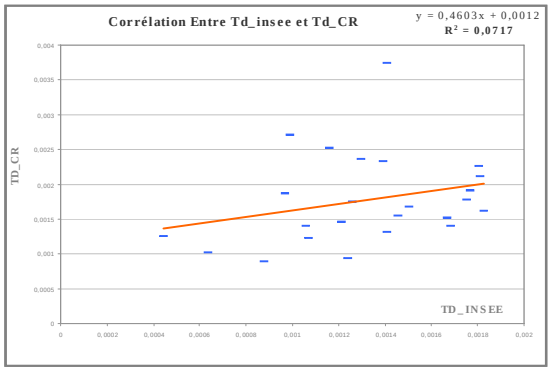



**Fig.3.2 Corrélation entre le taux de défaillance d'INSEE et le taux de défaillance du Crédit Agricole**

Le coefficient de corrélation entre deux jeux de données est de 0.07, ce qui est relativement faible.

En effet, la définition de la défaillance est différente entre ces deux entités. Pour l'INSEE, il correspond au dépôt de bilan de l'entreprise. Tandis que, pour le Crédit Agricole, il correspond au passage à la note interne : Z (i.e. en situation de quasi défaut). Ainsi, en moyenne, le décalage de temps entre l'abaissement de la note et le dépôt de bilan est de 3 mois. Ainsi, en comparant le taux de défaillance décalé de 3 mois pour celui du Crédit Agricole à celui de l'INSEE, nous obtenons le résultat suivant.

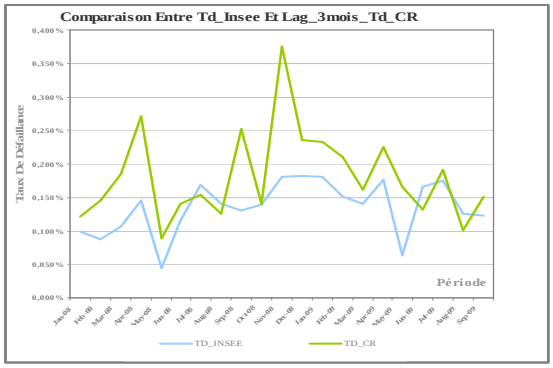



**Fig.3.3 Comparaison entre le taux de défaillance d'INSEE et le taux de défaillance du Crédit Agricole avec 3 mois de retard**

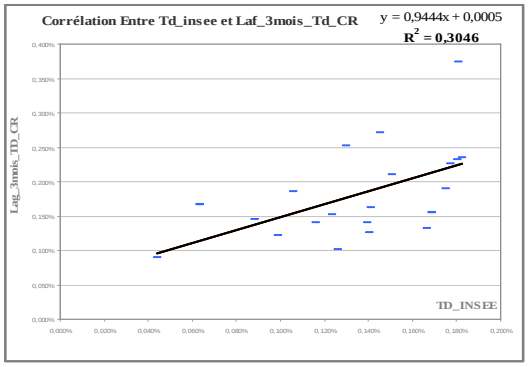



**Fig.3.4 Corrélation entre le taux de défaillance d'INSEE et le taux de défaillance du Crédit Agricole avec 3 mois de retard**

Nous remarquons que le coefficient de corrélation est de 31 %, ce qui est satisfaisant. L'adéquation est assez bonne entre les deux séries sur une période donnée. En conclusion, nous utiliserons par la suite les données de l'INSEE avec un décalage de 3 mois pour avoir un historique étendu.

L'objectif est d'analyser le comportement du taux de défaillance dans une équation, en relation notamment avec la structure des variables explicatives potentielles. Pour définir cette structure, deux facteurs ont été pris en considération. Il s'agit de l'ordre d'intégration des variables explicatives disponibles et du degré de colinéarité des variables. Une liste (*cf.***Tab.3.1**) de variables macroéconomiques est sélectionnée au préalable par les experts économiques au sein du Group Crédit Agricole :

| **Variables Macroéconomiques Sélectionnées** |
|---|
| taux d'endettement entreprises au sens large |
| taux de 3 mois |
| taux de 10 ans |
| différence entre tx 3 m et tx 10 ans |
| taux d'inflation sur 12 mois |
| taux de chômage |
| TC de PIB |
| TC d'investissement d'entreprises |
| TC d'exports |
| TC d'imports |
| TC de crédit investissement entreprises |
| TC de crédit trésorie entreprises |
| TC de CAC 40 |
| TC de production industrielle |

**Tab.3.1 La liste des variables**

# Chapitre 2 Mise En Œuvre Du Modèle À Correction D'erreurs

L'estimation du taux de défaillance par la procédure de Engle-Granger suppose de considérer successivement un certain nombre détapes. Il s'agit tout d'abord de définir l'ordre d'intégration des variables macroéconomiques potentielles (section 1). Si les variables sont stationnaires en premières différenciations, alors il est possible d'estimer le modèle en relation long terme et court terme. L'estimation de la relation de long terme est rendu possible par la méthode de moindres carrés (section 2). Cette estimation doit être validée par le test de cointégration sur les résidus. La deuxième relation consiste à étudier les impacts des mouvements macroéconomiques sur la variation du taux de défaillance pour une période plus courte (section 3). La relation à court terme permet de prédire la variation du taux de défaillance en fonction des variables macroéconomiques sélectionnées dans un cadre de stress tests (section 4).

## Section 1 Test de stationnarité

Les principes du test de racine unitaire proposé par Dickey et Fuller sont déjà expliqués dans la partie 2, nous nous concentrons maintenant sur l'application de ce test.

Rappelons que le test consiste à vérifier :

$$H_0 : \quad \rho = 1$$
$$H_1 : \quad |\rho| < 1$$

L'hypothèse nulle correspond au cas de marche aléatoire ($Y_t = Y_{t-1} + \varepsilon_t$) qui est non stationnaire stochastique et l'hypothèse alternative correspond au cas d'un modèle *AR(1)* stationnaire.
Du point de vue pratique, les tests de racine unitaire sont construits par Dickey-Fuller en réécrivant $Y_t = \rho Y_{t-1} + \varepsilon_t$ sous la forme :

$$\Delta Y_t = Y_t - Y_{t-1} = \rho^* Y_{t-1} + \varepsilon_t$$

avec :

$$\rho^* = \rho - 1$$

Par conséquent, les tests de Dickey-Fuller sont transformés de manière suivante :

$$H_0 : \rho^* = 0 \Leftrightarrow \rho = 1$$
$$H_1 : \rho^* < 0$$

Comme nous avons aussi discuté de différentes formes de la non stationnarité dans la partie précédente, nous remarquons que ce test ne répond pas aux attentes de détection du type de non stationnarité. Ainsi, les séries économiques sont caractérisées par l'autocorrélation.
Pour prendre en compte, d'une part la présence d'autocorrélation dans les séries économiques, nous pouvons généraliser la procédure $Y_t = \rho Y_{t-1} + \varepsilon_t$ à l'ordre $\rho$. Le modèle devient :

$$Y_t = \rho Y_{t-1} + ... + \phi_p Y_{p-1} + \varepsilon_t$$

D'autre part, en raison de l'hypothèse de plusieurs types de non stationnarité, nous allons réécrire la procédure précédente sous les trois régressions suivantes :

$$\Delta Y_t = \rho^* Y_{t-1} + \alpha + \beta t + \sum_{j=1}^{p} \phi_j \Delta Y_{t-j} + \varepsilon_t \qquad (1)$$

$$\Delta Y_t = \rho^* Y_{t-1} + \alpha + \sum_{j=1}^{p} \phi_j \Delta Y_{t-j} + \varepsilon_t \qquad (2)$$

$$\Delta Y_t = \rho^* Y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta Y_{t-j} + \varepsilon_t \qquad (3)$$

avec :

$p$ : est le nombre de retards
$\varepsilon_t$ : est le bruit blanc faible

Cette réécriture des régressions permet de détecter le type de régression dans lesquelles nous menons le test de racine unitaire.
La mise en œuvre des tests de racine unitaire est de procéder de manière emboitée, selon la stratégie suivante : nous testons la racine unitaire dans le modèle le plus général (i.e. modèle (1)), puis nous testons si le modèle utilisé pour mener le test était pertinent. Si tel n'est pas le cas, nous devons mener à nouveau le test de racine unitaire dans le modèle contraint. Nous ajoutons le schéma (*cf.***Fig.3.5**) afin de visualiser le principe de raisonnement :

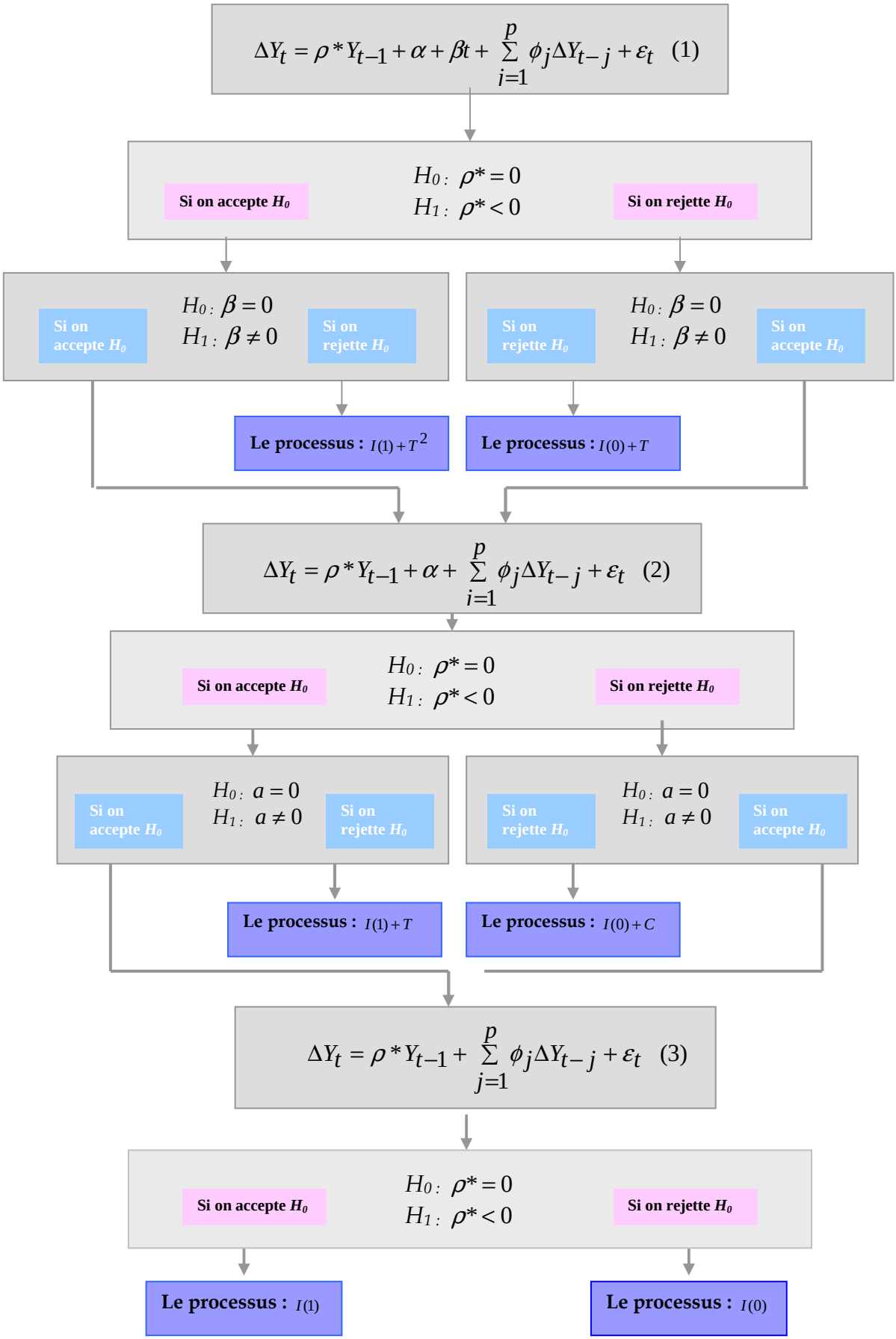



**Fig.3.5 Stratégie de test de racine unitaire**

Comme le montre dans le schéma, les réalisations de la stratégie de test de Dickey-Fuller consistent à tester l'hypothèse de racine unitaire dans les trois modèles.
- Pour le modèle (1), nous sommes dans le modèle le plus général. Nous commençons par le test de présence de la racine unitaire.
Sous l'hypothèse $H_0$ (i.e. $\rho^* = 0 \Leftrightarrow$ une racine unitaire $\Leftrightarrow$ non stationnaire), le deuxième test est mené $(H_0 : \beta = 0)$ afin de s'assurer que le test de racine unitaire a été appliqué dans un bon modèle.

- Si nous rejetons l'hypothèse $(H_0 : \beta = 0)$, nous ne pouvons constater que la différenciation de $Y_t$ (i.e. $\Delta Y_t$) est $I(0) + T$, c'est-à dire que $Y_t$ est $I(1) + T^2$. La série $Y_t$ a l'ordre d'intégration 1 et la composante déterministe est un peu plus complexe que la linéaire.
- Si nous acceptons l'hypothèse $(H_0 : \beta = 0)$, ceci est traduit par le fait que nous ne sommes pas dans un modèle approprié. Nous devons donc continuer à avancer dans la procédure jusqu'à ce que nous trouvions un bon modèle.

Sous l'hypothèse $H_1$, il peut ainsi y avoir deux possibilités : soit la série est stationnaire, soit nous ne sommes pas dans un modèle correct et nous pouvons rien conclure. Pour déterminer la raison exacte, il convient de mener le test de la significativité de la tendance $(H_0 : \beta = 0)$.

- Si nous rejetons l'hypothèse $(H_0 : \beta = 0)$, la série $Y_t$ est $I(0) + T$.
- Si nous acceptons l'hypothèse $(H_0 : \beta = 0)$, nous allons continuer le test dans le deuxième type de régression.

- Pour le modèle (2), nous sommes dans un modèle avec l'absence de la partie tendancielle. D'abord, nous testons la racine unitaire comme dans l'étape précédente.
Sous l'hypothèse $H_0$, nous allons tester la significativité de la partie constante $(H_0 : a = 0)$.

- Si nous rejetons l'hypothèse $(H_0 : a = 0)$, nous pouvons constater que $Y_t$ est $I(1) + T$, c'est-à-dire que $Y_t$ est une marche aléatoire avec dérive.
- Si nous acceptons l'hypothèse $(H_0 : a = 0)$, ceci est traduit par le fait que nous ne sommes pas dans un modèle approprié pour conclure la stationnarité. Nous devons mener le test dans le modèle (3).

Sous l'hypothèse $H_1$, alors nous devons vérifier la pertinence d'avoir testé la racine unitaire dans un modèle avec constante à l'aide de deuxième test $(H_0 : a = 0)$.

- Si nous rejetons l'hypothèse $(H_0 : a = 0)$, la série $Y_t$ est stationnaire avec dérive $I(0) + C$.
- Si nous acceptons l'hypothèse $(H_0 : a = 0)$, nous allons continuer le test dans le modèle (3).

- Pour le modèle (3), nous sommes dans le modèle le plus simple avec absence de la partie constante et tendancielle. Nous testons la racine unitaire $(H_0 : \rho^* = 0)$.

- ❖ Si nous rejetons l'hypothèse $(H_0 : \rho^* = 0)$ , la série $Y_t$ stationnaire centrée $I(0)$.
- ❖ Si nous acceptons l'hypothèse $(H_0 : \rho^* = 0)$, la série $Y_t$ a l'ordre d'intégration 1, c'est-à-dire que $Y_t \rightarrow I(1)$.

Les valeurs critiques du test de racine unitaire dépendent de la présence ou non d'une tendance. Ainsi, il faut comparer la statistique de Student non standard à la valeur critique pertinente. La raison d'utilisation de la statistique de Student non standard a été déjà expliquée dans la partie théorique de la non stationnarité. Nous notons qu'au seuil de 5%, nous devons considérer les valeurs critiques -3.45 dans le modèle (1), -2.89 dans le modèle (2) et -1.95 dans le modèle (1). Sous SAS, les tables sont intégrées dans le logiciel, et SAS donne directement les P-values calculées selon les bonnes valeurs critiques.
Nous détaillions par la suite les étapes du test de Dickey Fuller à l'aide de SAS pour la variable macroéconomique : taux de 3 mois.

**Etape 1 : Etudier la non stationnarité de la série initiale à l'aide**

- du graphique de la série pour visualiser une tendance.
La représentation graphique (*cf.***Fig.3.6**) de la série « taux 3 mois» nous permet de constater la présence d'une tendance au cours de l'année. L'allure croissante de la série montre la non stationnarité du taux de 3 mois.

- du diagramme en bâtons de la fonction d'autocorrélation empirique.
Le graphique (*cf.***Fig.3.7**) de la fonction d'autocorrélation empirique de la série « taux 3 mois» nous indique la présence de la non stationnarité. Puisque la non stationnarité se traduit souvent par une décroissance très lente de la fonction.

-du test de bruit blanc. Le résultat du test de bruit blanc indiqué dans les tests de DF nous permet de vérifier si la série est un bruit blanc. Si la P-value associée à la statistique du Khi-Deux est supérieur au seuil fixé du risque (eg 5%), nous allons accepter l'hypothèse $H_0$ (la série est un bruit blanc), ce qui met un terme à l'analyse.

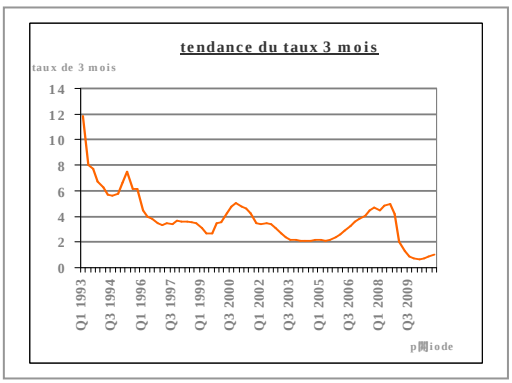



**Fig.3.6 Evolution du taux de 3 mois**

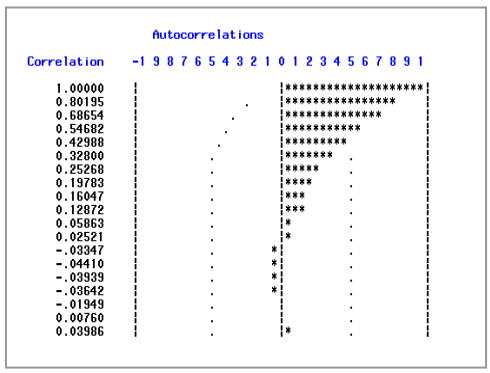



**Fig.3.7 Fonction d'autocorrélation**

Si l'un de ces critères suggère la non stationnarité, il convient de spécifier la nature *de* la non stationnarité dans les étapes suivantes.

## Etape 2 : Détecter la non stationnarité déterministe

Nous supposons que la série étudiée non stationnaire est caractérisée par une tendance déterministe. Nous régressons donc la série sur une tendance et nous récupérons le résidu de cette régression afin d'analyser sa stationnarité. Comme les méthodes sont résumées ci-dessus, nous pouvons vérifier la non stationnarité à l'aide du graphique (*cf.***Fig.3.7**) et du diagramme (*cf.***Fig.3.8**) de la fonction d'autocorrélation empirique.
Selon les graphiques, le fait d'avoir retiré une tendance semble avoir créé une autocorrélation très forte. Nous constatons donc que la série n'est pas caractérisée par la non stationnarité déterministe. Dans ce contexte, nous allons essayer de conclure la non stationnarité stochastique avec la méthode de différenciation.

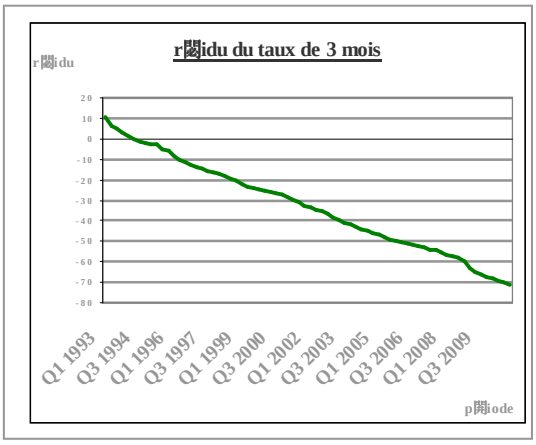



**Fig.3.8 Evolution du résidu du taux de 3 mois**

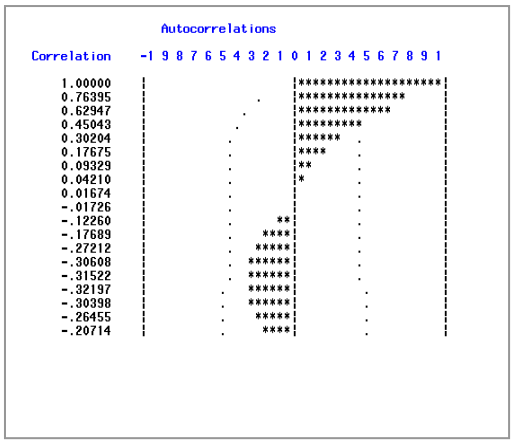



**Fig.3.9 Fonction d'autocorrélation après régressé une tendance**

## Etape 3 : Détecter la non stationnarité stochastique

D'après la nature de la non stationnarité stochastique, la série devient stationnaire après l'avoir différenciée. Nous commençons à stationnariser le taux de 3 mois par une première différenciation et nous analysons sa stationnarité à l'aide du diagramme de la fonction d'autocorrélation (*cf.***Fig.3.9**).

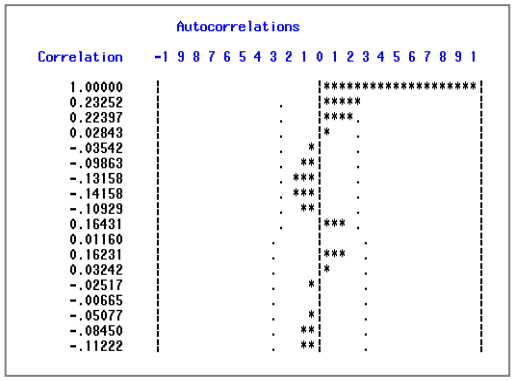



**Fig.3.10 Fonction d'autocorrélation après une différenciation**

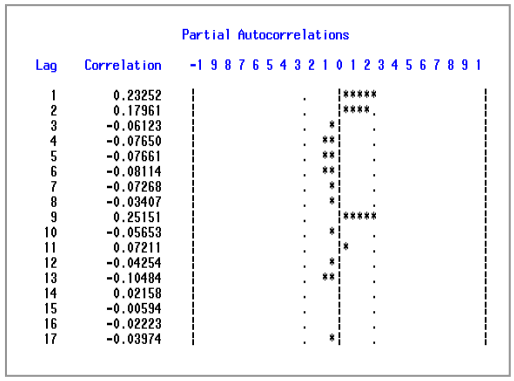



**Fig.3.11 Fonction d'autocorrélation partielle après une différenciation**

La série différenciée semble stationnaire à cause d'une décroissance rapide, mais nous ne pouvons pas conclure la stationnarité de la série avec la fonction d'autocorrélation. Nous pouvons penser que la série est un processus *I(1)*. Pour confirmer cette intuition, nous menons les tests de Dickey et Fuller avec la méthode emboitée que nous avons présentée précédemment. Il faut pour cela choisir au préalable le nombre de retards à introduire dans la régression du test de DF. Il convient de regarder le diagramme de la fonction des autocorrélations parti (*cf.***Fig.3.11**), de la série différenciée, nous pouvons penser à un *AR(1)*. Alors le test de DF est mené pour *p=1*.

Le résultat du test de Dickey-Fuller pour *p=1* est comme le suivant (*cf.***Tab.3.2**) :

| Tests De Dickey-Fuller Augumenté | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Retards | Rho | P-Value | Tau | P-Value | F | P-Value |
| Modèle (3): *Zéro Mean* | 1 | -2,29 | 0,296 | -1,75 | 0,076 | | |
| Modèle (2): *Single Mean* | 1 | -5,65 | 0,363 | -1,84 | 0,358 | 2,28 | 0,497 |
| Modèle (1): *Trend* | 1 | -9,55 | 0,439 | -2,28 | 0,439 | 2,63 | 0,656 |

**Tab.3.2 Résultat détaillé du test de Dickey-Fuller**

A priori, nous fixons un seuil de risque à 5%, ceci signifie que si la P-value[24] est inférieur à 5%, alors $H_0$ est rejetée avec un risque de 5%. Au vu de la P-value (P-value < Tau=0.439) associée au test de l'ADF dans le modèle le plus général (*Trend*), le résultat conduit à accepter l'hypothèse nulle de racine unitaire $(H_0 : \rho^* = 0)$. Nous testons alors conjointement la racine unitaire et la significativité du coefficient de la tendance $(H_0 : \beta = 0)$ par le test F3 afin de s'assurer que le test a été mené dans un bon modèle. La P-value de ce test (P-value > F=0.656) conduit à accepter l'hypothèse de nullité jointe. Il faut donc refaire le test de DF dans le modèle (2) (*Single mean*) : la P-value du test DF (P-value < Tau=0.363) conduit à accepter l'hypothèse nulle de racine unitaire. Nous testons ensuite conjointement la racine unitaire et la nullité du

[24] P-value : Dans un test statistique, la valeur P (P-value) est le plus petit niveau auquel on rejtte l'hypothèse nulle.

coefficient de la constante pour le test F2 afin de s'assurer que le modèle (2) était pertinent pour le test de racine unitaire. La P-value de ce test (P-value > F = 0.497) conduit à accepter l'hypothèse de nullité jointe. Enfin, nous allons mener le test de racine unitaire pour dans le modèle (3) (*Zero Mean*) : la P-value du test DF (P-value < Tau = 0.076) nous permet d'accepter l'hypothèse nulle de racine unitaire. Selon la stratégie du test de racine unitaire, nous constatons que le taux de 3 mois est $I(1)$, ceci signifie que la série possède une racine unitaire. En d'autres termes, le taux de 3 mois est non stationnaire de type stochastique (*TS*). Il suffit de différencier la série pour la rendre stationnaire.
De même manière, nous appliquons ce principe sur toutes les variables macro économiques concernées et sur le taux de défaillance afin de déterminer leur ordre d'intégration (*cf.***Tab.3.3**) Les résultats détaillés des tests de DF sont indiqués en annexe (*cf.* **Annexe.5**).

| Variables Macroéconomiques | L'ordre D'intégration |
|---|---|
| taux d'endettement entreprises au sens large | I(1) |
| taux de 3 mois | I(1) |
| taux de 10 ans | I(1) |
| différence entre tx 3 m et tx 10 ans | I(1) |
| taux d'inflation sur 12 mois | I(1) |
| taux de chômage | I(1) |
| TC de PIB | I(1) |
| TC d'investissement d'entreprises | I(0)+T |
| TC d'exports | I(1) |
| TC d'imports | I(1) |
| TC de crédit investissement entreprises | I(1) |
| TC de crédit trésorie entreprises | I(1) |
| TC de CAC 40 | I(0) |
| TC de production industrielle | I(0) |
| taux de défaillance INSEE grande entreprise | I(1) |

**Tab.3.3 L'ordre d'intégration des variables concernées**

Après avoir appliqué de façon systématique les tests sur l'hypothèse de racine unitaire à un ensemble de 14 séries macroéconomiques françaises, trimestrielles sur des durées de 16 ans (voir tableau ci-dessus), nous considérons que : à l'exception du taux de croissance d'investissement d'entreprises, du CAC 40 et de production industrielle, les autres variables sont non-stationnaires de type DS (*Difference stationary*). Pour le reste, il n'est pas possible de conclure sur le type de non-stationnarité.

Les conclusions des tests de racine unitaire ont deux principales conséquences : d'une part de déterminer le type de la non stationnarité, d'autre part de prendre en compte seulement des variables macroéconomiques ayant l'ordre d'intégration 1 pour alimenter le modèle de Engle et Granger.

Dans ce qui suit, nous allons mettre en œuvre le modèle à long terme en utilisant les variables décrites ci-dessus avec la méthode de la régression linéaire.

## Section 2 Estimation de relation de long terme

L'approche théorique de l'estimation de relation de long terme était présentée dans la partie 2. Dans cette section, nous nous intéressons à la mise en place du modèle avec SAS et aux tests principaux effectués.
Cette section met l'accent sur la pratique de la régression linéaire multiple. Nous avons déjà vu l'aspect théorique des principes de la régression linéaire qui n'est pas très compliquée. Cependant, l'apparente simplicité d'utilisation des programmes de calcul, qui servent principalement pour la régression, masque les nombreux problèmes pour la stabilité du modèle. Nous allons aborder ce sujet dans ce qui suit.

### 2.1 Régression linéaire multiple avec Pro REG

L'objectif du modèle de Engle et Granger est de trouver une relation entre la variable réponse et les variables explicatives dans une durée longue. Le fichier de données est composé de 15 variables, une variable de réponse (i.e. taux de défaillance trimestriel de l'INSEE pour une durée de 16 ans) et les variables explicatives de type *I(1)*.
Nous étudions le modèle linéaire multiple de la variable taux de défaillance en fonction de ses 14 répresseurs. Le modèle linéaire multiple est une extension du modèle à régression linéaire simple. Après avoir présenté les problèmes majeurs dans une régression linéaire en première et deuxième étapes, nous envisageons une procédure d'estimation des paramètres en explicitant deux méthodes proposées en troisième et quatrième étape. Les différents tests d'hypothèses concernant la pertinence du modèle sont proposés en dernière étape.

#### <u>Etape 1 Analyse des corrélations entre les variables</u>

La première étape de la construction du modèle linéaire est d'étudier des corrélations entre les variables afin d'avoir une image des liaisons entre toutes les variables.

La table (*cf.***Tab.3.4**) de matrice de corrélation relève une des difficultés de la régression linéaire : la corrélation entre les variables. Selon une des hypothèses de la méthode des moindres carrés, les variables explicatives doivent être non corrélées. Lorsque cette hypothèse n'est pas vérifiée, les estimations des MCO seraient biaisées. Ceci implique que le modèle ne sera pas robuste dans le temps. Cependant, dans le milieu économique, cette hypothèse d'indépendance apparaît très forte et difficile à satisfaire. A titre d'exemple, dans notre contexte, nous observons que le taux d'endettement des entreprises est relativement lié aux indices économiques comme : le taux de chômage avec un coefficient de corrélation de Pearson[25] de 70.3%, le taux de croissement du PIB avec un coefficient de corrélation de 66.8%...

[25] Coefficient de corrélation de Pearson : $C_{Pearson} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ avec $\sigma_{xy}$ est la covariance entre $x$ et $y$, $\sigma_x, \sigma_y$ sont l'écart-type respectivement de $x$ et $y$. Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus la corrélation entre les variables est forte.

Nous allons voir, par la suite, comment optimiser le modèle en minimisant la liaison entre les variables.

| Coefficients De Corrélation De Pearson | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **tx défaill** | **tx endettemt** | **tx 3 m** | **tx 10 ans** | **diff tx 3 et 10** | **tx inflat** | **tx chômage** | **tc PIB** | **tc Exp** | **tc Imp** | **tc crédit inv** | **tc crédit trés** |
| **tx défaill** | 100,0% | -44,6% | 36,5% | 68,3% | 37,4% | 0,4% | 74,5% | -8,2% | 17,6% | 1,4% | -84,8% | -47,6% |
| **tx endettemt** | -44,6% | 100,0% | -27,7% | -43,9% | -16,2% | 8,7% | **-70,3%** | **-66,0%** | **-69,0%** | **-65,4%** | 49,9% | -13,4% |
| **tx 3 m** | 36,5% | -27,7% | 100,0% | **77,8%** | -49,2% | 33,2% | 14,2% | 32,4% | 49,0% | 42,6% | -17,4% | 34,3% |
| **tx 10 ans** | 68,3% | -43,9% | 77,8% | 100,0% | 16,5% | 13,5% | 53,0% | 28,1% | 51,3% | 36,3% | **-60,3%** | -12,5% |
| **diff tx 3 et 10** | 37,4% | -16,2% | -49,2% | 16,5% | 100,0% | -33,3% | 51,1% | -12,0% | -5,8% | -16,5% | -57,9% | **-71,4%** |
| **tx inflat** | 0,4% | 8,7% | 33,2% | 13,5% | -33,3% | 100,0% | -41,6% | 7,9% | 14,3% | 13,6% | 12,0% | 18,4% |
| **tx chômage** | 74,5% | -70,3% | 14,2% | 53,0% | 51,1% | -41,6% | 100,0% | 20,9% | 37,6% | 24,9% | **-79,7%** | -30,8% |
| **tc PIB** | -8,2% | -66,0% | 32,4% | 28,1% | -12,0% | 7,9% | 20,9% | 100,0% | **86,7%** | **93,1%** | -8,0% | 42,6% |
| **tc Exp** | 17,6% | -69,0% | 48,7% | 51,3% | -5,8% | 14,3% | 37,6% | 86,7% | 100,0% | **92,4%** | -26,1% | 35,2% |
| **tc Imp** | 1,4% | -65,4% | 42,6% | 36,3% | -16,5% | 13,6% | 24,9% | 93,1% | 92,4% | 100,0% | -11,0% | 45,4% |
| **tc crédit inv** | -84,8% | 49,9% | -17,4% | -60,3% | -57,9% | 12,0% | -79,7% | -8,0% | -26,1% | -11,0% | 100,0% | 57,1% |
| **tc crédit trés** | -47,6% | -13,4% | 34,3% | -12,5% | -71,4% | 18,4% | -30,8% | 42,6% | 35,1% | 45,4% | 57,1% | 100,0% |

**Tab.3.4 Matrice des coefficients de corrélation de Pearson**

## Etape 2 Régression multiple

Comme nous l'avons vu la régression multiple permet de mettre en relation une variable à expliquer avec les variables explicatives à l'aide de la méthode des moindres carrés, qui consiste à chercher les coefficients minimisant les erreurs. Dans un premier temps, nous utilisons la procédure Proc REG sans option pour obtenir les estimations des coefficients du modèle (*cf.***Tab.3.5**)

:

| Le Système SAS<br>The REG Procdure | | |
|---|---|---|
| **Variable à expliquer:** | *taux de défaillance INSEE grande entreprise* | |
| **Valeur F** | 34,74 | |
| **Pr > F** | <.0001 | |
| **R-Square** | 0,8742 | |
| | | |
| **Résultats estimés des paramètres** | | |
| **Variable** | **Résltat estimé des paramètres** | **Pr > \| t \|** |
| Intercept | 0,001230 | 0,402700 |
| *taux d'endettement entreprises au sens large* | -0,000012 | 0,168100 |
| *taux de 3 mois* | -0,000021 | 0,670800 |
| *taux de 10 ans* | 0,000147 | 0,013500 |
| *différence entre tx 3 m et tx 10 ans* | 0,000000 | 0,000000 |
| *taux d'inflation sur 12 mois* | 0,000113 | 0,018300 |
| *taux de chômage* | 0,000128 | 0,075400 |
| *TC de PIB* | -0,012120 | 0,005000 |
| *TC d'exports* | -0,000597 | 0,582400 |
| *TC d'imports* | 0,001200 | 0,385500 |
| *TC de crédit investissement entreprises* | -0,003510 | 0,004900 |
| *TC de crédit trésorie entreprises* | -0,000066 | 0,900100 |

**Tab.3.5 Sortie SAS de PROC REG sans options**

Cette sortie SAS est simplifiée par rapport à son original. Nous avons juste pris en compte les tests les plus significatifs et les estimations du modèle. Dans la première partie de la table, nous avons :

- F value= 37.74 avec P-Value <0.0001, ceci nous conduit à rejeter l'hypothèse $H_{0:}$ tous les paramètres ne sont pas tous nuls.

- $R^2$ =0.8742, ceci montre qu'il y a 87,42% de variabilité de taux de défaillance qui est expliqué par le modèle.

Dans la deuxième partie, nous observons que toutes les P-Value associées aux estimateurs sont >0.05, ceci nous conduit à accepter l'hypothèse de nullité pour chacun des coefficients de la régression. Nous constatons que le modèle avec toutes les variables ne fonctionne pas. La raison est que le dysfonctionnement du modèle est surement provoqué par une forte corrélation entre les variables prédictives. Dans la sous section suivante, nous allons essayer de proposer des méthodes pour diagnostiquer le problème de la présence de corrélation entre les répresseurs.

Il y a deux problèmes principaux dans la régression linéaire. D'une part, la liaison entre les variables explicatives comme nous avons vu précédemment. D'autre part, la colinéarité entre les variables, c'est-à-dire certaines variables sont approximativement, ou exactement des combinaisons linéaires d'autres variables. Ainsi, la colinéarité des variables implique les fortes corrélations entre les variables. En effet, la colinéarité des variables régresseurs a les conséquences suivantes :

- Une instabilité des estimations des paramètres, qui en rendent l'interprétation difficile.
- Une instabilité du modèle lui-même, dont les prédictions deviennent peu fiables.
- Les variances des estimateurs sont élevées.

Il existe aussi plusieurs solutions pour résoudre le problème de la colinéarité :

- La méthode la plus générale est de réduire le nombre de prédicateurs pris en compte dans la construction du modèle. Ceci fait l'objet du paragraphe suivant.
- La Régression Ridge a été proposé pour combattre la colinéarité. Le lecteur intéressé peut se reporter à la bibliographie.
- La création de nouveaux prédicteurs décorrélés par construction peut pallier ces inconvénients

Nous nous intéressons ici à la première méthode : choix des variables explicatives les plus pertinentes. Nous allons commencer par utiliser l'option de SAS/INSIGHT.

**Etape 3 Choix des régresseurs avec SAS/INSIGHT**

Lorsque l'effet de colinéarité se présente dans les variables à expliquer, il est nécessaire de faire des choix parmi des variables proposées. Comme nous l'avons vu dans l'étape précédente, si nous mettons toutes les variables dans la construction du modèle, ceci va entrainer une instabilité des coefficients.
Nous verrons d'abord l'équation fondamentale de l'analyse de la variance. Cette analyse détermine les principes de fonctionnement de l'option SAS/INSIGHT[26].
En un point d'observation $(X_i, Y_i)$, nous allons décomposer l'écart entre $Y_i$ et la moyenne des $Y$ : $\overline{Y}$ en ajoutant la valeur estimée de $Y$ par la MCO : $\hat{Y}$

$$Y_i - \overline{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \overline{Y})$$

En sommant sur toutes les observations *i*, nous obtenons :

$$\sum_i (Y_i - \overline{Y})^2 = \sum_i (\hat{Y}_i - \overline{Y})^2 + \sum_i (Y_i - \hat{Y}_i)^2$$

[26] SAS/INSIGHT : SAS/INSIGHT est à la fois un tableur un grapheur et un analyseur.

Nous obtenons les définitions suivantes :

SS totale = SS modèle + SS erreur

avec :

SS (*sum of squares*) totale : les variations de *Y* autour de sa moyenne.
SS modèle : la somme des carrés des écarts dus au modèle augmentée.
SS erreur : la somme des carrés des écarts dus aux erreurs.

Rappelons que la définition du coefficient de détermination $R^2$ est le rapport entre la variance due au modèle et la variance totale, qui peut s'écrire :

$$R^2 = \frac{SSMod\grave{e}le}{SSTotale} = \frac{Variance\ explicative\ par\ le\ mod\grave{e}le}{Variance\ de\ la\ variable\ à\ expliquer}$$

Après avoir donné la formule générale de l'analyse de la variance, il convient de présenter les tests des rapports à SS modèle d'une variable dans la table « Type III Tests » de SAS/INSIGHT afin d'évaluer la pertinence de la variable régresseur.
Définissons un modèle restreint. Ceci est un modèle sans *r* variables, par opposition à un modèle dit complet qui est à *p* variables. Les tests de significativité du modèle sont réalisés à l'aide d'une statistique F dans le tableau « Type III Tests ». Nous allons définir F par la formule suivante :

$$F = \frac{(RRSS - URSS)/r}{URSS/(n-p-1)}$$

avec :

RRSS (*Restricted Residuel Sum of Squares*) : Somme des carrés des résidus du modèle restreint.
URSS (*Unrestricted Residual Sum of Squares*) : Somme des carrés des résidus du modèle complet.

Dans notre cas, nous essayons d'éliminer une seule variables chaque fois, pour r = 1, F peut se récrire :

$$F = \frac{(RRSS - URSS)/1}{URSS/(n-p-1)} = \frac{SS_{Mod\grave{e}le\ complet} - SS_{Mod\grave{e}le\ sans\ variable\ i}}{SS_{Mod\grave{e}le\ sans\ variable\ i}/(n-p-1)}$$

$$= \frac{SS_{apport?\ par\ i}}{SS_{Mod\grave{e}le\ sans\ variable\ i}/(n-p-1)}$$

En pratique, les sorties de SAS/INSIGHT nous permettent de tester la validité de l'apport des sommes de carrés dans le tableau « Type III Tests ». L'idée est d'éliminer progressivement la variable dont l'apport est le plus petit. Suite à l'explication de la corrélation des variables causant l'inadaptation du modèle, nous allons répartir les variables macroéconomiques en 3 groupes de 10 variables : chaque groupe possède seulement une des trois variables : taux de 3 mois, taux de 10 ans, la différence entre le taux de 3mois et le taux de 10 ans. Ensuite, nous appliquons le SAS/INSIGHT par groupe en examinant la valeur de statistique F. Dans la partie qui suit, nous allons utiliser cette application sur les données issues du group 1, c'est-à-dire le group sans la variable du taux de 3 mois.

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| Model | 10 | 1.036E-05 | 1.036E-06 | 34.74 | <.0001 |
| Error | 50 | 1.491E-06 | 2.983E-08 | | |
| C Total | 60 | 1.185E-05 | | | |

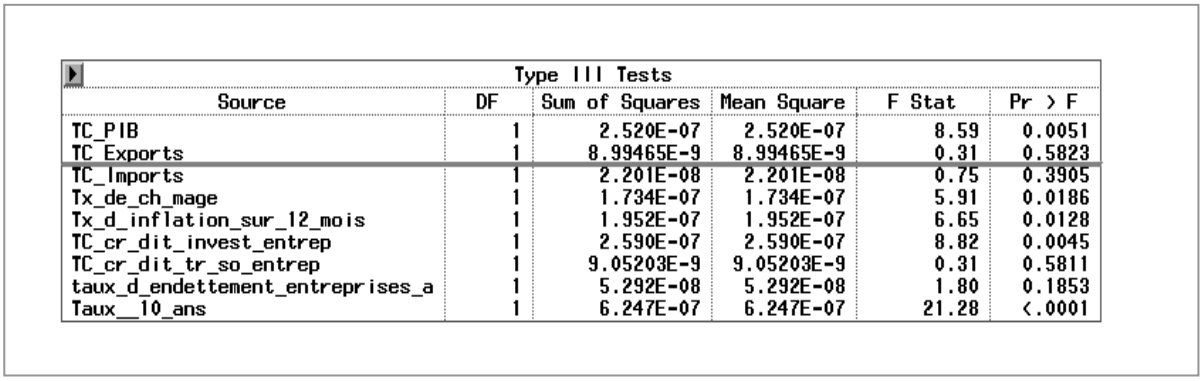
**Tab.3.6 Analyse de la variance Group 1**

Type III Tests

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| TC_PIB | 1 | 2.572E-07 | 2.572E-07 | 8.62 | 0.0050 |
| TC_Exports | 1 | 9.13694E-9 | 9.13694E-9 | 0.31 | 0.5824 |
| TC_Imports | 1 | 2.286E-08 | 2.286E-08 | 0.77 | 0.3855 |
| Tx_de_ch_mage | 1 | 9.837E-08 | 9.837E-08 | 3.30 | 0.0754 |
| Tx_d_inflation_sur_12_mois | 1 | 1.775E-07 | 1.775E-07 | 5.95 | 0.0183 |
| TC_cr_dit_invest_entrep | 1 | 2.590E-07 | 2.590E-07 | 8.68 | 0.0049 |
| TC_cr_dit_tr_so_entrep | 1 | 4.7506E-10 | 4.7506E-10 | 0.02 | 0.9001 |
| taux_d_endettement_entreprises_a | 1 | 5.834E-08 | 5.834E-08 | 1.96 | 0.1681 |
| Taux__10_ans | 1 | 6.289E-07 | 6.289E-07 | 21.08 | <.0001 |
| diff_tx_10a_3m | 1 | 5.45229E-9 | 5.45229E-9 | 0.18 | 0.6708 |

**Tab.3.7 Type III Tests du SAS/INSIGHT Etape 1.Group 1**

Nous remarquons que le modèle est globalement bon, car la P-Value associé à la valeur de la statistique est <0.0001 (*cf.***Tab.3.7 :**Analyse de la variance), ceci nous conduit à rejeter l'hypothèse de nullité pour tous les coefficients. Or, les coefficients de la plupart des régresseurs ne sont pas significatifs (*cf.***Tab.3.7 :** Type III Tests du SAS/INSIGHT Etape 1). Le modèle est donc inadapté. Par conséquent, il faut éliminer des régresseurs. Pour cela nous effectuons les tests sur les apports de sommes de carrés. Nous essayons d'éliminer la variable diff_tx_ 10a_3m dont la valeur de F est la plus petite (*F=0.18*). Ensuite, nous relançons le SAS/INSIGHT, le tableau suivant (*cf.***Tab.3.8**: Type III Tests du SAS/INSIGHT Etape 2) nous montre un nouveau modèle avec une variable de moins:

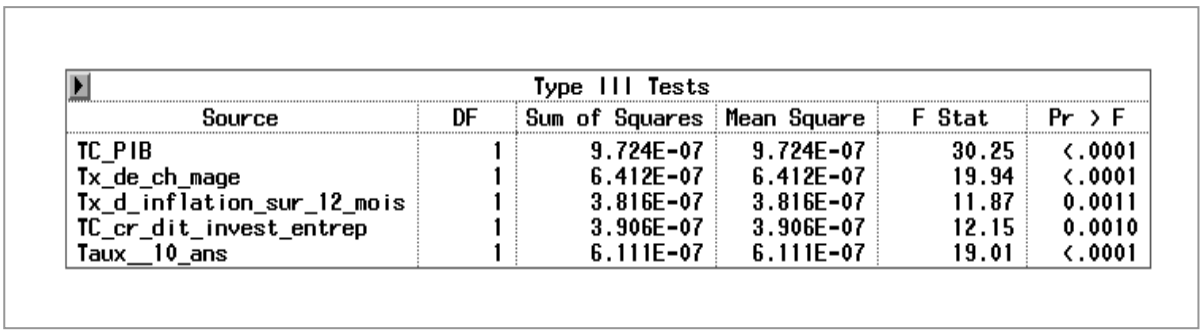


Type III Tests

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| TC_PIB | 1 | 2.520E-07 | 2.520E-07 | 8.59 | 0.0051 |
| TC_Exports | 1 | 8.99465E-9 | 8.99465E-9 | 0.31 | 0.5823 |
| TC_Imports | 1 | 2.201E-08 | 2.201E-08 | 0.75 | 0.3905 |
| Tx_de_ch_mage | 1 | 1.734E-07 | 1.734E-07 | 5.91 | 0.0186 |
| Tx_d_inflation_sur_12_mois | 1 | 1.952E-07 | 1.952E-07 | 6.65 | 0.0128 |
| TC_cr_dit_invest_entrep | 1 | 2.590E-07 | 2.590E-07 | 8.82 | 0.0045 |
| TC_cr_dit_tr_so_entrep | 1 | 9.05203E-9 | 9.05203E-9 | 0.31 | 0.5811 |
| taux_d_endettement_entreprises_a | 1 | 5.292E-08 | 5.292E-08 | 1.80 | 0.1853 |
| Taux__10_ans | 1 | 6.247E-07 | 6.247E-07 | 21.28 | <.0001 |

**Tab.3.8 Type III Tests du SAS/INSIGHT Etape 2.Group1**

Au vu de la valeur de la statistique F (*F=0.31*), nous éliminons la variable TC_Exports. De manière récurrente, nous éliminons les variables pour obtenir un modèle (Tab 9 Type III Tests du SAS/INSIGHT Etape finale) dont les apports de toutes les variables sont tous significatifs. Nous arrêtons donc le processus.



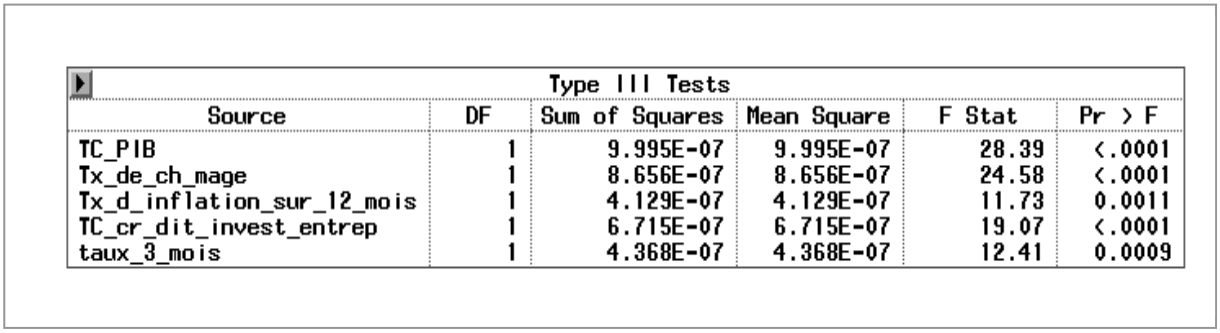
Type III Tests

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| TC_PIB | 1 | 9.724E-07 | 9.724E-07 | 30.25 | <.0001 |
| Tx_de_ch_mage | 1 | 6.412E-07 | 6.412E-07 | 19.94 | <.0001 |
| Tx_d_inflation_sur_12_mois | 1 | 3.816E-07 | 3.816E-07 | 11.87 | 0.0011 |
| TC_cr_dit_invest_entrep | 1 | 3.906E-07 | 3.906E-07 | 12.15 | 0.0010 |
| Taux__10_ans | 1 | 6.111E-07 | 6.111E-07 | 19.01 | <.0001 |

**Tab.3.9 Type III Tests du SAS/INSIGHT Etape Finale.Group1**

Il nous reste les variables comme le taux de croissance de PIB et de crédit investissement d'entreprises, le taux de chômage, d'inflation et de 10 ans. Au lieu d'avoir onze régresseurs dans le modèle de départ, nous obtenons un modèle à 5 variables explicatives.

Ainsi, nous appliquons ce processus sur les deux autres groupes (i.e. group 2 : le modèle initial sans le taux de 10 ans, group 3 : le modèle initial sans la différence entre le taux de 3 mois et celui de 10 ans). L'ensemble des résultats de type III Tests du SAS/INSIGHT est présenté dans les tableaux suivants :



Type III Tests

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| TC_PIB | 1 | 9.995E-07 | 9.995E-07 | 28.39 | <.0001 |
| Tx_de_ch_mage | 1 | 8.656E-07 | 8.656E-07 | 24.58 | <.0001 |
| Tx_d_inflation_sur_12_mois | 1 | 4.129E-07 | 4.129E-07 | 11.73 | 0.0011 |
| TC_cr_dit_invest_entrep | 1 | 6.715E-07 | 6.715E-07 | 19.07 | <.0001 |
| taux_3_mois | 1 | 4.368E-07 | 4.368E-07 | 12.41 | 0.0009 |

**Tab.3.10 Type III Tests du SAS/INSIGHT Etape Finale.Group2**

| Type III Tests | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
| TC_PIB | 1 | 9.724E-07 | 9.724E-07 | 30.25 | <.0001 |
| Tx_de_ch_mage | 1 | 6.412E-07 | 6.412E-07 | 19.94 | <.0001 |
| Tx_d_inflation_sur_12_mois | 1 | 3.816E-07 | 3.816E-07 | 11.87 | 0.0011 |
| TC_cr_dit_invest_entrep | 1 | 3.906E-07 | 3.906E-07 | 12.15 | 0.0010 |
| Taux__10_ans | 1 | 6.111E-07 | 6.111E-07 | 19.01 | <.0001 |

**Tab.3.11 Type III Tests du SAS/INSIGHT Etape Finale.Group3**

Les résultats des variables sélectionnées présentées ci-dessus montrent que les variables concernées sont les même dans le group 1 et le group 2 alors que le group 2 partage les mêmes variables à l'exception de la variable taux de 10 ans. Ceci vérifie le fait que le peu de changement de variables du modèle initial peut avoir un impact sur le choix des variables explicatives du modèle final.
Une autre méthode alternative à la régression multiple pour réduire l'effet de la colinéarité fait l'objet de l'étape suivante.

**Etape 4 Choix des régresseurs avec STEPWISE**

Dans cette partie, nous allons introduire une autre méthode de sélection des régresseurs à l'aide de l'option de PROC REG/STEPWISE. La méthode STEPWISE est une combinaison de la méthode ASCENDANTE et la méthode DESCENDANTE. Le principe de la méthode ASCENDANTE est de commencer par un modèle à une seule variable et faire entrer à chaque fois la variable qui apportera l'augmentation la plus significative de la somme des carrés du modèle. Pour la méthode DESCENANTE, nous partons avec un modèle qui a toutes les variables, nous éliminons à chaque pas la variable ayant $SS_{\text{apporté par i}}$ minimum. La procédure de la méthode STEPWISE est d'effectuer une sélection ASCENDANTE, en laissant la possibilité de faire sortir du modèle, à chaque pas, une variable non significative.

Pour ces méthodes, l'introduction d'une variable (et la sortie d'une variable) se fait en fixant un seuil de significativité.
L'ensemble des variables retenues dans notre étude correspond à des niveaux de probabilité de l'ordre de 5%. Les tableaux suivants donnent, pour les trois groupes envisagés, les variables retenues, les $R^2$, la P-Value associée, par la méthode STEPWISE.

Groupe 1

**Summary of Stepwise Selection**

| Variables Sélectionnées | R carré du modèle | Pr > F |
|---|---|---|
| *TC de crédit investissement entreprises* | 0,715 | <0,0001 |
| *taux de 10 ans* | 0,058 | 0,0003 |
| *TC de PIB* | 0,048 | 0,0003 |
| *taux d'endettement entreprises au sens large* | 0,033 | 0,0009 |

| Groupe 2<br>**Summary of Stepwise Selection** | | |
|---|---|---|
| **Variables Sélectionnées** | **R carré du modèle** | **Pr > F** |
| *TC de crédit investissement entreprises* | 0,715 | <0,0001 |
| *taux de 3 mois* | 0,742 | 0,0171 |
| *TC de PIB* | 0,783 | 0,0017 |
| *différence entre tx 3 m et tx 10 ans* | 0,822 | 0,0010 |
| *taux d'endettement entreprises au sens large* | 0,854 | 0,0009 |

| Groupe 3<br>**Summary of Stepwise Selection** | | |
|---|---|---|
| **Variables Sélectionnées** | **R carré du modèle** | **Pr > F** |
| *TC de crédit investissement entreprises* | 0,715 | <0,0001 |
| *taux de 10 ans* | 0,772 | 0,0003 |
| *TC de PIB* | 0,820 | 0,0003 |
| *taux d'endettement entreprises au sens large* | 0,853 | 0,0009 |

**Tab.3.12 Méthode STEPWISE**
**Groupe 1,2,3**

Nous remarquons que les variables significatives sont identiques dans le groupe 1 et le groupe 3. Les régresseurs sont un peu différents dans le groupe 2 en comparant aux deux autres. La comparaison des méthodes de sélection des variables nous permet d'avoir plus de choix et de prendre la meilleure décision. Au total, nous disposons de quatre modèles pour prédire le taux de défaillance dans une relation longue.

| **Modèle 1** | **Modèle 2** |
|---|---|
| *TC de PIB*<br>*TC de crédit investissement entreprises*<br>*taux de chômage*<br>*taux d'inflation sur 12 mois*<br>*taux de 10 ans* | *TC de PIB*<br>*TC de crédit investissement entreprises*<br>*taux de chômage*<br>*taux d'inflation sur 12 mois*<br>*taux de 3 mois* |
| **Modèle 3** | **Modèle 4** |
| *TC de PIB*<br>*TC de crédit investissement entreprises*<br>*taux d'endettement entreprises au sens large*<br>*taux de 10 ans* | *TC de PIB*<br>*TC de crédit investissement entreprises*<br>*taux d'endettement entreprises au sens large*<br>*taux de 3 mois*<br>*différence entre tx 3 m et tx 10 ans* |

**Tab.3.13 Modèles potentiels**

A ce stade, nous ne pouvons pas conclure que le modèle est adapté à la situation réelle. Nous allons présenter les différents éléments nécessaires à la validation d'une régression au paragraphe suivant.

**Etape 5 Validation du modèle**

Dans cette partie, nous présentons les différents éléments nécessaires à la validation d'une régression, c'est-à-dire la vérification des suppositions de base du modèle, l'étude de la robustesse au niveau des variables. L'étude de la robustesse au niveau des variables consiste à analyser ses corrélations, notamment ses colinéarités.

### 5.1 Vérification des hypothèses sut les erreurs

Lors de la partie théorique sur la régression linéaire, nous supposons que les erreurs sont des aléas, indépendants, d'espérance nulle, de variance constante, et de même loi.

- **Indépendance des erreurs**

Tout d'abord, il convient de vérifier l'indépendance sur les erreurs à partir de leurs résidus. De façon générale, nous mettons en œuvre le test de Durbin-Watson (*DW*) pour tester l'indépendance qui est intégré dans l'option de Proc REG.
Nous calculons le coefficient de Durbin-Watson à partir des résidus $e_i$ de la manière suivante :

$$DW = \frac{\sum_i (e_{i+1} - e_i)^2}{\sum_i e_i^2}$$

avec :

$$e_i = Y_i - \widehat{Y}_i$$

*DW* permet de vérifier si le résidu en *i* est non corrélé aux résidus en (*i+1*). La règle de décision consiste à dire : s'il n'y pas d'autocorrélation entre les résidus, alors DW est proche de 2.

- **Egalité des variances des erreurs**

Dans un deuxième temps, nous nous concentrons sur la vérification de l'hypothèse d'égalité des variances (homoscédasticité). Trois méthodes sont proposées pour vérifier l'égalité des variances.
Premièrement, l'étude des graphiques des résidus contre les régresseurs, nous permet de détecter qu'elle est la variable responsable de l'hétéroscédastivité. Ensuite, l'option SPEC de Proc REG permet de tester s'il y un problème d'hétéroscédastivité.
Enfin le test de Chow (1960) permet de mettre en évidence l'hétéroscédastivité (non égalité des variances) de la série, notamment pour la série chronique.

- **Normalité des erreurs**

La supposition de normalité est indispensable pour effectuer les tests sur les coefficients et les tests sur les sommes de carrés. En pratique, l'option de SAS « QQ-Plot » nous permet de vérifier la normalité des erreurs graphiquement.

## 5.2 Colinéarité des régresseurs

Dans étape 1, nous avons vu que la source principale d'instabilité d'une régression linéaire est la colinéarité. Nous allons détecter l'effet de cette colinéarité entre les variables explicatives à l'aide de SAS après une brève présentation sur l'aspect théorique.
L'idée est basée sur la méthode de la matrice carrée symétrique des régresseurs, nous récrivons le modèle multiple linéaire $Y_t = a_1 X_{1t} + a_2 X_{2t} + ...a_k X_{kt} + b + +\varepsilon_t$ sous forme matricielle :

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 X_{11} X_{12} \dots X_{1p} \\ 1 \\ \dots \\ 1 X_{n1} X_{n2} \dots X_{np} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_p \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_p \end{bmatrix}$$

L'estimation des paramètres par la méthode des moindre carrée est le vecteur :

$$\hat{b} = (X'X)^{-1} X'Y$$

La matrice variance de $\hat{b}$ est en effet :

$$V(\hat{b}) = V(X'X)^{-1} X'Y) = (X'X)^{-1} X'V(Y) X (X'X)^{-1} \text{ et } V(Y) = V(\varepsilon) = \sigma_\varepsilon^2 1_n$$

D'où $V(\hat{b}) = \sigma_\varepsilon^2 (X'X)^{-1}$

Il sera commode pour la suite de supposer que toutes les variables sont centrées et réduites. Nous notons donc $(X'X) = nR$, $R$ est une matrice identité.

Comme $V(\hat{b}) = \sigma_\varepsilon^2 (X'X)^{-1} = \sigma_\varepsilon^2 \frac{R^{-1}}{n}$, si les variables régresseurs ne sont pas corrélés entre elles, la matrice $X'X$ est une matrice identité. Sinon, la matrice aura des valeurs propres proches de 0 et son inverse aura des termes élevés. Dans ce cas les paramètres du modèle seront estimés avec imprécision et les prédictions pourront être entachées d'erreurs considérables même si les $R^2$ ont des valeurs élevées. Il est donc important de mesurer l'effet de la colinéarité entre les prédicteurs, cela s'effectue au moyen des facteurs d'inflation de la variance (VIF) et des valeurs

propres de la matrice de corrélation $V(\hat{b})$. REG possède des options TOL, VIF et COLLIN pour détecter des problèmes de colinéarité.

- **Le facteur d'inflation de la variance (VIF)**

Nous avons $V(\hat{b}) = \sigma_\varepsilon^2 (X'X)^{-1} = \sigma_\varepsilon^2 \frac{R^{-1}}{n}$, pour chaque variable explicative $X_j$, nous nommons l'élément $VIF_j$ (Variance inflation Factor) : $VIFj = V(\hat{b}_j) = \sigma_\varepsilon^2 \frac{(R^{-1})_{jj}}{n}$ et $(R^{-1})_{jj} = \frac{1}{1-R_j^2}$ où $R_j^2$ est le carré du coefficient de corrélation multiple de la régression avec constante de $X_j$ sur les *(p-1)* autres variables explicatives. S'il y a colinéarité, alors $R_j^2$ est proche de 1, donc $VIF_j$ est grand. Le seuil de limite est établit par Belsley et al. La règle de décision est la suivante : une valeur de VIF plus grand que 10 révèle un problème de colinéarité.

- **Le rôle des valeurs propres de R**

Posons $R = U\Lambda U'$ avec $\Lambda$ est la matrice diagonale des valeurs propres et $U$ la matrice des vecteurs propres de $R$. Nous pouvons en déduire que:

$$V(\hat{b}_j) = \sigma_\varepsilon^2 \frac{(R^{-1})_{jj}}{n} = \frac{\sigma_\varepsilon^2}{n} \sum_{k=1}^{p} \frac{(u_{jk})^2}{\lambda_k}$$

Pour la démonstration, le lecteur intéressé peut se reporter à la bibliographie. Nous voyons donc que $V(\hat{b}_j)$ dépend des inverses de valeurs propres de R lorsqu'il y a une forte colinéarité entre les prédicteurs les dernières valeurs propres sont proches de zéro. Les valeurs propres faibles entrainent donc de grandes variances des coefficients. D'où l'instabilité des estimations des paramètres. Dans SAS, si les proportions de variance de plusieurs variables sont plus grandes que 0.50 pour un « condition index »grand, les variables correspondantes ont un problème de colinéarité entre elles.
Après avoir expliqué les tests principaux pour les vérifications des hypothèses sur erreurs et variables, nous allons analyser les résultats obtenus sur les modèles potentiels.

### 5.3 Analyse des résultats

Nous présentons ici les résultats retenus pour le modèle 1 issu de Méthode SAS/INSIGHT du group 2. Les variables présentées ci-dessus constituent les variables macroéconomiques qui semblent les plus adaptées pour le SAS/INSIGHT.

Afin de valider et choisir le modèle, nous effectuons les tests de statistiques concernant les vérifications des hypothèses sur l'erreur et les variables du modèle.

| Variable A Expliquer | Variables Explicatives Modèle 1 | VIF |
|---|---|---|
| *taux de défaillance* | *taux de chômage* | 4,857 |
| | *taux de 10 ans* | 1,913 |
| *INSEE grande* | *TC de PIB* | 1,250 |
| | *taux d'inflation sur 12 mois* | 1,785 |
| *entreprise* | *TC de crédit investissement entreprises* | 3,849 |
| Nombre d'observations | 63 | |
| R2 | 0,863 | |
| Durbin-Watson | 1,669 | |
| Test d'hétéroscédastivité: P-Value | 0,695 | |

**Tab.3.14 Résultats des tests du modèle 1**

Dans un premier temps, nous regardons les tests associés à la vérification des hypothèses des erreurs. Le test de l'indépendance des erreurs, la valeur de DW (DW=1.669) proche de 2 nous permet de constater qu'il n'y pas d'auto-corrélation d'ordre 1 entre les erreurs. Ensuite, nous pouvons juger l'égalité des variances à l'aide du test d'hétéroscédastivité ($H_0$ : non égalité des variances des erreurs). La P-Value du test est bien supérieure. Elle nous conduit à rejeter $H_0$. Enfin, la supposition de normalité est validée par le tracé QQ-Plot ci-dessous (*cf.***Fig.3.12** : Tracé QQ-Plot de la normalité des erreurs du modèle 1) Nous observons qu'il y a un assez bon ajustement à la normale pour les erreurs du modèle 1.

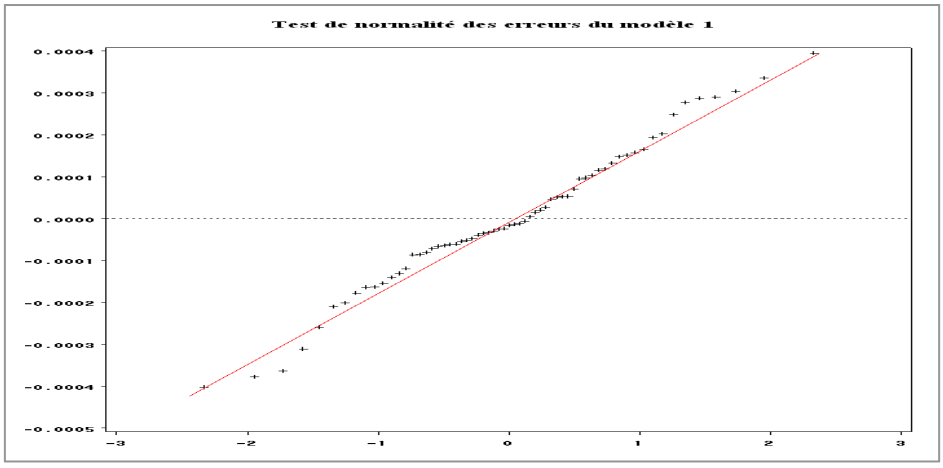



**Fig.3.12 Fonction d'autocorrélation partielle après une différenciation**

Dans un second temps, nous allons analyser les hypothèses concernant le problème de colinéarité des régresseurs. Nous restons toujours sur le modèle 1 : si aucune des valeurs VIF ne est supérieure à 10, ceci ne relève donc pas un problème de colinéarité pour toute suite. Il convient de regarder les valeurs propres de la matrice (*cf.***Tab.3.15** : Proportion De Variations du modèle 1) de corrélation pour approfondir nos analyses.

| Proportion De Variation Modèle 1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Nombre** | **Valeur Propre** | **Index de condition** | ***taux de chômage*** | ***taux de 10 ans*** | ***TC de PIB*** | ***taux d'inflation sur 12 mois*** | ***TC de crédit investissement entreprises*** |
| **1** | 2,410 | 1,000 | 0,030 | 0,053 | 0,016 | 0,007 | 0,035 |
| **2** | 1,250 | 1,389 | 0,009 | 0,059 | 0,140 | 0,262 | 0,000 |
| **3** | 0,874 | 1,661 | 0,001 | 0,026 | 0,628 | 0,103 | 0,031 |
| **4** | 0,347 | 2,637 | 0,033 | 0,860 | 0,053 | 0,227 | 0,157 |
| **5** | 0,119 | 4,503 | **0,927** | 0,001 | 0,163 | 0,402 | **0,776** |

**Tab.3.15 Proportion De variations du modèle 1**

En se référant à la règle de décision sur l'index de condition, une grande valeur met en évidence les problèmes de colinéarité entre les variables. Dans notre cas, il n'y pas de grande valeur, c'est-à-dire que toutes les valeurs sur la colonne d'index de condition sont inférieures à 30 avec l'option de SAS COLLINOINT. Si nous regardons la dernière valeur sur la 5ième ligne, nous avons deux variables taux de chômage et TC de crédit investissement entreprise qui apportent le plus de variance au modèle. Ceci traduit le fait que les deux variables sont responsables de la faiblesse de la dernière valeur propre. Comme nous l'avons montré précédemment, lorsque les proportions de variance de plusieurs variables sont plus grandes que 0.50 pour un « condition index »grand, les variables correspondantes ont un problème de colinéarité entre elles. Malgré une forte proportion de variation pour certaines variables, nous ne pouvons pas constater un problème de colinéarité dans notre cas grâce à la faible valeur d'index de condition.

Pour conclure, le modèle 1 est à la fois validé par une série des tests sur ses erreurs et par les tests sur les variables. Nous devons procéder de même pour les trois autres modèles afin de choisir les prédicteurs les mieux adaptés. Les résultats et les graphiques des tests effectués sur les autres modèles sont en annexe (*cf.* **Annexe.6**).

Lorsque les modèles potentiels sont tous validés par les tests présentés ci-dessus, le critère du choix de modèle repose plutôt sur le coefficient de détermination : $R^2$. Selon la règle générale, nous pouvons choisir le modèle qui fournir le $R^2$ maximum. Par ailleurs, il faut remarquer que $R^2$ varie en fonction du nombre de variables explicatives, c'est-à-dire qu'il augmente si nous ajoutons un prédicteur même peu corrélé avec la variable à expliquer. Nous pouvons donc choisir la taille de sous-ensemble de prédicteurs pour augmenter la valeur de $R^2$ . Pour cette raison, nous

allons avoir recours aux autres indicateurs : le coefficient $R^2$ ajusté (Adj R-sq) et le coefficient de variation.

- **Le coefficient $R^2$ ajusté**

Le coefficient $R^2$ ajusté tient compte du nombre de paramètres du modèle, il y a deux formes :

$$R^2 ajusté = 1 - \frac{n(1-R^2)}{n-p} \quad \text{s'il n'y pas d'intercept}$$

$$R^2 ajusté = 1 - \frac{(n-1)(1-R^2)}{n-p} \quad \text{s'il y a un intercept}$$

avec :

$n$ : est le nombre d'observations
$p$ :est le nombre de paramètres

Le coefficient $R^2$ ajusté tient compte du rapport entre le nombre de paramètres du modèle et le nombre d'observations.

- **Le coefficient de variation (CV)**

Le coefficient de variation peut s'écrire par la formule suivante :

$$CV = \frac{\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 / (n-p-1)}}{\bar{Y}} \times 100$$

Le CV est un indicateur sans dimension -exprimé en %- permettant de comparer l'écart moyen résiduel à la moyenne de la variable dépendante Y. Ce pourcentage est plutôt utilisé pour comparer 2 modèles portant sur le même jeu de données. Lorsqu'il est petit, les estimations des paramètres sont plus précises.

Les résultats issus des coefficients des modèles sont présentés dans le tableau **Tab.3.16**

| Modèles | $R^2$ | $R^2$ ajusté | Coefficient de Variation |
|---|---|---|---|
| **1** | **0,863** | **0,851** | **11,310** |
| **2** | 0,850 | 0,837 | 11,835 |
| **3** | 0,853 | 0,842 | 11,331 |
| **4** | 0,854 | 0,841 | 11,360 |

**Tab.3.16 Récapitulatif des coefficients des modèles**

L'analyse des résultats des coefficients nous montre que les valeurs du coefficient $R^2$ ajusté se rapprochent des valeurs du coefficient $R^2$. Par ailleurs, les valeurs du coefficient diminuent tandis que leurs valeurs relatives au coefficient $R^2$ ajusté sont plus proches de 1. Dans le cas du modèle 1, les coefficients $R^2$ et $R^2$ ajusté sont les plus élevés par rapport à ceux qui sont dans les trois autres modèles. Ainsi la valeur du CV est la plus faible également. Cette constatation nous permet de conclure que le modèle 1, associé aux prédicateurs : taux de chômage, taux de 10 ans, taux de croissement de PIB, taux d'inflations sur 12 mois et taux de crédit investissement entreprises, est le plus adapté pour estimer le taux de défaillance du périmètre.

| Résultats estimés des paramètres du modèle 1 | | |
|---|---|---|
| **Variable** | **Estimation** | **P-Value** |
| Intercept | -0,00070845 | 0,1402 |
| taux de chômage | 0,00020176 | <0,0001 |
| taux de 10 ans | 0,00011106 | <0,0001 |
| TC de PIB | -0,00848 | <0,0001 |
| taux d'inflation sur 12 mois | 0,00014345 | 0,0011 |
| TC de crédit investissement entreprises | -0,0032 | 0,001 |

**Tab.3.17 Résultats estimés des paramètres du modèle 1**

Le principe du modèle est basé sur l'hypothèse que le taux de défaillance des entreprises dépend du niveau d'activité de l'économie nationale et du niveau des taux d'intérêt du pays. Une économie plus vigoureuse, c'est-à-dire une plus forte valeur du PIB et un taux de chômage peu élevé s'accompagnerait d'un nombre faible de défaillance. En revanche, une hausse des taux d'intérêt pourrait réduire la capacité des emprunteurs à respecter leurs obligations et se traduit par une augmentation du nombre de défaillances.
Si une entreprise arrive à répercuter l'augmentation des prix à son propre prix de vente quand elle subit une hausse d'inflation, alors il y a peu d'impact négatif sur sa santé économique. Dans le cas contraire, la société pourrait subir une perte à cause de l'augmentation du prix lié à l'inflation. De plus, le taux de croissance des crédits aux investissements des entreprises s'accroît lorsque l'ensemble des facteurs économiques sont positifs.

Ces phénomènes sont confirmés par notre modèle. En effet, dans le tableau ci-dessus, les variables macroéconomiques qui ont un signe positif (voir estimation) augmente en même temps que le taux de défaillance. Inversement, les variables négatives ont un comportement contraire.
Pour conclure, notre modèle semble cohérent.

Dans ce qui suit, nous allons effectuer le test de cointégration afin de valider le modèle de long terme du taux de défaillance dans le cadre du MCE.

### 5.4 Test de cointégration

Dans cette partie, notre objectif est d'effectuer le test de cointégration sur le modèle sélectionné (modèle 1) pour valider si la corrélation entre la variable à expliquer et les variables explicatives est une vraie corrélation. En effet, nous avons vu qu'il pourrait y avoir une fausse corrélation dite « **régression fallacieuse** ». Il est donc important d'effectuer un test de cointégration.
Dans le test de cointégration à la Engle et Granger, l'ordre d'intégration et la forme de la relation de cointégration doivent être connus. Dans notre cas, l'ordre d'intégration des variables explicatives et à expliquer est 1. La relation entre elles est déterminée au préalable par la régression multiple linéaire. Ce test revient alors à vérifier l'hypothèse de racine unitaire des résidus de cointégration et est donc mené à l'aide du test de Dickey-Fuller que nous avons déjà détaillé dans la section précédente.
Pour cela, nous considérons une représentation qui autorise l'existence de la relation obtenue à l'aide de la régression multiple linéaire : le taux de défaillance dépend des variables macroéconomiques endogènes.

Formellement :

$$\begin{aligned} &Taux\ de\ d閒aillance = 0.00020176 \times taux\ de\ ch魗age \\ &+ 0.00011106 \times taux\ à\ 10\ ans \\ &- 0.00848 \times TC\ de\ PIB \\ &+ 0.00014345 \times taux\ d'\text{inf}\ lation\ sur\ 12\ mois \\ &- 0.0032 \times TC\ de\ cr閐it\ investissement\ entreprises \\ &+ \varepsilon_t \end{aligned}$$

Nous devons vérifier la stationnarité du résidu $e_t$ du modèle afin d'accepter la relation de cointégration :

$$e_t = Taux\ \ de\ \ d閒aillance\ \ r閍l - Taux\ \ de\ \ d閒aillance\ \ estim? \rightarrow I(0)$$

Le tableau suivant rend compte des résultats du test de racine unitaire appliqué sur le résidu :

| Tests De Dickey-Fuller Augumenté | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Retards | Rho | P-Value | Tau | P-Value | F | P-Value |
| Modèle (3): *Zéro Mean* | 7 | 0,20 | 0,725 | **0,45** | 0,809 | | |
| Modèle (2): *Single Mean* | 7 | 131,86 | 1,000 | **-2,95** | 0,047 | **4,58** | 0,058 |
| Modèle (1): *Trend* | 7 | 130,42 | 1,000 | **-2,92** | 0,164 | **4,26** | 0,336 |

**Tab. 3.18 Résultat détaillé du test de Dickey-Fuller**

Comme nous l'avons déjà évoqué dans la partie théorique des tests de Dickey-Fuller, nous devons utiliser le tableau des valeurs critiques de MacKinnon (*cf.* **Annexe.2**) pour juger la stationnarité du résidu selon le nombre d'observations du modèle.

Dans notre cas, le nombre d'observation est *n=63*. La table suivant donne les valeurs critiques du test en Tau, qui sont calculées à partir de MacKinnon.

| **Valeurs Critiques De MacKinnon** | |
|---|---|
| **nombre d'obervations** | 63 |
| **seuil** | 5% |
| **Moèles** | **Tau** |
| 1 | -1,946 |
| 2 | -2,908 |
| 3 | -3,485 |

**Tab.3.19 Valeurs Critiques De MacKinnon Pour n=63**

Selon le test de Dickey-Fuller sur le résidu pour le modèle (1), nous avons la valeur de Tau égale à -2.92. Avec 63 observations, la table de McKinnon nous donne une valeur, à 5% de risque, égale à -3.485 dans le modèle 3. Le fait que la valeur -2.92 est inférieure à la valeur critique de MacKinnon -3.485 nous conduit à rejeter l'hypothèse de la racine unitaire dans le modèle 3. Il y a donc une confirmation de la stationnarité pour le résidu issu de la régression.

Toutes les variables entrant dans la relation de long terme ont du être passées au crible des tests de racine unitaire. Toutes les variables utilisées se sont alors révélées non stationnaires initialement et deviennent stationnaires en différences premières. Le même test a été utilisé pour effectuer les tests de stationnarité des résidus afin de valider l'estimation de long terme.
Dans la relation à long terme, le fait que les variables soient cointégrées et non stationnaires soulève un problème d'estimation. La bonne qualité statistique du modèle ($R^2$ élevé et coefficient significatifs) est non seulement du au fait qu'il est proche de la réalité économique. Ceci est aussi du au fait que les variables sont non stationnaires puisqu'elles partagent une tendance commune. Le problème est donc, d'une part de retirer la relation commune de cointégration, d'autres part, de rechercher la liaison la plus réelle entre les variables. Ceci fait l'objet du modèle à correction d'erreurs.

Nous allons poursuivre par l'estimation du modèle à correction d'erreurs qui intègre les les variations des variables et le résidu issu de la relation à long terme.

## Section 3 Estimation du modèle à correction d'erreurs

Dans le cadre de la procédure Engle-Granger, nous avons tout d'abord estimé la relation entre le taux de défaillance et les variables macroéconomiques à long terme. Nous avons vu que les coefficients de l'équation de long terme ont les signes attendus. En ce qui concerne l'équation de court terme, la méthode des moindres carrés doit être appliquée sur le delta du taux de défaillance et sur celui des variables macroéconomiques tout en intégration le résidu estimé de la relation à long terme.

$$\Delta Yt = \gamma_1 \Delta X_{1t} + \gamma_2 \Delta X_{2t} + ... + \gamma_k \Delta X_{kt} + \delta e_{t-1} + \nu_t$$

avec :

$\Delta Yt$ : est le delta du taux de défaillance.
$\Delta X_{1t} \ldots \Delta X_{kt}$ : sont les deltas de la variable macroéconomique.
$e_{t-1}$ : est le retard du résidu issu de la relation du long terme.

Les principaux résultats, obtenus avec la méthode de Stepwise et INSIGHT figurent dans les tableaux ci-dessous :

| Variable A Expliquer | Variables Explicatives | VIF |
|---|---|---|
| Δ *taux de défaillance INSEE grande entreprise* | Δ *taux de chômage* | 1,063 |
| | Δ *TC de PIB* | 1,035 |
| | *lag résidu de la relation long terme* | 1,028 |
| Nombre d'observations | 62 | |
| R2 | 0,412 | |
| Durbin-Watson | 1,808 | |
| Test d'hétéroscédastivité: P-Value | 0,569 | |

**Tab.3.20 Résultats des tests du modèle MCE**

La statistique de Dubin-Watson révèle qu'il n'y pas d'auto-corrélation d'ordre 1 étant donné sa valeur proche de 2 (DW=1.808). Ainsi, le test d'hétéroscédastivité nous permet de rejeter l'hypothèse H0 (ie H0 : il n'y pas d'égalité de variances des erreurs) à 5%. Enfin, nous constatons qu'il y a un bon ajustement à la loi normale des erreurs du modèle MCE à l'aide de la graphique du QQ-Plot (*cf.***Fig.3.13**)

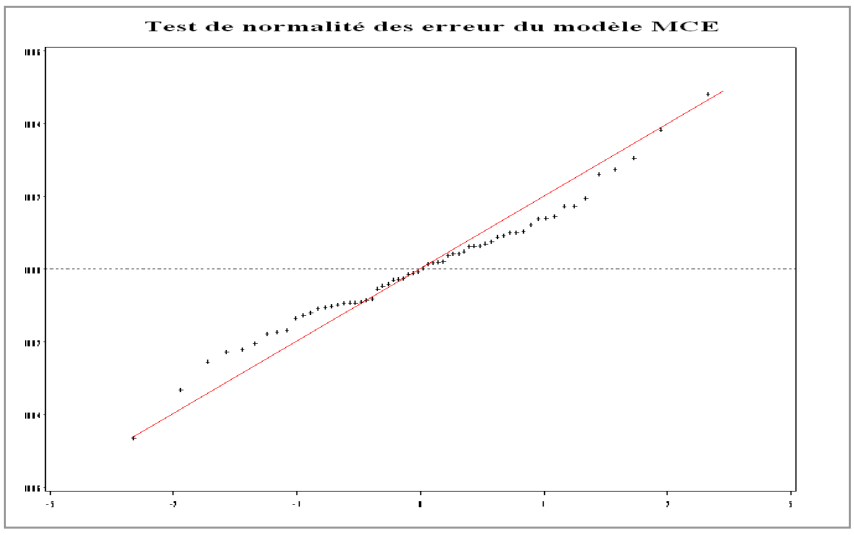



**Fig.3. 13 QQ-Plot : test de normalité du modèle MCE**

En ce qui concerne le test sur les hypothèses de colinéarité, nous observons les valeurs propres de matrice du modèle MCE (*cf.***Tab.3.21**): les valeurs VIF sont toutes inférieures à 10, ceci nous permet de constater qu'il n'y pas de problème de colinéarité entre les régresseurs du modèle.

| Proportion De Variation Modèle MCE | | | | | |
|---|---|---|---|---|---|
| **Nombre** | **Valeur Propre** | **Index de condition** | **Δ *taux de chômage*** | **Δ *TC de PIB*** | ***lag résidu de la relation long terme*** |
| **1** | 1,248 | 1,000 | 0,372 | 0,216 | 0,177 |
| **2** | 0,994 | 1,121 | 0,000007 | 0,434 | 0,542 |
| **3** | 0,758 | 1,283 | 0,628 | 0,350 | 0,281 |

**Tab.3.21 Proportion De variations du modèle MCE**

L'estimation du modèle à correction d'erreurs associée à la relation de cointégration donne le résultat suivant :

| Résultats estimés des paramètres du modèle retenu | | |
|---|---|---|
| **Variable** | **Estimation** | **P-Value** |
| Intercept | -0,00053441 | <0,0001 |
| Δ taux de chômage | 0,0001643 | 0,059 |
| Δ TC de PIB | -0,00686 | 0,0331 |
| lag résidu de la relation long terme | -0,72948 | <0,0001 |

**Tab.3.22 Résultats estimés des paramètres du modèle MCE**

En dépit d'une faible valeur du coefficient $R^2$, le modèle décrivant la dynamique de court terme du taux de défaillance peut être validé comme le modèle à correction d'erreurs, puisqu'il fait ressortir pour le lag d'erreur un coefficient négatif (-0.72948) et significatif (P-Value <0.0001) qui traduit bien la notion de force de rappel. De plus, le délai moyen de retour à l'équilibre est de 1/0.72948= 1.37 trimestres.
Du point du vue macroéconomique, les régresseurs ont des singes attendus, une augmentation du taux de chômage pourrait engendrer une hausse du taux de défaillance. En même temps, une diminution du taux de croissance du PIB alourdirait le bilan du nombre d'entreprises défaillantes.
Ainsi, le retour à l'équilibre en cas de déformation de la relation de cointégration se caractérise par un mouvement d'ensemble des variables issues du modèle de long terme (le lag résidu est issu du modèle long terme).

En conclusion, la variation du taux de défaillance est obtenu par le modèle court terme et n'est pas seulement expliqué par les variations du taux de chômage et du TC de PIB mais par l'ensemble des variables présentes dans le modèle long-terme.

Les coefficients du modèle étant estimés, la prévision peut être calculée à un horizon donné dans le cadre des stress scénarii afin d'atteindre l'objectif des études. La construction du modèle des stress tests découle directement de l'estimation du modèle à correction d'erreurs. Il a pour but de mesurer une éventuelle augmentation

du taux de défaillance lorsque l'environnement économique se dégrade, c'est-à-dire le stress scénario.

Nous allons proposer dans la partie suivante une évaluation du taux de défaillance à l'aide des estimations de l'ensemble des variables prédicteurs dans le cadre d'un stress scénario.

## Section 4 Prévision des modèles à correction d'erreurs

Une fois que les modèles ont été construits, testés et analysés, ils sont prêts à être utilisés pour réaliser des prévisions. Les prévisions consistent à faire fonctionner le modèle, dans le futur, pour un horizon donné. Dans le cadre du modèle interne des fonds propres réglementaires en cas de stress scénarii, il faut évaluer le risque de crédit pour une période de deux ans. Cette évaluation nécessite principalement des prévisions des variables macroéconomiques exogènes fournies par le service des études économiques du Crédit Agricole SA.
Afin de parvenir à une prévision convenable, plusieurs étapes sont nécessaires. Premièrement, il convient d'analyser la performance du modèle dans le passé récent, ce qui permet d'apprécier la validité du modèle estimé. Ensuite, une bonne prévision sur les variables endogènes est primordiale pour atteindre la meilleure estimation de la variable à expliquer. Dans le cas des stress tests, la prévision consiste à donner la variation du taux de défaillance en cas de choc. Enfin, l'analyse, la discussion et si c'est nécessaire la retouche des résultats prévisionnels du modèle sont recommandées.

Dans ce qui suit, nous allons présenter la première phase : mesurer la stabilité du modèle estimé sur une période postérieure.

- **Performance dans le passé récent**

La première phase consiste à s'assurer que le modèle estimé est bien capable de prévoir le taux de défaillance proche de la réalité. En d'autres termes, nous nous assurons des performances du modèle en comparant les taux de défaillance estimés et les taux de défaillance réels sur une période. Cette période est postérieure à la période d'estimation du modèle, mais pour laquelle nous disposons des observations de variables macroéconomiques. Il faut analyser comment l'équation du modèle retrace l'évolution du taux de défaillance sur cette période, ensuite tester la stabilité des paramètres estimés et donc la robustesse de l'équation. Ainsi, l'analyse graphique nous permettra de tester et de visualiser cette performance.

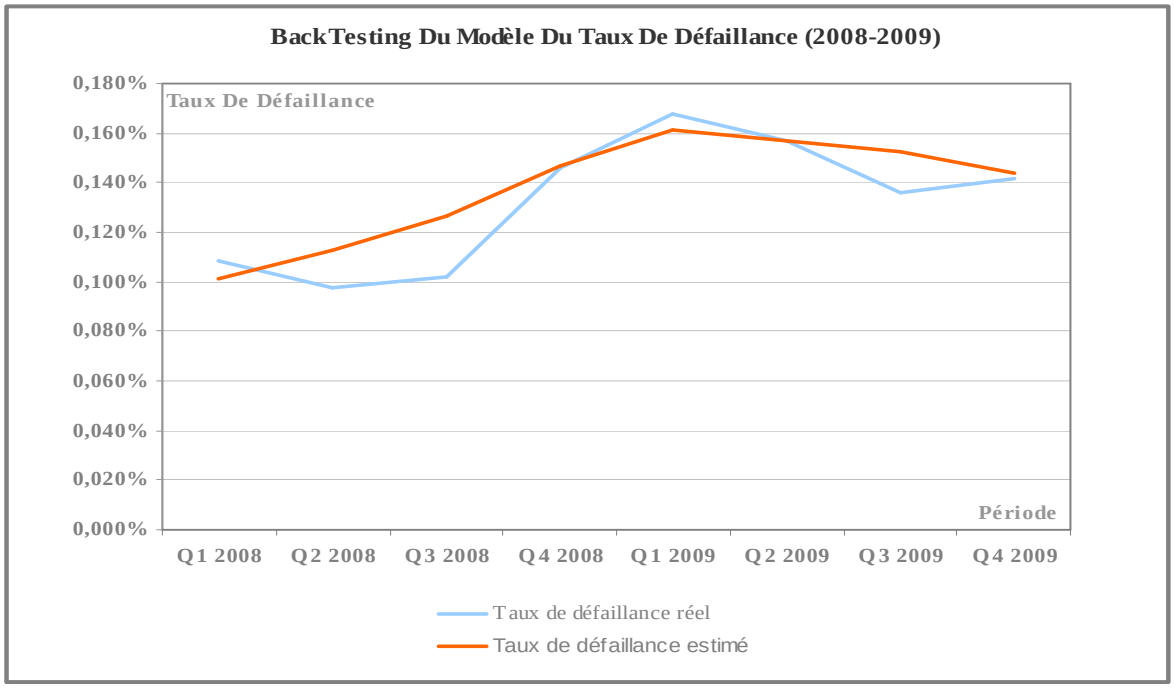



**Fig. 3.14 Comparaison entre le taux de défaillance réel et le taux de défaillance estimé**

Le graphique (*cf.***Fig.3.14**) présente l'évolution du taux de défaillance estimé à l'aide du modèle à correction d'erreurs et celle du taux de défaillance réel de la France entre 2008 et 2010. Les volumes sont exprimés en pourcentage. L'espace entre les deux courbes donne une idée de la performance du modèle utilisé. Globalement, le taux de défaillance estimé exprime bien le taux de défaillance réel. La courbe du taux de défaillance estimé se caractérise par une croissance continue durant la période. Néanmoins, il y a un grand choc sur ce taux pendant la période de crise financière des subprimes. Il est alors possible, sur cette période de deux ans, d'identifier trois phases distinctes où les deux taux se différencient. Entre début 2008 et fin 2008, le taux de défaillance croît faiblement et de manière non monotone. Durant cette période, le taux de défaillance calibré par le modèle surestime le taux de défaillance réel.

Entre fin 2008 et le 3ième trimestre 2009, la crise financière se traduit par une forte croissance en termes de nombres d'entreprises défaillantes. Le taux de défaillance estimé et réel évolue de manière presque parallèles au cours de la période de crise. Cependant si la courbe orange est au dessous de la courbe bleue, cela signe que la valeur du taux de défaillance réel est sous estimé par le modèle.
Entre le 3ième trimestre 2009 et fin 2009, la courbe du taux de défaillance se caractérise par une tendance décroissante lors du démarrage de la reprise économique. En terme général, il faut souligner que le taux de défaillance estimé a tendance à surestimer le taux de défaillance réel.
Après avoir exposé les trois phases, nous constatons que le modèle calibré ne diverge pas de la réalité. Nous allons alors effectuer les prévisions dans le cadre de stress scénarii.

- **Prévision des variables exogènes**

Les variables exogènes pour le modèle nécessitent de l'information sur les économies nationales et internationales. Ainsi les discussions avec les experts sur des différents secteurs sont importantes. Lors de la construction du modèle, il est bon d'avoir à l'esprit que nous devrons réaliser la prévision en incluant des variables accessibles, c'est-à-dire que les variables sont faciles à prédire. Le plus souvent, il convient d'utiliser des variables exogènes, car il est impensable de prévoir indépendamment l'activité économique et financière, par exemple, le taux de 10 ans et taux de chômage. C'est la raisons pour laquelle, nous avons exclut les variables comme les encours CDL (client douteux litigieux[27]) entreprises et les ratios CDL entreprises[28], qui dépendent presque du taux de défaillance et qui sont difficiles à prédire de manière indépendante. Les variables macroéconomiques issues de l'EBA (*European Banking Authrity*) sont sélectionnées et modifiées par les experts de la Direction des études économiques du Crédit Agricole S.A pour mieux s'adapter aux portefeuilles français dans le cadre des prévisions des stress scénarii.

- **Prévision du taux de défaillance des stress scénarii**

Nous avons décidé de conserver une équation du modèle MCE après avoir observé des écarts entre la réalité et les valeurs calculées par l'équation sur la période récente et vérifié la fiabilité des variables macroéconomiques. Les scénarios de stress recommandés par le régulateur sont basés sur deux scénarii. D'une part, le scénario central (*BASELINE*) consiste à prédire le taux de défaillance dans une situation économique générale. D'autre part, le scénario stressé (*ADVERSE*) a pour l'objectif de prévoir des situations économiques dégradées, par exemple, les scénarios de crise économiques.
Les deux scénarios économiques détaillés par ECO sont présentés ci-dessous (*cf.***Tab.3.23 & 3.24**)

[27] CDL : Créances Douteuses Litigieuses sont des créances pour lesquelles certaines faits permettent de douter de la solvabilité ou du paiement à l'échéance du débiteur. Il est indispensable d'opérer une distinction entre les créances douteuses et les créances litigieuses. Une créance douteuse est une créance dont le recouvrement incertain est dû aux difficultés financières du client. Une créance litigeuse est une créance faisant l'objet d'une mésentente entre l'entreprise et son client portant sur le montant de la dette.
[28] Ratio CDL Entreprises : Le nombre d'entreprises en CDL divise par le nombre d'entreprises totales.

| SCENARIO M0 | | BASELINE | | | |
|---|---|---|---|---|---|
| | | | | | |
| **Nom** | | **Indicateur** | | **2011** | **2012** |
| **France** | | | | | |
| (1) : taux de croissance annuel moyen | 1 | PIB | | **2,1** | **1,8** |
| (2) : taux de croissance sur un an en fin d'année | 1 | inflation | | **2,2** | **1,5** |
| (3) : valeur annuelle moyenne | 1 | conso ménages | | **1,7** | **1,4** |
| (4) : valeur en fin d'année | 1 | invest ménages | | **2,0** | **2,6** |
| | 1 | invest entreprises | | **5,0** | **4,0** |
| | 3 | taux de chômage | | **9,1** | **8,8** |
| | 2 | encours crédits conso | | **0,2** | **1,8** |
| | 2 | encours crédits immo | | **7,1** | **6,4** |
| | 2 | encours tréso entreprises | | **4,0** | **6,0** |
| | 2 | encours invest entreprises | | **3,2** | **4,0** |
| **UE** | | | | | |
| | 4 | taux courts | | **2,10** | **2,85** |
| | 4 | taux longs | | **4,00** | **4,25** |

**Tab.3.23 Scénario BASELINE**

| SCENARIO M1 | | ADVERSE | | | |
|---|---|---|---|---|---|
| | | | | | |
| **Nom** | | **Indicateur** | | **2011** | **2012** |
| **France** | | | | | |
| (1) : taux de croissance annuel moyen | 1 | PIB | | **1,6** | **-0,8** |
| (2) : taux de croissance sur un an en fin d'année | 1 | inflation | | **1,7** | **0,8** |
| (3) : valeur annuelle moyenne | 1 | conso ménages | | **1,3** | **0,2** |
| (4) : valeur en fin d'année | 1 | invest ménages | | **1,0** | **-1,0** |
| | 1 | invest entreprises | | **4,1** | **-2,0** |
| | 3 | taux de chômage | | **9,3** | **10,0** |
| | 2 | encours crédits conso | | **-0,4** | **-1,8** |
| | 2 | encours crédits immo | | **5,5** | **2,0** |
| | 2 | encours tréso entreprises | | **1,0** | **-2,0** |
| | 2 | encours invest entreprises | | **1,5** | **0,0** |
| **UE** | | | | | |
| | 4 | taux courts | | **1,80** | **1,30** |
| | 4 | taux longs | | **4,00** | **4,00** |

**Tab.3.24 Scénario ADVERSE**

En scénario central, l'économie mondiale reste en phase de reprise en 2011-2012. En revanche, cette reprise est sous certaines contraintes : crise de la zone euro, crise de la dette américaine. De plus, l'économie mondiale fait face à de nouveaux chocs : séisme au Japon, choc pétrolier, tension au Proche-Orient et en Afrique du nord.
En scénario dégradé, les chocs affecteraient plus fortement la situation économique que dans le scénario central pour les raisons suivants : dégradation de la confiance des entreprises et des ménages à cause de l'ensemble de l'environnement économique et financier, remontée du chômage national et international, freinage sensible de la croissance économique dans les principaux pays (Allemagne, Etats-Unis…), récession plus marquée dans certains pays (Grèce, Portugal, Italie …).

Les résultats d'estimation du taux de défaillance associé au scénario central et au scénario dégradé sont présentés dans les figures suivantes (*cf.***Fig.3.15**)

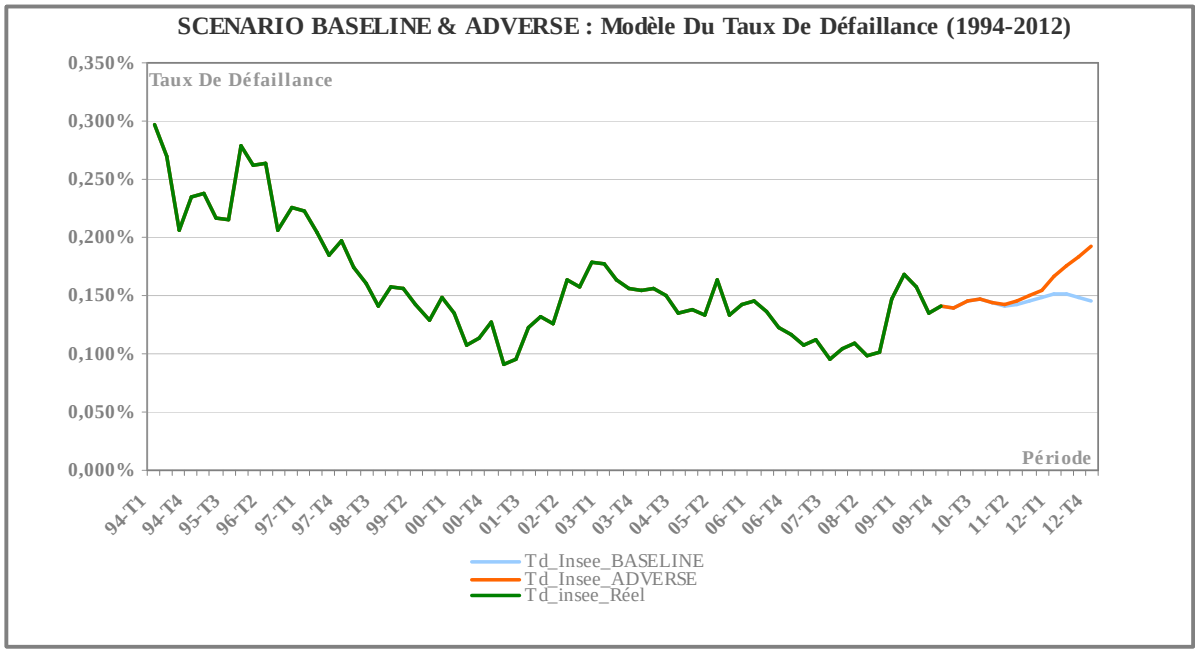



**Fig. 3.15 Evolution du taux de défaillance des entreprises françaises de 1994 à 2009 et estimation du taux de défaillance dans les deux scénarii de 2010 à 2012**

Les données issues de l'INSEE présentent l'évolution du taux de défaillance des entreprises françaises ayant le chiffre d'affaires annuel minimum de 5 millions d'Euro entre 1994 et 2009. Les volumes sont exprimés en pourcentage. Pendant cette période, le graphique permet d'identifier trois grandes phases. Entre 1994 et 2000, après la crise immobilière, nous observons une période de redémarrage économique qui se traduit par une accélération de la baisse du taux de défaillance en France. Entre 2000 et 2008, cette période est caractérisée par une phase durable de ralentissement économique français (e.g. explosion de la bulle internet). En conséquence, le taux de défaillance des entreprises en France reste globalement élevé malgré un baissement du taux de défaillance sur une courte période avant la crise des subprimes. En 2008, l'économie française subit de plein fouet les effets des différents chocs financiers. Ces chocs se caractérisent par l'augmentation du taux de défaillance en France selon les données de l'INSEE. Nous pouvons nous demander quel sera l'impact de la situation macroéconomique sur l'évolution du taux de défaillance des entreprises françaises.

A partir de l'année 2010, nous envisageons d'estimer le taux de défaillance dans deux scenarii. La courbe bleue donne une estimation du taux de défaillance, qui est censé être plus proche de la réalité. Cependant, la courbe orange nous montre le taux de défaillance dans une situation économique et financière dégradée, autrement dit, dans un stress scénario. Nous pouvons remarquer que l'espace entre les deux courbes donne une vision globale de l'évolution du taux de défaillance dans deux situations différentes. La courbe orange est au dessus de la courbe bleue, cela signifie

que la valeur du taux de défaillance du scénario Adverse est supérieure à la valeur du taux de défaillance du scénario Baseline ce qui semble logique. Les résultats des prévisions du taux de défaillance dans deux scénarii sont présentés en annexe (*cf.* **Annexe.7**).

**Conclusion**

La modélisation à l'aide du modèle à correction d'erreurs présentée dans ce mémoire a plusieurs inconvénients. Comme nous l'avons évoqué dans la partie théorique, l'approche de Engle et Granger ne permet pas de distinguer les relations de contégration entre les variables macroéconomiques. Ainsi, la qualité d'estimation du taux de défaillance est dépendante fortement de la prévision des variables macroéconomique qui sont principalement fournies par l'EBA et partiellement modifiées par les économistes au sein du Groupe Crédit Agricole. L'approche d' Engle et Granger présente deux avantages importants. Le premier est qu'elle est assez transparente et qu'elle produit des résultats qui restent applicables du point de vue opérationnel. Le deuxième avantage est qu'elle donne, jusqu'à présent, des résultats qui semblent d'être proches de la réalité et que nous pouvons les expliquer du point de vue économique.
Finalement, notre étude peut s'étendre à la deuxième modélisation qui nous permet d'avoir une comparaison des résultats entre différents modèles. Nous allons donc avoir recours au Modèle Linéaire Généralisé en tenant compte de la nature aléatoire du nombre d'entreprises défaillantes. Le chapitre présent fait l'objet de l'application de ce modèle linéaire généralisé.

# Chapitre 3 Mise En Œuvre Du Modèle Linéaire Généralisé

Nous avons vu, au chapitre précédent, que nous pouvons obtenir un modèle pour prévoir le taux de défaillance de l'entreprise dans le cadre de l'approche de Engle-Granger. Lorsque nous nous intéressons à des modèles qui tiennent compte de la nature aléatoire du taux de défaillance. Nous avons, dans ce cas, le plus souvent recours à des techniques de modélisation telles que la régression logistique ou la régression de Poisson. Ce chapitre a un double objectif. Tout d'abord, il s'agit d'exposer les résultats du modèle linéaire généralisé en expliquant les processus de mise en œuvre du modèle. D'autre part, nous allons établir une comparaison objective entre le modèle linéaire généralisé et le modèle à correction d'erreurs afin de choisir le modèle le plus adapté pour la prévision du taux de défaillance d'entreprise par secteur d'activité (Chapitre 4).
Pour l'application du modèle linéaire généralisé, la démarche est la suivante. Nous sélectionnons dans un premier temps l'ensemble des variables macroéconomiques à l'aide de la méthode des moindres carrés. Il s'agit ensuite de construire le modèle linéaire généralisé, plus précisément le modèle linéaire généralisé de Poisson avec des variables sélectionnées dans l'étape précédente.

## Section 1 Précisions des variables inputs

Contrairement au modèle à correction d'erreurs, le modèle linéaire généralisé consiste à estimer le nombre de défaillances au lieu du taux de défaillance. Au vu de la nature de la variable à expliquer, nous appliquons directement les variables macroéconomiques, c'est-à-dire que nous gardons la nature des variables originales sans les transformer sous forme de taux de croissance afin d'avoir une cohérence avec le nombre de défaillances à estimer. La liste des variables macroéconomiques proposées par l'expert économique est la suivante :

| Variable A Expliquer |
|---|
| Nombre d'entreprises en défaillance INSEE |
| **Variables Explicatives** |
| Taux d'endettement entreprises au sens large |
| Taux de 3 mois |
| Taux de 10 ans |
| Différence entre tx 3 m et tx 10 ans |
| Taux d'inflation sur 12 mois |
| Taux de chômage |
| PIB |
| Investissement d'entreprises |
| Exports |
| Imports |
| Crédit investissement entreprises |
| CAC 40 |
| Production Industrielle |

**Tab.3.25 Liste des variables à expliquer et explicatives**

## Section 2 Choix des régresseurs avec SAS/Insight et Stepwise

Nous rencontrons des situations dans lesquelles nous disposons de trop de variables explicatives, soit parce que le plan de recherche était trop vague au départ, c'est-à-dire que nous avions mesuré beaucoup de variables au cas où elles auraient eu un effet, soit parce que le nombre d'observations est très faible par rapport au nombre de variables explicatives intéressantes. Dans notre contexte, nous disposons de 13 variables macroéconomiques proposées par l'expert économique. En même temps, le nombre d'observations du taux de défaillance est relativement faible. Il est alors naturel de sélectionner un nombre réduit de variables permettant d'expliquer la variable cible. Autre raison importante de faire cette sélection des variables est d'éviter le problème de convergence des estimateurs du maximum de vraisemblance qui se base sur l'algorithme d'optimisation de type Newton-Raphson.

Pour commencer, nous faisons un bref rappel sur les méthodologies de la sélection des variables explicatives dans le cadre de la méthode des moindres carrés. Le principe de la première méthode employée est d'éliminer les régresseurs de manière progressive jusqu'à ce que les variables introduites soient toutes significatives. Nous procédons à l'aide de SAS/INSIGHT pour réaliser cette première sélection. Puis, nous pouvons utiliser la méthode alternative à l'aide de l'option SAS - STEPWISE. Cette procédure consiste à faire entrer les variables l'une après l'autre dans le modèle par sélection progressive et, à chaque étape, et vérifier si les corrélations partielles de l'ensemble des variables déjà introduites sont encore significatives. Cette approche tente donc de neutraliser les inconvénients des deux autres méthodes[29] en les appliquant alternativement au modèle en construction.

[29] Deux autres méthodes : méthode rétrograde (*backward selection*), méthode progressive (*forward selection*).

Résultat des variables sélectionnées :

Nous présentons dans le tableau **Tab.3.26** les résultats des variables sélectionnées par Stepwise :

| Summary of Stepwise Selection | | |
|---|---|---|
| **Variables Sélectionnées** | **R carré du modèle** | **Pr > F** |
| *CAC 40* | 0,792 | <0,0001 |
| *Taux de 10 ans* | 0,089 | <0,0001 |

**Tab.3.26 Liste des variables explicatives par la méthode de Stepwise**

Comme le montre le tableau **Tab.3.27**, Option INSIGHT sélectionne les régresseurs les plus pertinents pour expliquer le nombre de défaillance dans le sens de MCO.

| Type Ⅲ Tests | | | | | |
|---|---|---|---|---|---|
| **Source** | **DF** | **Sum of squares** | **Mean Square** | **F Stat** | **Pr > F** |
| Investissement d'entreprise | 1 | 74525,91 | 74525,91 | 24,86 | <0,0001 |
| CAC 40 | 1 | 213411,92 | 213411,95 | 71,19 | <0,0001 |

**Tab.3.27 Liste des variables explicatives par la méthode d'INSIGHT**

Une fois que nous avons les variables explicatives appropriées fournies par les deux modèles, nous pouvons avoir recours aux tests statistiques afin de valider ces variables explicatives. L'ensemble des résultats issus des tests statistiques est présenté en annexe (*cf.* **Annexe.8**).
Les hypothèses de la méthode des moindres carrés sont principalement vérifiées par les tests statistiques. Nous constatons alors que les variables sont significatives. Par la suite, nous allons détailler les procédures de la prévision du nombre de défaillance en fonction des variables sélectionnées dans un cadre du modèle GLM.

## Section 3 Application du GLM

L'étape suivante consiste à modéliser la trajectoire des nombres d'entreprises défaillantes selon sa nature aléatoire à l'aide du modèle linéaire généralisé, plus précisément la régression de Poisson ou la régression binomiale négative.

Le schéma suivant illustre les étapes principales pour la construction pratique d'un modèle linéaire généralisé

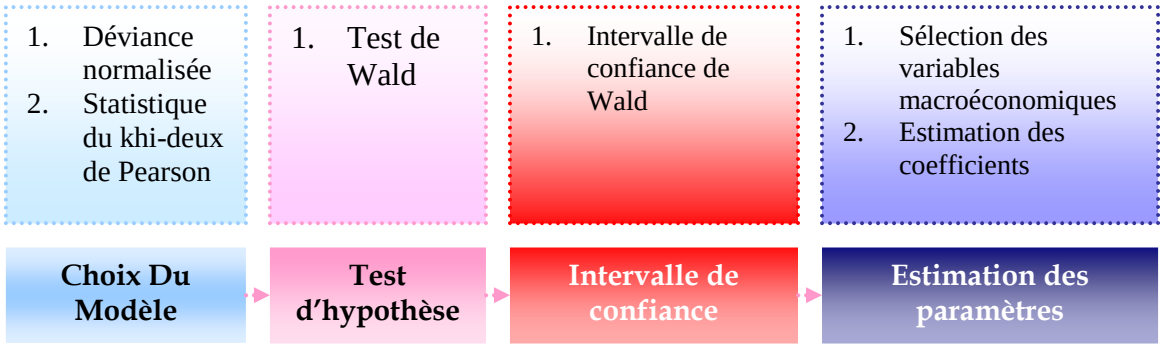



**Fig. 3.16 Principe de construction du modèle linéaire généralisé**

L'estimation des paramètres dans un modèle linéaire généralisé nécessite l'estimation de loi du modèle à l'aide des tests statistiques et la validation de l'estimateur selon l'intervalle de confiance. La méthodologie retenue et les résultats de l'application du GLM font l'objet du paragraphe suivant.

### 3.1 Choix du modèle

Le plus souvent le choix de la loi de probabilité de la fonction de réponse découle naturellement de la nature de la loi du problème étudié. Comme précisé dans le chapitre précédent, la défaillance d'une entreprise est un événement rare qui peut être éventuellement modélisée par la loi de Poisson. Nous pouvons alors choisir la fonction de lien logit pour associer, à la loi de probabilité Poisson, la fonction de réponse étudiée. La modélisation de la fonction de réponse est une combinaison linéaire des covariables et de leurs paramètres $\beta$ . Dans notre contexte, la formule peut s'écrire de la manière suivante :

$$\log(\mu_i) = \beta_0 + \beta_1 MarcoVar_{i,i} + \beta_2 MacroVar_{2,i} + ... + \beta_k MacroVar_{k,i}$$

avec :

$\mu_i : E(Y_i)$, $Y_i$ est le nombre d'entreprises tombant en défaillance par trimestre.
*MarcoVar* : CAC 40, Taux de 10 ans ou CAC 40, Investissement d'entreprise.

La modélisation du GLM est faite par SAS- GENMOD. Les tableaux (*cf.***Tab.3.28 & Tab.3.29**) donnent un récapitulatif des statistiques du SAS-GENMOD sur les deux modèles proposés.

| SAS GENMOD Procedure | | | |
|---|---|---|---|
| **Distribution** | Poisson | | |
| **Link Function** | Log | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | |
| **Macro Variable** | *Taux de 10 ans* | *CAC 40* | |
| **Critère d'évaluation de l'adquation** | | | |
| **Critère** | **DDL** | **Valeur** | **Valeur/DDL** |
| **Deviance** | 61 | 349,20 | **5,72** |
| **Scaled Deviance** | 61 | 58,73 | 0,96 |
| **Pearson Chi-Square** | 61 | 362,68 | 5,95 |
| **Log Likelihood** | | 21090,65 | |
| **AIC** | | 851,81 | |
| **BIC** | | 858,29 | |

**Tab.3.28 Résultat détaillé du test de SAS-GENMOD pour la loi de Poisson avec les variables macroéconomiques Taux de 10 ans et CAC 10**

| SAS GENMOD Procedure | | | |
|---|---|---|---|
| **Distribution** | Poisson | | |
| **Link Function** | Log | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | |
| **Macro Variable** | *Investissement d'entreprise* | *CAC 40* | |
| **Critère d'évaluation de l'adquation** | | | |
| **Critère** | **DDL** | **Valeur** | **Valeur/DDL** |
| **Deviance** | 61 | 307,32 | **5,04** |
| **Scaled Deviance** | 61 | 59,38 | 0,97 |
| **Pearson Chi-Square** | 61 | 315,72 | 5,18 |
| **Log Likelihood** | | 24232,21 | |
| **AIC** | | 809,00 | |
| **BIC** | | 816,41 | |

**Tab.3.29 Résultat détaillé du test de SAS-GENMOD pour la loi de Poisson avec les variables macroéconomiques Investissement d'entreprise et CAC 40**

avec :

*DDL* : Nombre d'observations – Nombre de paramètres.

*Déviance* : $\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2$ .

*Scaled Deviance* : $2In\frac{l(y,y)}{l(\hat{\mu},y)}$ .

*Pearson Chi-Square* : $\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$ .

Nous avons évoqué, dans la partie théorique, qu'une petite valeur de la *Scaled déviance* peut expliquer un bon ajustement du modèle. Nous nous intéressons davantage à la quantité « Valeur /DDL ». La quantité « Valeur/DDL » proche de 1 nous permet de juger l'adéquation du modèle. Au vu de la valeur de « Deviance $_{Valeur/DDL}$=5.72 », nous constatons qu'il y a un problème de la sur-dispersion du modèle de Poisson. Il y a sur-dispersion lorsque la déviance normalisée ou le khi-deux de Pearson normalisé sont nettement supérieurs à 1. La présence de sur-dispersion est du à la caractéristique de la loi de probabilité de Poisson qui exige une égalité entre la variance et la moyenne des observations.

Il convient de recourir à la loi de probabilité de Binomiale Négative afin de remédier au problème de la sur-dispersion du modèle. Par la suite, nous allons montrer et analyser le résultat des estimations du modèle Binomiale Négative.

| **SAS GENMOD Procedure** | | | |
|---|---|---|---|
| **Distribution** | Negative Binomial | | |
| **Link Function** | Log | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | |
| **Macro Variable** | *Taux de 10 ans* | *CAC 40* | |
| **Critère d'évaluation de l'adquation** | | | |
| **Critère** | **DDL** | **Valeur** | **Valeur/DDL** |
| **Deviance** | 61 | 63,20 | **1,04** |
| **Scaled Deviance** | 61 | 58,10 | 0,95 |
| **Pearson Chi-Square** | 61 | 66,36 | 1,09 |
| **Log Likelihood** | 115356,14 | 115356,14 | |
| **AIC** | 670,21 | 670,21 | |
| **BIC** | 678,84 | 678,84 | |

**Tab.3.30 Résultat détaillé du test de SAS-GENMOD pour la loi Binomiale Négative avec les variables macroéconomiques Taux de 10 ans et CAC 40**

| **SAS GENMOD Procedure** | | | |
|---|---|---|---|
| **Distribution** | Negative Binomial | | |
| **Link Function** | Log | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | |
| **Macro Variable** | *Investissement d'entreprise* | *CAC 40* | |
| **Critère d'évaluation de l'adquation** | | | |
| **Critère** | **DDL** | **Valeur** | **Valeur/DDL** |
| **Deviance** | 61 | 63,05 | **1,03** |
| **Scaled Deviance** | 61 | 58,76 | 0,96 |
| **Pearson Chi-Square** | 61 | 65,46 | 1,07 |
| **Log Likelihood** | | 116941,64 | |
| **AIC** | | 662,67 | |
| **BIC** | | 671,31 | |

**Tab.3.31 Résultat détaillé du test de SAS-GENMOD pour la loi Binomiale Négative avec les variables macroéconomiques Investissement d'entreprise et CAC 40**

Les résultats fournis par SAS-GENMOD (*cf.***Tab.3.30 & Tab.3.31**) (la valeur de « Valeur / DDL » est proche de 1) nous permettent de constater que l'approche proposée qui consiste à remplacer la loi de Poisson par la loi Binomiale Négative semble conduire à des résultats tout à fait acceptables sur le plan pratique. Cependant, nous devons remarquer que la fonction de lien associée à la loi de Probabilité Binomiale Négative est log au lieu de logit. Puisque la fonction de lien logit ne nous permet pas d'avoir les résultats satisfaisants, nous avons sélectionné la fonction de lien log pour la loi BN dans notre étude. Les résultats détaillés du test de la loi de Probabilité BN associée à la fonction de lien logit sont montrés en annexe (*cf.* **Annexe.9**).

Cela nous ramène à prochaine étape de la modélisation par GLM : Test de Wald.

### 3.2 Test d'hypothèse concernant les paramètres

Après avoir choisi la fonction de lien et la loi paramétrique, nous nous concentrons sur la vérification des hypothèses de départ à l'aide du Test de Wald.
Le test de Wald est le test de nullité de paramètres sous l'hypothèse $H_0$ : $L'\beta = 0$
Sous l'hypothèse $H_0$, la statistique $S = (L'b)'(L'J^{-1}L)^{-1}(Lb)$ suit une loi du $\chi^2$ à r degrés, où r = rang de L

avec :

$L : \prod_i f(y_i, \theta_i(\beta)), i = 1,2,...,n$

$b$ :fonction spécifiée de la densité $f(y_i, \theta_i, \phi, \omega_i) = \exp(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi, \omega_i))$

Nous présentons, les tableaux (*cf.***Tab.3.32 & Tab.3.33**), les résultats du test de Wald correspondent aux modèles avec les variables macroéconomiques : CAC 40, Taux à 10 ans et CAC 40, Investissement d'entreprise.

| **SAS GENMOD Procedure** | | | | |
|---|---|---|---|---|
| **Distribution** | Negative Binomial | | | |
| **Link Function** | Log | | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | | |
| **Macro Variable** | *Taux de 10 ans CAC 40* | | | |
| **Critère d'évaluation de l'adquation** | | | | |
| **Paramètre** | **DDL** | **Valeur estimée** | **Khi-2 de Wald** | **Pr > Khi-2** |
| **Intercept** | 1 | 6,40 | 3499,10 | < 0,0001 |
| **CAC 40** | 1 | -0,00018 | 183,71 | < 0,0001 |
| **Taux de 10 ans** | 1 | 0,0588 | 17,18 | < 0,0001 |

**Tab.3.32 Résultat détaillé du test de SAS-GENMOD pour la loi Binomiale Négative avec les variables macroéconomiques Investissement d'entreprise et CAC 40**

| SAS GENMOD Procedure | | | | |
|---|---|---|---|---|
| **Distribution** | Negative Binomial | | | |
| **Link Function** | Log | | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | | |
| **Macro Variable** | *Investissement d'entreprise CAC 40* | | | |
| **Critère d'évaluation de l'adquation** | | | | |
| **Paramètre** | **DDL** | **Valeur estimée** | **Khi-2 de Wald** | **Pr > Khi-2** |
| **Intercept** | 1 | 7,14 | 6790,95 | < 0,0001 |
| **CAC 40** | 1 | -0,000158 | 116,01 | < 0,0001 |
| **Investissement d'entreprise** | 1 | 0,000016 | 27,19 | < 0,0001 |

**Tab.3.33 Résultat détaillé du test de SAS-GENMOD pour la loi Binomiale Négative avec les variables macroéconomiques Investissement d'entreprise et CAC 40**

A l'issue des tests de Wald (Pr > Khi-2 < 0.0001), nous pouvons constater que l'hypothèse de nullité des paramètres est rejetée avec une probabilité de 0,5%. Les deux modèles potentiels associés à la loi de probabilité Binomiale Négative vérifient donc l'hypothèse concernant les paramètres. Le calcul de l'intervalle de confiance fait l'objet de la section suivante.

### 3.3 Intervalle de confiance

L'estimation des intervalles de confiance pour le paramètre $\beta$ est conduite par la procédure suivante :

$$IC = [b_j - z_{1-\frac{\alpha}{2}} S_j ; b_j + z_{1-\frac{\alpha}{2}} S_j]$$

avec :

$Z_{1-\frac{a}{2}}$ : fractile d'ordre $1-\frac{a}{2}$ d'une loi normale réduite

$S_j$ :j-ième terme de la diagonale de la matrice variance –covariance de b

Etant donné que b est approximativement distribué selon une loi normale $N(\beta, b)$, nous pouvons construire cet intervalle de confiance.
L'ensemble des résultats relatifs à l'intervalle de confiance de Wald est présenté dans les tableaux (*cf.* **Tab. 3.34 & Tab3.35**)

| SAS GENMOD Procedure | | | | |
|---|---|---|---|---|
| **Distribution** | Negative Binomial | | | |
| **Link Function** | Log | | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | | |
| **Macro Variable** | *Taux de 10 ans CAC 40* | | | |
| **Critère d'évaluation de l'adquation** | | | | |
| **Paramètre** | **DDL** | **Valeur estimée** | **95% Lower Confidence Limit** | **95% Upper Confidence Limit** |
| **Intercept** | 1 | 6,40 | 6,1293 | 6,5493 |
| **CAC 40** | 1 | -0,00018 | -0,0002 | -0,0001 |
| **Taux de 10 ans** | 1 | 0,0588 | 0,0310 | 0,0865 |

**Tab.3.34 Résultat détaillé de l'intervalle de confiance de Wald avec les variables macroéconomiques CAC 40 et Taux à 10 ans**

| SAS GENMOD Procedure | | | | |
|---|---|---|---|---|
| **Distribution** | Negative Binomial | | | |
| **Link Function** | Log | | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | | |
| **Macro Variable** | *Investissement d'entreprise CAC 40* | | | |
| **Critère d'évaluation de l'adquation** | | | | |
| **Paramètre** | **DDL** | **Valeur estimée** | **95% Lower Confidence Limit** | **95% Upper Confidence Limit** |
| **Intercept** | 1 | 7,14 | 6,9700 | 7,3100 |
| **CAC 40** | 1 | -0,000158 | -0,0002 | -0,0001 |
| **Investissement d'entreprise** | 1 | 0,000016 | 0,00001 | 0,00002 |

**Tab.3.35Résultat détaillé de l'intervalle de confiance de Wald avec les variables macroéconomiques CAC 40 et Investissement d'entreprise**

Dans le cas où les estimateurs des deux modèles potentiels sont compris dans l'intervalle de confiance de Wald, les estimations sont jugées adéquates avec le modèle choisi.

Afin d'estimer la distribution du nombre d'entreprises défaillantes de France, nous présenterons la démarche suivante.

### 3.4 **Estimation des paramètres**

Nous avons proposé deux modèles potentiels correspondants à des variables macroéconomiques différentes.
Dans le cas où plusieurs modèles concurrents sont en compétition, le but est de choisir le plus adéquat. Il est possible d'utiliser le test de déviance entre modèles ou des critères de choix de modèles tels l'AIC (Akaike, 1974) ou le BIC (Schwarz, 1978).
Dans un premier temps, nous allons comparer la déviance des modèles potentiels. Si la différence de déviance est grande entre deux modèles potentiels, celui qui possède

une déviance plus proche de 1 est en général plus acceptable. En revanche si l'écart est faible, nous pouvons comparer les critères AIC et BIC pour sélectionner le modèle le plus adapté.
Rappelons par définition que le critère de l'AIC ( *Akaike Informative Criterion*) pour un modèle à $p$ paramètres est défini par

$$AIC = -2L + 2p$$

avec :

$L$ : log-vraisemblance.
$p$ : paramètre du modèle.

Le mécanisme est simple : plus la vraisemblance est grande, plus la log-vraisemblance $L$ est importante et meilleur est le modèle. Cependant si nous mettons le nombre maximum de paramètres (ce qui est le modèle saturé) alors $L$ sera maximal. Il suffit donc de rajouter des paramètres pour la faire augmenter. Pour obtenir un modèle de taille raisonnable, il convient de la pénaliser par une fonction du nombre de paramètres, ici $2p$.

Un autre critère de choix de modèle le BIC (*Bayesian Informative Criterion*) pour un modèle à $p$ paramètres estimé sur n observations est défini par

$$BIC = -2L + 2p\log(n)$$

La règle de l'utilisation de ces critères est la suivante : le critère de choix de modèle est de sélectionner le modèle qui présente les plus faibles valeurs de déviance, d'AIC et de BIC.

| SAS GENMOD Procedure | | | | |
|---|---|---|---|---|
| **Distribution** | | Negative Binomial | | |
| **Link Function** | | Log | | |
| **Depend Variable** | | Nombre d'entreprises défaillantes | | |
| **Critère de choix** | | | | |
| **Modèle** | **Variable Explicative** | **Déviance** | **AIC** | **BIC** |
| **1** | ***CAC 40*** | 1,04 | 670,21 | 678,84 |
| | ***Taux de 10 ans*** | | | |
| **2** | ***CAC 40*** | **1,03** | **662,67** | **671,31** |
| | ***Investissement d'ent*** | | | |

**Tab.3.36 Résultat détaillé de critère de choix pour les modèles potentiels**

Selon le tableau (*cf.***Tab.3.36**), le modèle 2 nous semble le plus adéquat par rapport au modèle 1. Nous justifions notre choix par la plus faible valeur de déviance, de l'AIC et du BIC dans le second modèle.

| Résultats estimés des paramètres du modèle 2 par GLM | | |
|---|---|---|
| **Variable** | **Estimation** | **P-Value** |
| Intercept | 7,1373 | < 0,0001 |
| *CAC 40* | -0,000158 | < 0,0001 |
| *Investisement d'entreprise* | 0,000016 | < 0,0001 |

**Tab.3.37 Résultats estimés des paramètres du modèle GLM**

Une fois que le choix du modèle est effectué, nous obtenons la formule avec les paramètres estimés correspondants au modèle 2 (*cf.***Tab.3.37**) :

$$\log(\mu_i) = 7.14 - 0.000158 \times CAC40_i + 0.000016 \times Investissement\ \ d'entreprise_i$$

Le nombre d'entreprises défaillantes estimés pour le trimestre *i* est donné alors par :

$$\mu_i = \exp(\log(\mu_i))$$
$$= \exp(7.14 - 0.000158 \times CAC40_i + 0.000016 \times Investissement\ \ d'entreprise_i)$$

L'idée de départ de notre étude est de modéliser la défaillance des entreprises par les variables macroéconomiques. Les variables retenues par le modèle GLM sont le CAC40 et l'investissement d'entreprise. Nous notons que le signe des variables suit la réalité économique. La performance de CAC 40 peut être interprétée comme la bonne santé des entreprises françaises. En revanche, l'augmentation des investissements des entreprises peut provoquer un accroissement du nombre de défaillance pour les corporates.
Afin de parvenir à prédire le taux de défaillance sur le périmètre corporate, l'estimation du dénominateur du taux : nombre d'entreprises totales en France est nécessaire.

Nous tentons de modéliser l'évolution du nombre des entreprises totales françaises par un modèle ARMA

## Section 4 Application de l'ARMA

Dans cette section, nous allons présenter la méthodologie de la procédure ARMA de manière brève et l'application du modèle afin de parvenir à la prévision de l'évolution de la variable recherchée.
La modélisation du nombre d'entreprises françaises totales par trimestre est basée sur l'hypothèse que ce dernier est une série chronologique.

### 4.1 **Principe du modèle ARMA**

Rappelons qu'une série chronologique $(Y_t)_t$ suit un processus ARMA$(p,q)$ si $Y$ est stationnaire et $\varphi(B)Y_t = \theta(B)\varepsilon_t$

avec :

$\varphi(z) = 1 + \varphi_1(z) + ... + \varphi_p(z)^p$

$\theta(z) = 1 + \theta_1(z) + ... + \theta_q(z)^q$

$(\varepsilon_t)_t$ : bruit blanc de variance $\sigma_\varepsilon^2$ pour tout $t \in Z$

$B$ : opération de retard

Après avoir présenté la définition d'une série chronologique, nous allons décrire les étapes de l'application la procédure ARMA pour modéliser l'évolution du nombre d'entreprises totales :

**Etape 1 : Identification**

Cela signifie trouver les valeurs appropriées de *p*, *q* et *d* (nombre de différenciation de la série). Nous avons vu que le corrélogramme et le corrélogramme partiel sont des aides pour cette tâche.

**Etape 2 : Estimation**

Après avoir identifié les valeurs de *p* et *q* adéquates, il faut estimer les paramètres des termes autorégressifs et de moyennes mobiles inclus dans le modèle.

**Etape 3 ; Choix du modèle**

Après le choix d'un modèle ARMA et de ses paramètres, nous vérifions si le modèle ajuste bien les données car il se peut qu'un autre modèle ARMA soit aussi satisfaisant. Plusieurs tests nous permettront de définir le modèle le plus adapté.

**Etape 4 : Prévision**

Nous verrons, par la suite, les prévisions qui peuvent être effectuées à l'aide du modèle identifié aux étapes précédentes.

### 4.2 Application du modèle ARMA

Les données que nous allons analyser proviennent du nombre d'entreprises totales de l'INSEE qui couvre la période de 1993 à 2009. Comme le modèle ARMA requiert un nombre d'observations, nous avons ajusté les donnes annuelles aux données trimestrielles à l'aide de la méthode linéaire. Cela nous permet d'avoir 64 observations au lieu de 16 observations au départ. Néanmoins, nous risquons de créer des points dits artificiels qui peuvent éventuellement biaiser le résultat des estimations.

Par la suite, nous allons suivre les étapes mentionnées précédemment pour atteindre l'objectif de la prévision.

## 4.2 Choix du modèle

### 4.2.1 Identification

Cette étape consiste à observer la nature des fonctions d'autocorrélation, d'autocorrélation partielle, et d'autocorrélation inverse, de la série, pour chercher à identifier le modèle ARMA tout en s'aidant de sa propre expertise.
Le modèle ARMA requiert la stationnarité de la variable, nous allons effectuer l'étude sur la stationnarité de celle-ci.

**Etape 1 : Etude de la stationnarité de la série étudiée**

En se référant à la méthodologie décrite dans la partie précédente, l'analyse de la stationnarité est présentée en annexe (*cf.* **Annexe.10**). Selon la stratégie du test de racine unitaire, nous constatons que le nombre des entreprises défaillantes est un processus $I(1)$, ceci signifie que la série est non stationnaire de type stochastique (*TS*). Il suffit de différencier la série pour la rendre stationnaire. Dans notre étude, nous obtenons $d$=1. La série sur laquelle nous travaillons est donc :

$$X_t = (1-B)Y_t$$

**Etape 2 : Identification des modèles potentiels**

Une fois que la série est stationnaire, nous proposons un ou des modèles à l'aide des tests et les diagrammes de la fonction d'autocorrélation.
Pour identifier le modèle, il faut visualiser les fonctions d'autocorrélation, simples et partielles: en vertu des propriétés établies de la série ARIMA : si l'ACF (fonction d'autocorélation) s'écrase après le rang $q$, cela suggère une composante *MA(q)*. Si la PACF (fonction d'autocorrélation partielle) s'écrase après le rang $p$, cela suggère une composante *AR(p)*.
Il faut prendre garde à ne pas sur-paramétrer en même temps la composante *MA* et la composante *AR*. Considérer un *AR* pur d'ordre élevé, ou bien un *MA* pur d'ordre élevé, ne pose par contre pas de problème de stabilité dans l'estimation.
Les sorties SAS correspondantes nous suggèrent deux modèles potentilles :

Autocorrelations

| Lag | Covariance | Correlation | -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1 |
|---|---|---|---|
| 0 | 4044818 | 1.00000 | \|********************\| |
| 1 | 3325904 | 0.82226 | \|****************  \| |
| 2 | 2042288 | 0.50491 | \|**********  \| |
| 3 | 758671 | 0.18757 | \|****  \| |
| 4 | -203310 | -.05026 | \|  \| |
| 5 | -317855 | -.07858 | \|  \| |
| 6 | -92001.232 | -.02275 | \|  \| |
| 7 | 133853 | 0.03309 | \|*  \| |
| 8 | 288418 | 0.07131 | \|*  \| |
| 9 | 173348 | 0.04286 | \|*  \| |
| 10 | -60612.068 | -.01499 | \|  \| |
| 11 | -294572 | -.07283 | \|  \| |
| 12 | -543175 | -.13429 | \|  \| |
| 13 | -726189 | -.17954 | \|  \| |
| 14 | -868141 | -.21463 | \|  \| |
| 15 | -1010092 | -.24973 | \|  \| |

**Fig. 2.17 Fonction d'autocorrélation empirique du nombre des entreprises**

The ARIMA Procedure

Partial Autocorrelations

| Lag | Correlation | -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1 |
|---|---|---|
| 1 | 0.82226 | \|  \| |
| 2 | -0.52859 | \|  \| |
| 3 | -0.05323 | \|  \| |
| 4 | -0.01659 | \|  \| |
| 5 | 0.43579 | \|  \| |
| 6 | -0.32121 | \|  \| |
| 7 | 0.00443 | \|  \| |
| 8 | -0.00491 | \|  \| |
| 9 | 0.10853 | \|  \| |
| 10 | -0.16760 | \|  \| |
| 11 | -0.04789 | \|  \| |
| 12 | -0.07323 | \|  \| |
| 13 | -0.01577 | \|  \| |
| 14 | -0.13027 | \|  \| |
| 15 | -0.09525 | \|  \| |

**Fig.2.18 Fonction d'autocorrélation empirique partielle du nombre des entreprises**

Le choix entre les différents modèles présenté ici (*i.e AR (p), MA (q), ARMA (p,q)*) ne peut généralement se faire à priori. Nous sommes le plus souvent réduits à des tâtonnements par système d'essais. Il s'agit d'une méthodologie « pas à pas » qui implique la remise en cause de chaque modèle envisagé jusqu'à obtenir un modèle acceptable. Un modèle est dit acceptable lorsqu'il prend en compte toute la structure de la partie aléatoire et ne laisse qu'un bruit blanc. Il faut bien noter qu'il est tout à fait possible d'obtenir plusieurs modèles satisfaisants. En effectuant cette méthodologie à l'aide du SAS, nous proposons les deux modèles suivants :

(1). ARMA (p=2, 4, 6 ; q=2) : $(1-\varphi_2(B)-\varphi_4(B)^4-\varphi_6(B)^6)X_t=(1+\theta_2(B)^2)\varepsilon_t$

(2). ARMA (p=1, 4, 5 ; q=1) : $(1-\varphi_1(B)^1-\varphi_4(B)^4-\varphi_5(B)^5)X_t=(1+\theta_1(B)^1)\varepsilon_t$

Avant de comparer ces deux modèles, nous allons estimer les paramètres de chacun des modèles et tester les significativités des paramètres.

### 4.2.2 Estimation

Nous obtenons alors les deux modèles ARMA particuliers et dont nous pouvons estimer les coefficients. Cette étape n'est pas indépendante de la précédente. Le résultat sous SAS inclut au minimum les informations suivantes: estimation des paramètres du modèle soumis et t-tests associés (tests de leur nullité), estimation de la variance des résidus, test de bruit blanc de Ljung-Box pour la série des résidus obtenue (variante de la statistique de Box-Pierce), valeurs des critères d'information AIC et BIC pour le modèle soumis.
Dans ce mémoire, nous n'allons pas détailler le plan théorique. Les lecteurs intéressés peuvent se référer aux nombreux documents sur les séries temporelles.

La sorties SAS correspondante pour le premier modèle potentiel ARMA (p=2,4,6 ; q= 2) est la suivante :

| The ARIMA Procedure<br>Conditional Least Squares Estimation | | | | | | |
|---|---|---|---|---|---|---|
| **Paramètre** | | **Valeur esimée** | **Erreur type** | **Valeur du test** | **P-Value** | **Retatd** |
| MA | 1 | -1,38 | 0,12 | -11,02 | <0,0001 | 1 |
| MA | 2 | -0,44 | 0,12 | -3,65 | 0,0006 | 2 |
| AR | 1 | 0,73 | 0,15 | 5,05 | <0,0001 | 2 |
| AR | 2 | -5,90 | 0,14 | -4,21 | <0,0001 | 4 |
| AR | 3 | 0,36 | 0,14 | 2,64 | 0,0107 | 6 |

**Tab.3.37 Résultats estimés des paramètres du modèle ARMA (6 ;2)**

La sorties SAS correspondante pour le premier modèle potentiel ARMA (p=1,4,5 ; q= 1) est ci-dessous :

| The ARIMA Procedure<br>Conditional Least Squares Estimation | | | | | | |
|---|---|---|---|---|---|---|
| **Paramètre** | | **Valeur esimée** | **Erreur type** | **Valeur du test** | **P-Value** | **Retatd** |
| MA | 1 | -0,52 | 0,12 | -4,23 | <0,0001 | 1 |
| AR | 1 | 0,86 | 0,09 | 9,56 | <0,0001 | 1 |
| AR | 2 | -0,59 | 0,12 | -4,82 | <0,0001 | 4 |
| AR | 3 | 0,45 | 0,12 | 3,76 | 0,0004 | 5 |

**Tab.3.38 Résultats estimés des paramètres du modèle ARMA (5 ;1)**

- **Test sur paramètres**

Sous l'hypothèse que le bruit blanc est gaussien, les estimateurs $\widehat{\varphi}_t$ et $\widehat{\theta}_t$ sont approximativement gaussiens. Nous pouvons donc effectuer des tests sur les paramètres en utilisant la loi de student :

$$\frac{\widehat{\varphi}_t}{\sqrt{\widehat{V}(\widehat{\varphi}_t)}} \approx T$$

Dans le cadre d'un modèle ARMA (*p,q*), nous pouvons remarquer le test sur un paramètre $\varphi_t$ :

$$H_o : \varphi_t = 0 \text{ contre } H_1 : \varphi_t \neq 0$$

Les résultats montrent que les p-valeurs (*cf.* **Tab.3.26 & Tab.3.26**) de chaque paramètre des deux modèles sont inférieures à 0.05 : les coefficients sont donc tous significatifs.

- **Test du bruit blanc**

Dans l'analyse des séries chronologiques par processus, le bruit blanc joue un rôle particulier puisque c'est un processus sans aucune structure. Quand pour un processus $X_t$ , nous avons éliminé toute tendance, toute saisonnalité et toute dépendance vis-à-vis du passé, il reste un processus bruit banc imprévisible sur lequel il n'y a plus grand chose à dire.
Le test préliminaire concerne justement ce processus. Quand nous étudions une série, l'hypothèse testée est :

$$H_o : \ X_t \text{ est un bruit blanc"}$$

Si nous acceptons cette hypothèse, l'analyse de la série est achevée : la série étudiée est imprévisible car elle n'a aucune structure. Sinon, il faut retourner aux étapes précédentes ou passer à un autre modèle. Pour nos deux modèles, l'hypothèse que la

série est un bruit blanc est rejetée. Les résultats du test de bruit blanc sont en annexe (*cf.* **Annexe.11**). Nous remarquons que les modèles concurrentiels sont adéquats sur le plan des tests statistiques. Le choix du modèle le mieux adapté fait l'objet de l'étape suivante.

### 4.2.3 Choix du modèle

Pour choisir le modèle le mieux adapté, nous avons recours aux critères AIC et BIC. Le règle de décision est de sélectionner le modèle qui minimise les critères AIC et BIC. Le tableau récapitulatif des critères sur les modèles potentiels :

| SAS ARIMA Procedure<br>Critère de choix | | | | |
|---|---|---|---|---|
| **Modèle** | **Retards** | **Déviance** | **AIC** | **BIC** |
| **1** | **AR=2,4,6**<br>**MA=2** | 872,5 | 1036,8 | 1047,5 |
| **2** | **AR=1,4,5**<br>**MA=1** | **866,3** | **1034,9** | **1043,5** |

**Tab.3.39 Résultat détaillé de critère de choix pour les modèles pontentiels**

Au vu de la valeur des critères AIC, BIC, le deuxième modèle semble plus adapté. L'équation du modèle ARMA (5,1) est la suivante :

$$1+0.86\times X_{t-1}-0.59\times X_{t-4}+0.45\times X_{t-5}=1-0.52\times\varepsilon_{t-1}$$

Nous pouvons envisager d'effectuer des prédictions une fois le modèle sélectionné.

## Section 5 Prévision du taux de Défaillance avec le GLM et ARMA

Le modèle de prévision étant composé de deux sous modèles : le modèle linéaire généralisé qui consiste à modéliser la distribution du nombre d'entreprises défaillantes, qui est le numérateur du taux de défaillance, et la procédure ARMA qui consiste à calibrer le nombre d'entreprises totales, qui est le dénominateur du taux.
Dans cette partie de l'étude, nous nous bornerons à procéder aux étapes de la prévision. Comme nous avons décrit le schéma de la prévision dans le chapitre précédent, elle se déroule selon les étapes suivantes : analyse de la performance du modèle, estimations des variables explicatives et détermination de la prévision.

### 5.1 Analyse de la performance du modèle

Le préalable de cette phase consiste à faire un « back-testing » avec les données réelles appliquées sur le modèle estimé afin d'analyser la pertinence du modèle calibré.

Nous commençons d'abord par analyser le nombre d'entreprises défaillantes estimés par le modèle GLM en comparant avec les données réelles (*cf.* **Fig.3.20**).

.

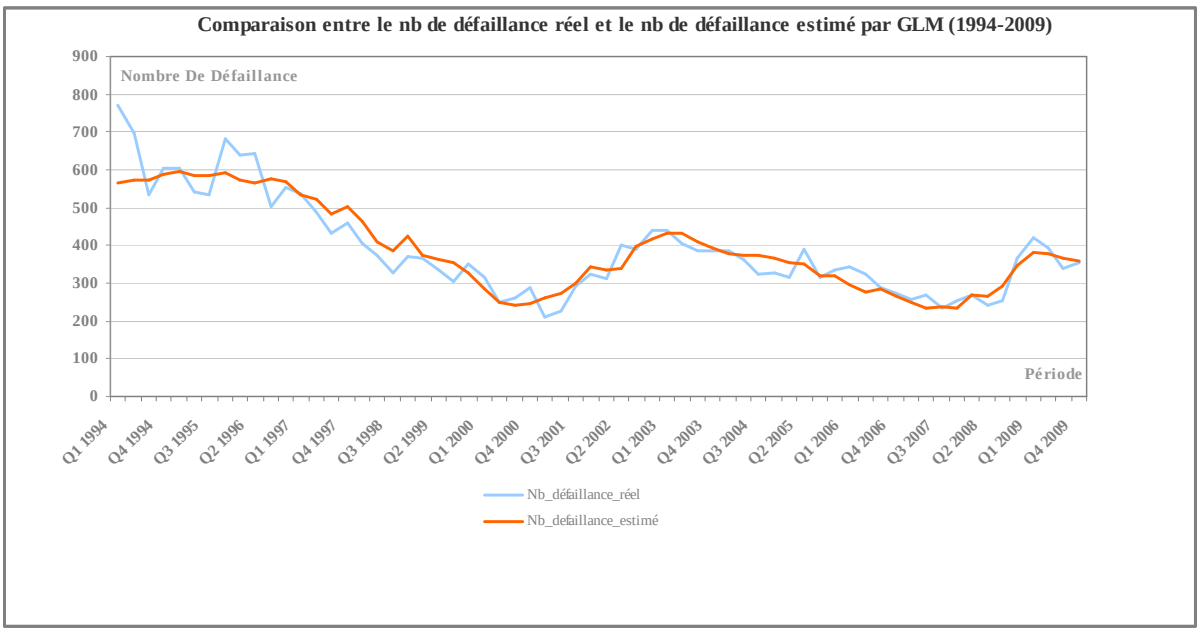



**Fig. 3.20 Comparaison du nombre de défaillance des entreprises réelle et celui de défaillance des entreprises estimé par GLM**

Nous remarquons que le modèle GLM utilisés dans le cas de l'estimation du nombre d'entreprises défaillantes reflète de manière approximative le nombre réel. Le modèle semble pouvoir refléter les tendances au travers toute la période.

Ensuite, nous allons examiner la performance de l'estimation du nombre des entreprises totales réalisée par la procédure ARMA (5,1) (*cf.* **Fig.3.21**).

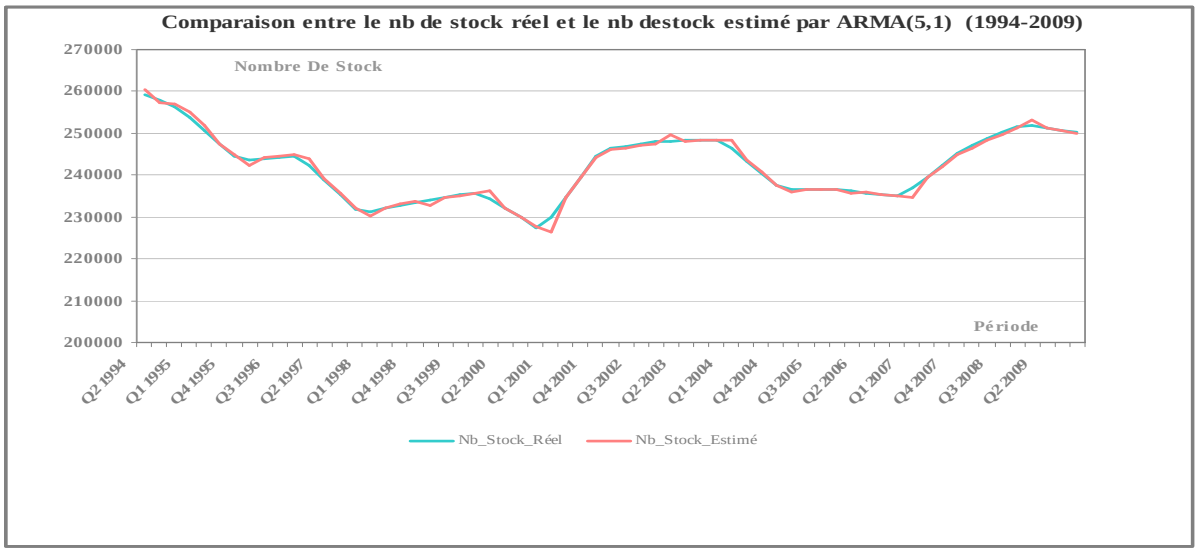



**Fig. 3.21 Comparaison du nombre d'entreprises réelles et du nombre d'entreprises estimées par ARMA(5,1)**

La comparaison illustrée par le graphique ci-dessus est plutôt satisfaisante, celle-ci nous permet de conclure que le modèle ARMA est valable pour l'estimation du nombre des entreprises en stock.
Enfin, nous allons terminer cette analyse de performance par comparer le taux de défaillance estimé avec le taux réel (*cf.***Fig.3.22**).

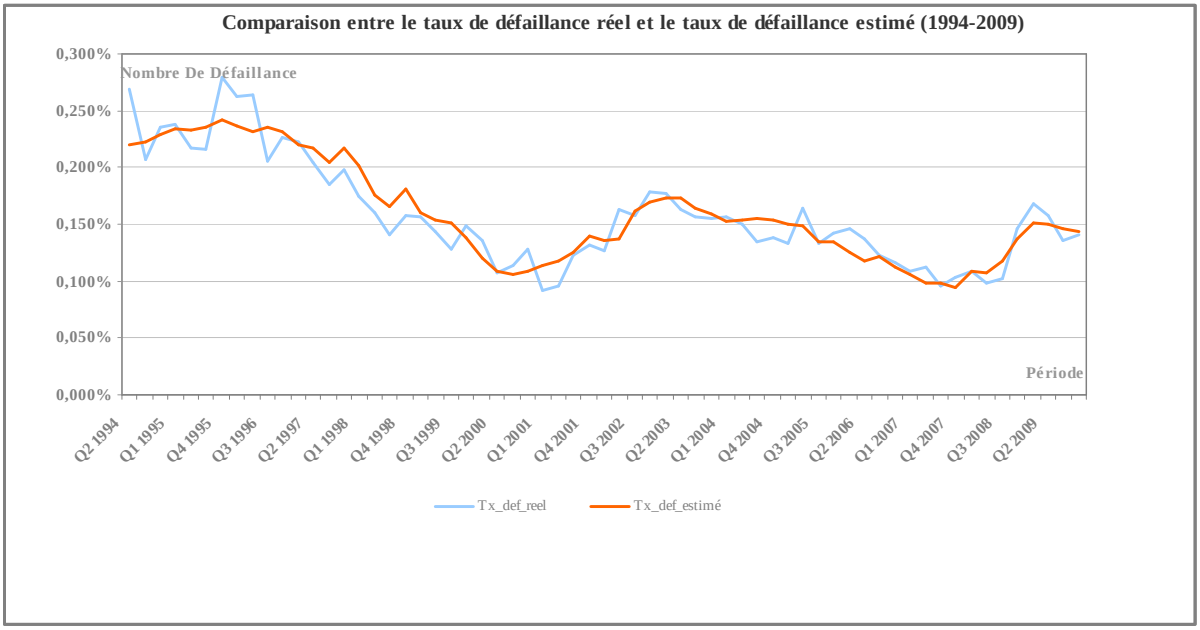



**Fig. 3.22 Comparaison du taux de défaillance réel taux de défaillance estimé pour la période 1994-Q1 et 2009-Q4**

La distance séparant entre les deux courbes montre que la tendance de fond est presque identique. Cette distance est principalement due à l'évolution du nombre d'entreprises défaillantes. Cependant, la courbe du taux de défaillance réel est plus volatile. Ceci montre l'inconvénient du modèle estimé, qui ne parvient pas à capter la fréquence de la série étudiée sur une période donnée. Par ailleurs, cette fréquence élevée de ce taux est principalement due à l'évolution du nombre d'entreprises défaillantes. Ainsi la différence entre le nombre d'entreprises totales réelles et celui d'entreprises estimées par ARIMA est faible. Enfin, nous pouvons constater que le modèle utilisé semble assez pertinent pour estimer l'évolution du taux de défaillance cherché.
Dans la section suivante, nous allons procéder de la même manière afin d'aboutir à la prévision du taux de défaillance.

- **Prévision des variables exogènes**

Les prévisions des variables macroéconomiques concernées (l'indice de CAC 40 et l'investissement des entreprises françaises) sont issues de l'EBA. Elles sont vérifiées et validées par les économistes au sein de Crédit Agricole S.A. Nous allons les appliquer directement.

- **Prévision du taux de défaillance des stress scénarii**

Nous nous fixons ici les mêmes objectifs que dans le chapitre précédent : prévision du taux de défaillance dans le cadre des stress scénarii Bâlois. Plus précisément, nous appliquons le modèle GLM permettant de prévoir l'évolution la variable endogène, soit le nombre d'entreprises défaillantes à l'aide des variables exogènes : CAC 40 et investissement des entreprises. Ainsi, l'approche exposée à la section 4 nous permet de faire la prévision sur le nombre des entreprises françaises totales. Dans la suite, il s'agit de montrer le résultat de la prévision du taux de défaillance dans le scénario Baseline et le scénario Adverse et de les interpréter.

Les deux scénarios économiques détaillés par ECO sont présentés ci-dessous (*cf.* **Tab.3.40 & Tab.3.41**)

| **SCENARIO M0 BASELINE** | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| **Nom** | | **Indicateur** | | **2011** | **2012** |
| **France** | | | | | |
| (4) : valeur en fin d'année | 4 | invest entreprises | | **45236,5** | **47423,7** |
| **UE** | | | | | |
| | 4 | CAC 40 | | **3933,2** | **4310,7** |

**Tab.3.40 Scénario BASELINE**

| **SCENARIO M1 ADVERSE** | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| **Nom** | | **Indicateur** | | **2011** | **2012** |
| **France** | | | | | |
| (4) : valeur en fin d'année | 4 | invest entreprises | | **45135,9** | **46305,5** |
| **UE** | | | | | |
| | 4 | CAC 40 | | **3641,8** | **3122,4** |

**Tab.341 Scénario ADVERSE**

Le stress test de la probabilité de défaut mené sous l'égide du régulateur comporte deux scénarios principaux. A chacun de ces scénarios macroéconomiques est associé un certain nombre de chocs. Le scénario central et le scénario adverse ont été bâtis sur les mêmes hypothèses que ceux du stress test bancaire mené par l'EBA, avec certaines adaptations sont menées par l'ECO toutefois afin de prendre en compte les spécificités de l'activité bancaire.
Comme vu dans la partie précédente, le premier scénario, dit « Scénario Baseline », reproduit une situation réaliste. Le second scénario, dit « Scénario Adverse », présente une détérioration plus importante des principales variables macroéconomiques.
A ces deux scénarios sont associés aux risques de marché et de crédits calqués en grande partie sur les chocs fournis par la BCE pour les exercices de stress tests européens. Néanmoins, pour notre modèle, nous n'appliquons que les deux variables macroéconomiques. Nous supposons que les entreprises sont généralement plus

affectées par une baisse des indices de CAC 40 et une diminution des investissements des entreprises.
Les chocs des indices de CAC 40 et celui des investissements des entreprises se déclinent en chocs sur les activités des entreprises françaises dans notre modèle.

Les résultats d'estimation du taux de défaillance associé au scénario central et au scénario dégradé sont présentés dans la figure suivante (*cf.* **Fig.3.22**).

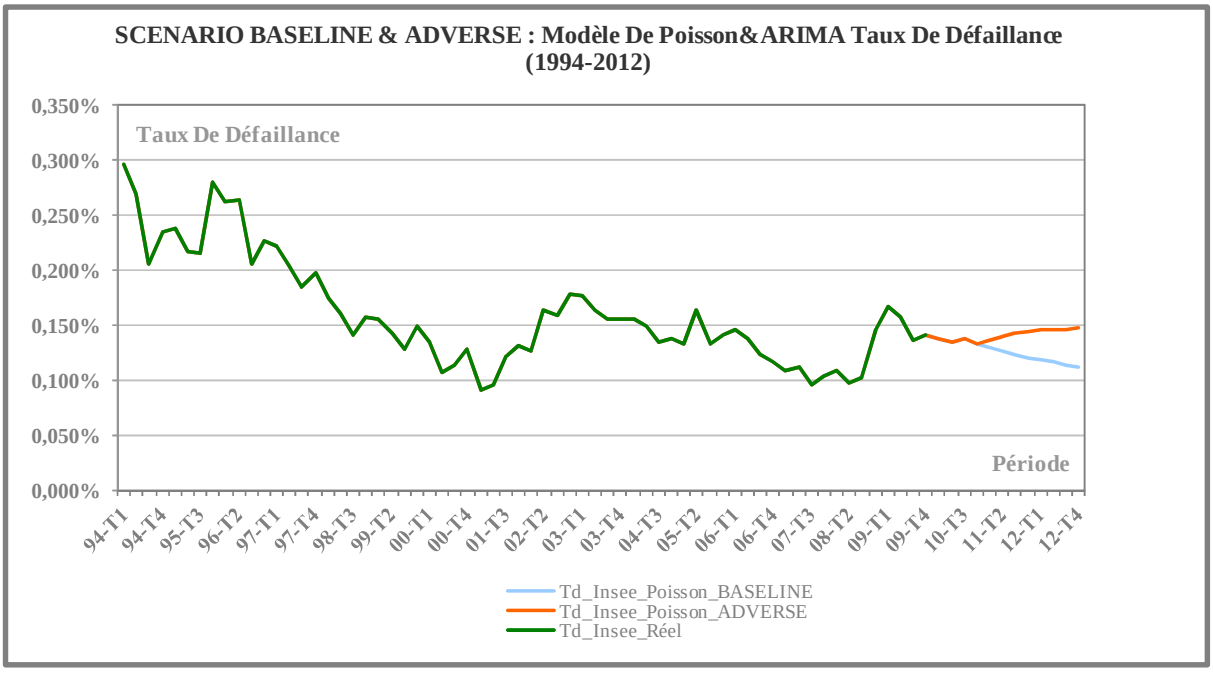



**Fig.3.22 Evolution du taux de défaillance des entreprises françaises de 1994 à 2009 et estimation du taux de défaillance dans les deux scénarii de 2010 à 2012**

La courbe bleue donne une estimation du taux de défaillance, qui a la tendance à la baisse pour les deux années à suivre. Cependant, la courbe orange nous montre le taux de défaillance élevé dans une situation économique et financière dégradée. Nous pouvons remarquer, que la distance entre les deux courbes s'écarte de manière considérable dans les deux scénarii différents. La courbe orange est au dessus de la courbe bleue, cela signifie que la valeur du taux de défaillance du scénario Adverse est supérieure à la valeur du taux de défaillance du scénario Baseline. Le résultat de la prévision du taux de défaillance modélisé par le modèle GLM et ARIMA de deux scénarii se présente en annexe (*cf.* **Annexe.12**).

Etant donné que l'évolution des marchés financiers impactes la santé financière des entreprises françaises : en période de hausse, elle peut encourager de nouveaux flux de placement, mais plus encore elle entraine un effet flux positif. Or, alors que les fondamentaux restent extrêmes fragiles en France, les cours des actions plafonnent autour de 3 200 points (2011-10-16) pendant la période de la préparation mon mémoire. Ainsi la volatilité des marchés boursiers reste très élevée. Conséquence des inquiétudes sur la zone euro et de l'ampleur des déficits publics, l'indice CAC 40 de volatilité s'était notamment de nouveau fortement élevé début août 2011. Ce sont

alors les difficultés de la Grèce et les pays périphériques qui avaient exacerbé la nervosité des investisseurs.
Ainsi, compte tenu de la crise économique qui frappe l'Europe, les projets des secteurs aux investissements lourds comme l'automobile, le transport ont considérablement chuté.
En conséquence, un résultat de la prévision du taux de défaillance modélisé par le modèle GLM en utilisant ces deux variables macroéconomiques serait plus décevant que prévu.

**Conclusion**

A l'aide de la modélisation linéaire généralisé, nous avons déterminé les paramètres caractérisant le nombre des entreprises défaillantes.
Un des avantages de la famille des modèles linéaires généralisés réside dans le fait que l'algorithme est la même pour touts les modèle, quels que soient le choix de la loi de probabilité de la variable à expliquer et de la fonction de lien. Ainsi les procédures offertes dans les logiciels tels que GENMOD dans SAS, permet de construire le modèle ciblé de manière très pratique.
En recourant le recours à la procédure ARIMA, nous avons estimé le nombre total des entreprises françaises en total qui nous permet d'obtenir le taux de défaillance recherché.
Quels que soient les scénarii, le taux de défaillance estimé issu de la combinaison du GLM et de l'ARIMA est significativement plus faible que le taux évalué par le modèle à correction d'erreurs. Cette différence peut expliquer par le nombre de variables macroéconomiques utilisées et la nature des variables macroéconomiques.
En comparant la complexité de deux modèles appliqués et la performance du taux de défaillance estimé, nous procéderons d'appliquer le modèle à correction d'erreurs pour estimer le taux de défaillance de différents secteurs d'activité. Nous allons présenter, ensuite, une application de l'estimation du taux de défaillance par le modèle à correction d'erreurs sur différents secteurs d'activité.

# Chapitre 4 Mise En Œuvre Du Modèle À Correction d'erreurs Sur Les Secteurs D'activité

La méthodologie retenue de la modélisation du taux de défaillance de notre mémoire s'applique à seize secteurs d'activité[30] desquels sont exclues les institutions financières.

Les 16 grands secteurs retenus, caractérisés par un type d'activité économique, sont détaillé ci-après :

| 16 secteurs retenus | |
|---|---|
| 1 | I.A.A. Collecte Approvisionnement |
| 2 | I.A.A. Production et première transformation |
| 3 | I.A.A. Eaux de vie et Champagne |
| 4 | I.A.A. Autres activités |
| 5 | Industries Extractives et Production et Distribution d'électricité, de gaz et d'eau |
| 6 | Construction - BTP |
| 7 | Industries Manufacturières |
| 8 | Négoce International de Matières Premières |
| 9 | Commerce de Gros |
| 10 | Commerce Distribution |
| 11 | Grande Distribution |
| 12 | Transports et Communications |
| 13 | Média Technologie de l'Information |
| 14 | Services |
| 15 | Hôtellerie Loisirs Immobilier hors promotion |
| 16 | Activités financières diverses (hors Banques et Assurances) |

**Tab.3.42 Liste des 16 secteurs retenus**

Nous remarquons le cas particulier parmi les 16 secteurs : le secteur agroalimentaire qui est constitué de quatre sous-secteurs.

Les caractéristiques de chacun des secteurs IAA :

- secteur 1 : « IAA Collecte Approvisionnement » : *simple activité de collecte et de stockage intermédiaire en attendant la livraison ou l'enlèvement, activité de transformation négligeable ; pour l'approvisionnement, activité purement commerciale d'achat en gros et revente en l'état, sans transformation*

- secteur 2 : « IAA Production et première transformation » : les produits végétaux et animaux sont produits et/ou subissent une première transformation du type écrasement, simple mélange (aliments du bétail en vrac ou en sac par exemple), abattage;

[30] En France, selon la définition de l'INSEE : un secteur d'activité regroupe des entreprises de fabrication, de commerce ou de service qui ont la même activité principale

- secteur 3 : « IAA Production Eaux de vie et Champagne » activités dont le procédé d'élaboration dans l'entreprise est de 3 ans minimum ;

- secteur 4 : « IAA Autres activités » : à partir de produits agricoles collectés ou ayant subi une première transformation, fabrication de produits agroalimentaires finis (crus ou cuits), disponibles chez les distributeurs en l'état sans que ces produits aient à subir une quelconque transformation avant leur commercialisation finale.

**Section 1 Périmètre**

Pour un secteur donné, nous sommes censés d'évaluer le taux de défaillance trimestriel à l'aide du modèle à correction d'erreurs. Nous avons vu cependant pour certains secteurs, le nombre des entreprises défaillantes sont assez faibles. En absence de la fréquence des événements, le modèle calibré serait moins robuste. Ainsi, de manière pratique, il est légitime d'optimiser le nombre de modèles à mise en place. A ce stade, il est toutefois possible de regrouper les secteurs en quarte grands groupes. Le choix du regroupement d'activité de secteurs est conseillé par les économistes au sein du Groupe Crédit Agricole, dans la mesure où les secteurs activité réagissent de manière similaire autant qu'en période favorable, qu'en période de ralentissement économique. Le tableau Tab. 3.42 résume les quartes principaux groupes est présenté ci-dessous :

| | |
|---|---|
| **Groupe 1** | |
| 1 | I.A.A. Collecte Approvisionnement |
| 2 | I.A.A. Production et première transformation |
| 3 | I.A.A. Eaux de vie et Champagne |
| 4 | I.A.A. Autres activités |
| 8 | Négoce International de Matières Premières |
| 9 | Commerce de Gros |
| 10 | Commerce Distribution |
| 11 | Grande Distribution |
| **Groupe 2** | |
| 6 | Construction - BTP |
| **Groupe 3** | |
| 5 | Industries Extractives et Production et Distribution d'électricité, de gaz et d'eau |
| 7 | Industries Manufacturières |
| **Groupe 4** | |
| 12 | Transports et Communications |
| 13 | Média Technologie de l'Information |
| 14 | Services |
| 15 | Hôtellerie Loisirs Immobilier hors promotion |

**Tab. 3.42 Liste des quatre groupes**

A noter que la branche regroupe des activités de secteurs n'est pas tout à fait homogène. Nous allons justifier notre groupement avec les graphiques du taux de défaillance des quatre branches pour la période 1994-2009.

- **Groupe « Agro alimentaire »**

Nous avons déjà vu que la définition du secteur agroalimentaire. Le secteur agroalimentaire est un secteur d'activité correspondant à l'ensemble des entreprises qui participent à la production de produits alimentaires.
Dans notre approche, il convient de regrouper tous les autres secteurs concernant les matières premières et les aliments pour former une branche, dite Agro alimentaire

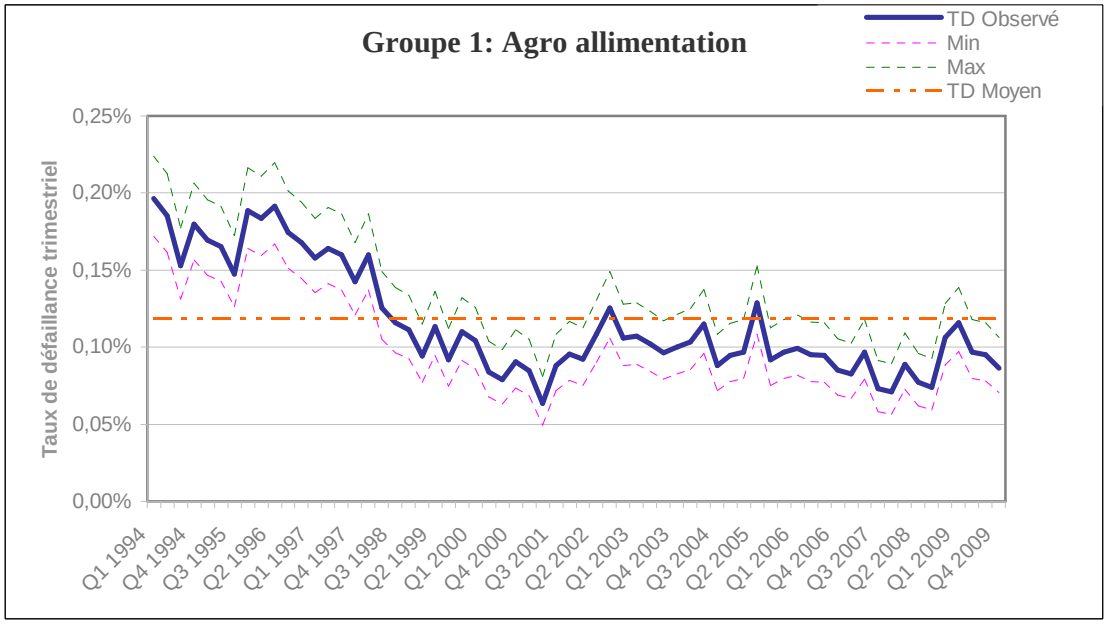



**Fig.3.23 Evolution du taux de défaillance des entreprises françaises du Groupe Agro alimentaire de 1994 à 2009**

- **Groupe « Construction »**

Le secteur du Bâtiment et des Travaux Publics regroupe toutes les activités de conception et de construction des bâtiments publics et privés, industriels et des infrastructures. Le caractéristique unique du secteur nous conduit à le placer dans une catégorie à part.
A noter que le secteur du BTP a été touché par la crise de l'immobilier durant le début des années 1990. En 1991, les ventes ont stagnées. Dans les années qui ont suivi, de 1992 à 1996, les prix de l'immobilier ont constamment baissé. Cette crise a ralenti le secteur et provoqué une forte hausse du taux de défaillance du secteur.

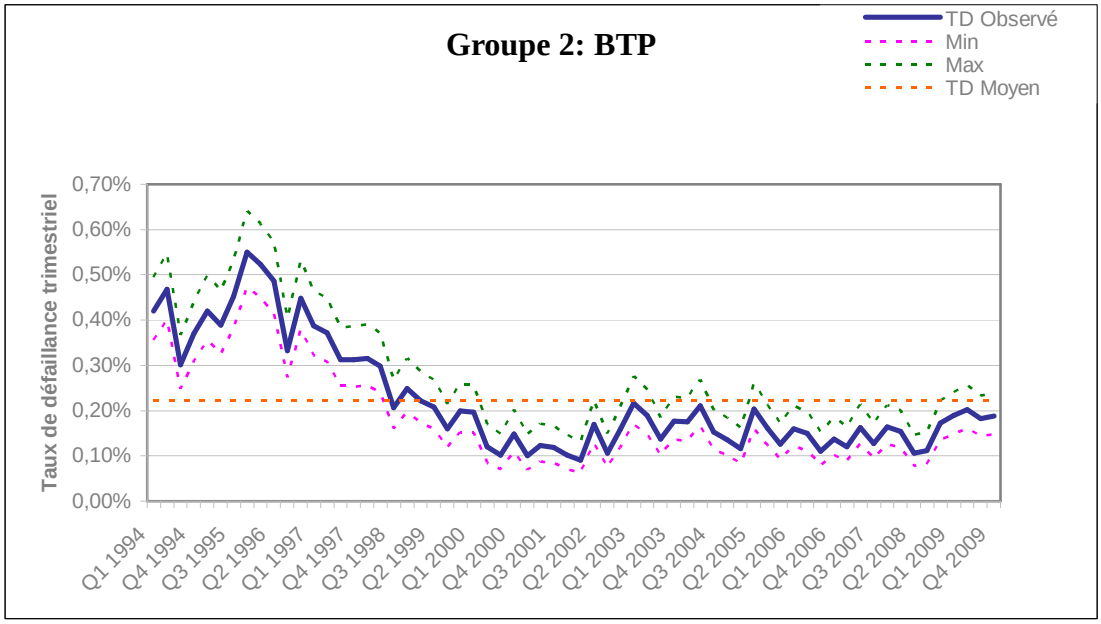



**Fig.3.24 Evolution du taux de défaillance des entreprises françaises du Groupe BTP de 1994 à 2009**

- **Groupe « Industrie »**

Dans le groupe « industrie », l'industrie manufacturière représente une grande partie, qui regroupe les industries de transformation de biens, mais aussi la réparation et l'installation d'équipements industriels. Un des sous secteurs joue un rôle très important dans l'ensemble des secteurs industriels. Selon l'analyse économique, le secteur automobile est frappé de plein fouet par la crise.

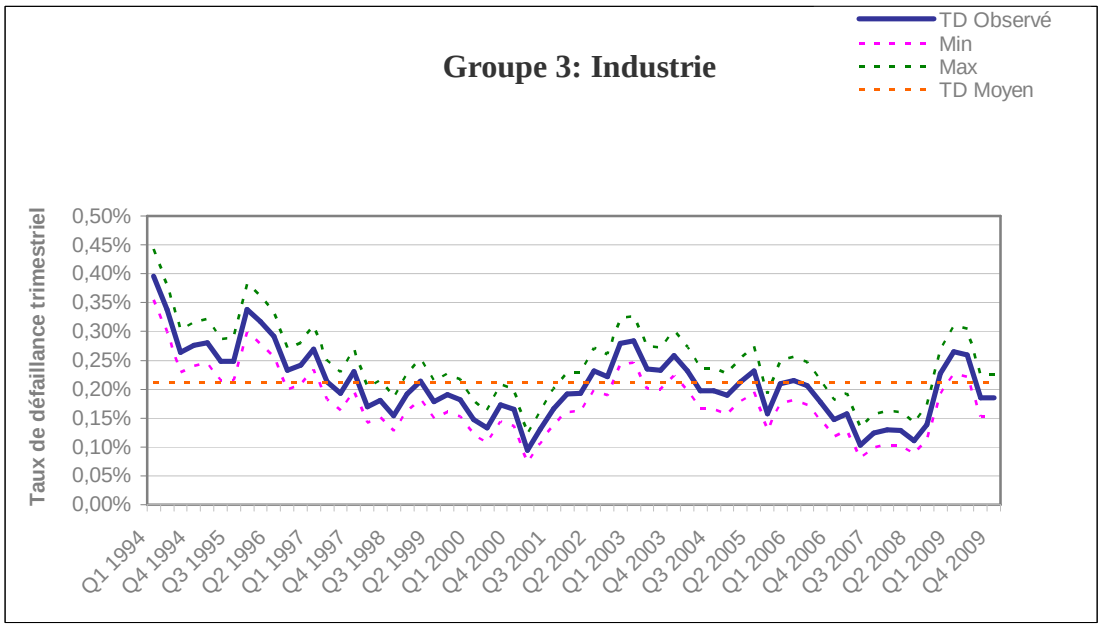



**Fig.3.25 Evolution du taux de défaillance des entreprises françaises du Groupe Industriel de 1994 à 2009**

- **Groupe « Service »**

Le dernier groupe couvre un vaste champ d'activité qui va du transport et communication à l'hôtellerie loisir en passant par les services aux entreprises et aux particuliers, le média technologie. Ce périmètre est de fait défini par complémentarité avec les activités agricoles, BTP et industriels.
Après l'industrie, le groupe « Service » atteint par la crise économique entre en récession.
Depuis le début 2009, la valeur du taux de défaillance a augmenté de manière significative.

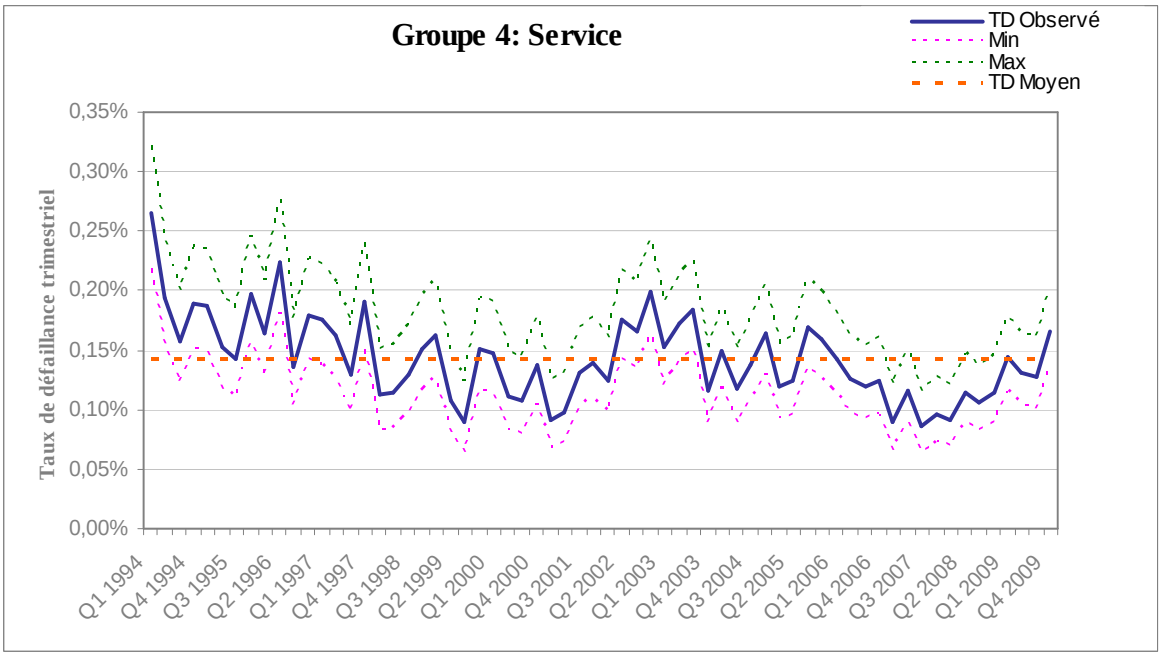



**Fig.3.25 Evolution du taux de défaillance des entreprises françaises du Groupe de service de 1994 à 2009**

Présentons le résultat des prévisions du taux de défaillance estimé par chacun des groupes dans la partie suivante.

## Section 2 Résultat de l'application du modèle sur les différents groupes

Les principaux résultats, obtenus avec soixante observations trimestrielles (i.e. 1994-2009) pour les quatre groupes concernés figurent dans les tableaux suivants (*c.f.* **Tab.3.43, Tab.3.44, Tab.3.45 et Tab.3.46**). La partie supérieure du tableau contient l'estimation obtenue du modèle à long terme. L'estimation du modèle à court terme se situe dans la partie inférieure du tableau, qui a notre préférence.

| groupe 1 | | | |
|---|---|---|---|
| | Modèle à long terme: taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | Intercept | -0,0016 | <0,0001 |
| | taux de 10 ans | 0,0001 | <0,0001 |
| | taux de chômage | 0,0002 | <0,0001 |
| | TC de PIB | -0,0056 | <0,0001 |
| | taux d'inflation sur 12 mois | 0,0001 | <0,0001 |
| | Modèle à court terme: delta du taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | delta taux de 3 mois | -6,0E-05 | 0,0016 |
| | Lag résidu de la relation long terme | -8,0E-01 | <0,0001 |

**Tab. 3.43 Résultat des estimations du MCE pour le Groupe Alimentaire**

| groupe 2 | | | |
|---|---|---|---|
| | Modèle à long terme: taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | Intercept | -0,0059 | <0,0001 |
| | TC de crédit investissement entreprises | -0,0052 | <0,0001 |
| | taux de 3 mois | 0,0003 | <0,0001 |
| | taux de chômage | 0,0008 | <0,0001 |
| | TC de PIB | -0,0202 | <0,0001 |
| | taux d'inflation sur 12 mois | 0,0003 | <0,0001 |
| | Modèle à court terme: delta du taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | intercept | -6,10E-05 | 0,001 |
| | delta taux d'inflation sur 12 mois | 4,87E-04 | 0,0035 |
| | delta taux de 10 ans | -5,91E-04 | 0,0057 |
| | delta taux de chômage | 4,40E-04 | 0,0068 |
| | Lag résidu de la relation long terme | -9,20E-01 | <0,0001 |

**Tab. 3.44 Résultat des estimations du MCE pour le Groupe BTP**

| groupe 3 | | | |
|---|---|---|---|
| | Modèle à long terme: taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | Intercept | 0,0027 | <0,0001 |
| | TC de crédit investissement entreprises | -0,0083 | <0,0001 |
| | TC de PIB | -0,0112 | <0,0001 |
| | Modèle à court terme: delta du taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | intercept | -9,38E-04 | 0,0025 |
| | delta taux de chômage | -5,28E-04 | 0,0003 |
| | delta TC de PIB | -2,09E-02 | 0,00054 |
| | Lag résidu de la relation long terme | -8,19E-01 | <0,0001 |

**Tab. 3.45 Résultat des estimations du MCE pour le Groupe Industrie**

| groupe 4 | | | |
|---|---|---|---|
| | Modèle à long terme: taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | Intercept | 0,0013 | <0,0001 |
| | taux d'inflation sur 12 mois | 0,0001 | <0,0001 |
| | TC de crédit investissement entreprises | -0,0034 | <0,0001 |
| | différence entre tx 3 m et tx 10 ans | 0,00009096 | <0,0001 |
| | Modèle à court terme: delta du taux de défaillance = | | |
| | **Variable** | **Estimation** | **P-Value** |
| | intercept | 0,0007 | 0,00027 |
| | delta taux de 3 mois | -0,0005 | 0,0068 |
| | Lag résidu de la relation long terme | -1,2987 | <0,0001 |

**Tab. 3.46 Résultat des estimations du MCE pour le Groupe Service**

En ce qui concerne les variables macroéconomiques sélectionnées, pour ce qui ressort des effets de long terme et de cour termes les plus significatifs, le taux de croissance du PIB et le taux de chômage sur 12 mois, le taux à 3 mois. L'influence de l'indice boursier CAC 40 n'est importante ni à long terme ni à court terme. Ceci n'est pas le cas pour le modèle global. Ainsi, les coefficients de l'équation de long terme et de court terme ont le signe attendu. Nous pouvons conclure, à présent, que les modèles estimés pour les quatre groupes semblent adéquats. Cette conclusion nous permet d'effectuer les prévisions du taux de défaillance. Cependant, la vérification de la performance du modèle est toujours nécessaire. Les résultats de cette étape de *back-testing* et les analyses sont présentés en annexe (*cf.* **Annexe.13**).

Résultat des prévisions du taux de défaillance pour les différents groupes

Nous exposons dans les graphes suivants (*cf.* **Fig.3.26, Fig.3.27, Fig.3.28 et Fig.3.29**) :

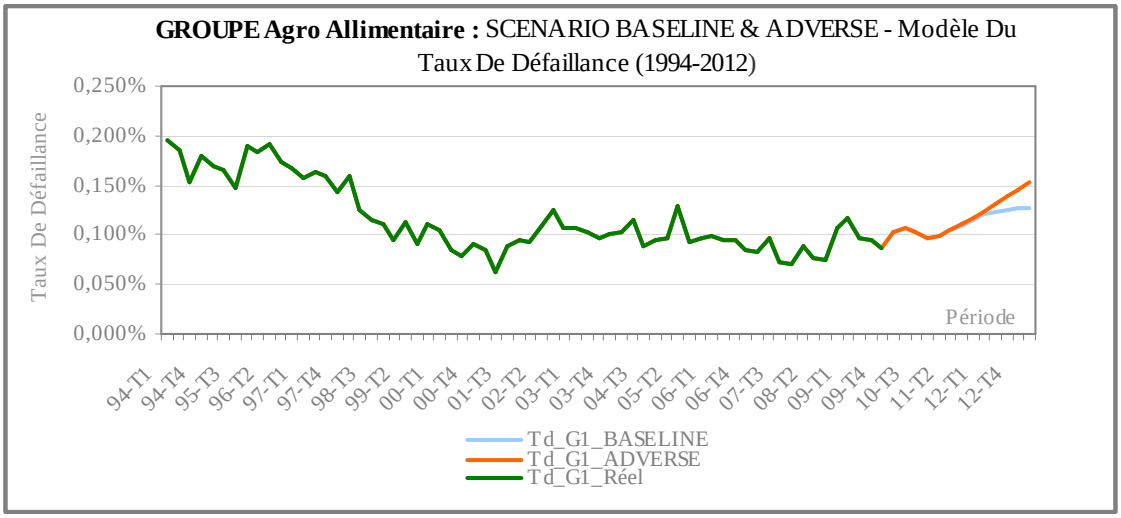



**Fig.3.26 Prévision du taux de défaillance des entreprises françaises du Groupe 1 de 1994 à 2009**

Pour le groupe agro alimentaire, la dégradation des variables macroéconomiques choisies (i.e. taux à 10 ans, taux d'inflation) par le modèle a peu de corrélation avec un éventuel affaiblissement de la santé des entreprises de ce groupe. C'est la raison pour laquelle, l'écart entre les deux scénarii est faible.

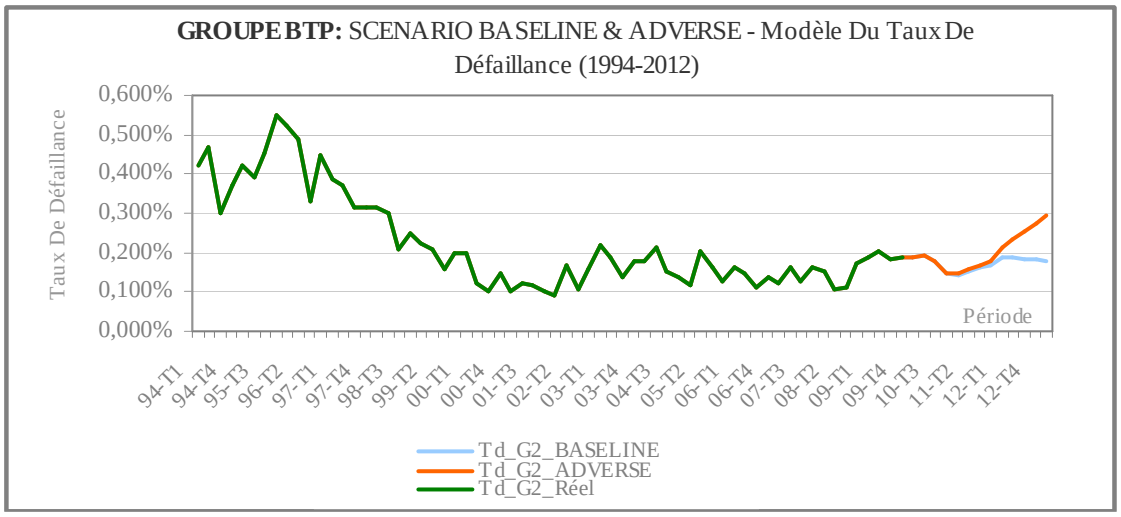



**Fig.3.27 Prévision du taux de défaillance des entreprises françaises du Groupe 2 de 1994 à 2009**

Dans le cas du groupe BTP, l'ensemble des variables macroéconomiques a plus d'impact sur la bonne santé des établissements relevant de ce secteur. A titre d'exemple, si le taux de chômage augmente, les investissements immobiliers des particuliers vont diminuer.

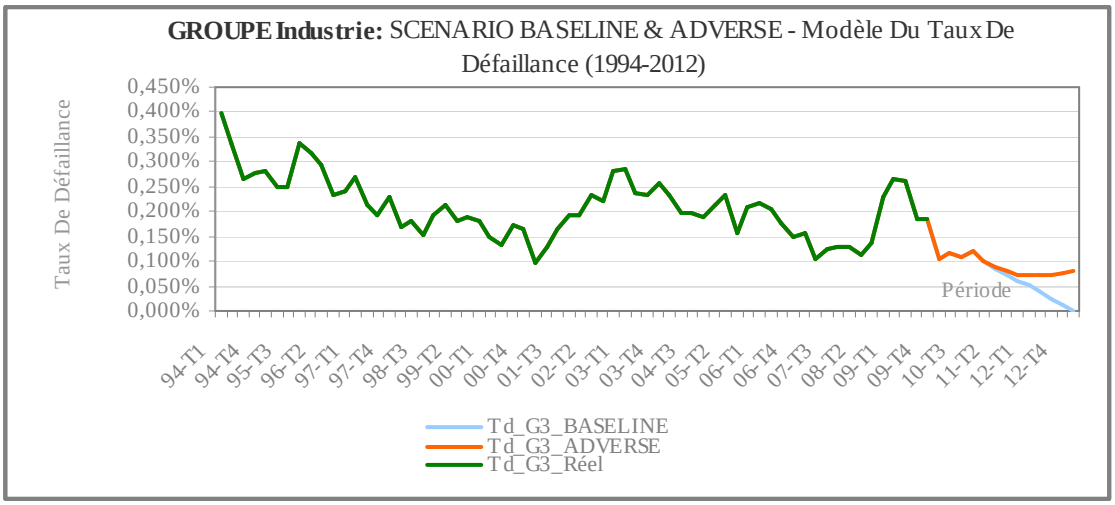



**Fig.3.28 Prévision du taux de défaillance des entreprises françaises du Groupe 3 de 1994 à 2009**

La survie des entreprises industrielles est largement corrélée avec les variables macroéconomiques que nous avons déterminées pour estimer la conjoncture globale. Quand l'économie se dégrade, les entreprises industrielles figurent parmi les plus touchées par la crise. Ceci explique la divergence des deux courbes au fil du temps.

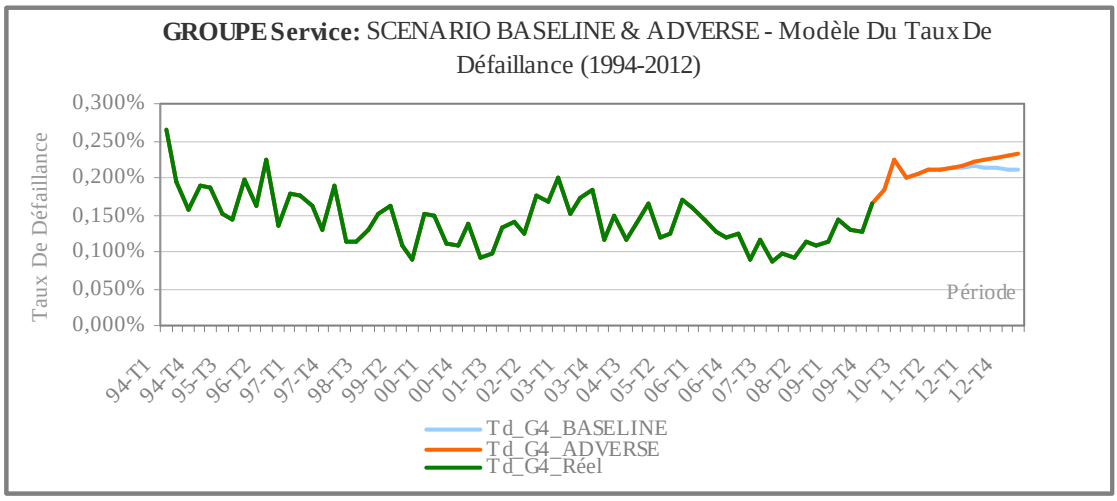



**Fig.3.29 Prévision du taux de défaillance des entreprises françaises du Groupe 4 de 1994 à 2009**

Le graphique ci-dessus présente l'évolution des estimations du taux de défaillance dans les deux scénarii, il met en évidence l'ampleur de la crise qui frappe ce secteur dans le cas de détérioration de la conjoncture économique global. Ce groupe est plus sensible à cette dégradation économique que les autres branches.

## Conclusion

La modélisation du taux de défaillance par groupe s'inscrit dans les mesures de notre étude pour l'amélioration de la qualité des estimations.
L'ensemble des graphiques présentés montre l'évolution du taux de défaillance réel et estimé du modèle linéaire généralisé. Il semble donc bien que les modèles appliqués soient appropriés pour quantifier le taux de défaillance dans le cadre des stress scénarii. L'intérêt de construire un modèle par groupe permet d'avoir une estimation plus sophistiquée en tenant compte des caractéristiques du groupe. Cependant, les données des échantillons sectoriels sont réparties de manière inégale. Ceci peut provoquer parfois une détérioration du modèle appliqué. Néanmoins, le modèle semble donner des résultats réalistes pour les 4 groupes.

# CONCLUSION

Les banques doivent disposer d'un modèle solide pour valider l'exactitude et la cohérence des systèmes et procédures d'estimation de tous les grands facteurs de risque. Dans ce contexte, l'intérêt porté au risque de crédit a été renforcé par les régulateurs et les banques. C'est dans le cadre de la recherche d'un outil, qui permet à la fois de modéliser le risque de crédit d'un portefeuille et de l'évaluer dans des stress scénarii avec respect des exigences réglementaires de Bâle II que nous avons effectué les travaux présentés dans notre mémoire.

Le principal risque inhérent au portefeuille des prêts du secteur bancaire aux entreprises est la défaillance possible des emprunteurs. Du point de vue réglementaire, chaque banque doit élaborer différentes méthodes applicables en fonction des circonstances. Il ne s'agit pas de demander aux banques de prévoir des situations catastrophiques, mais au moins d'envisager les effets de scénarios de récession pour en déterminer l'incidence sur la probabilité de défaut. Concernant les méthodes internes pour évaluer la probabilité de défaut, nous faisons l'hypothèse que le taux de défaillance des entreprises françaises du portefeuille du Groupe Crédit Agricole dépend du niveau d'activité de l'économie française et du niveau des taux d'intérêts du pays.

L'évolution du taux de défaillance est fonction du comportement dynamique des variables économiques. Nous modélisons donc cette dynamique à l'aide de l'approche Engle et Granger.

La construction du modèle s'organise autour d'un certain nombre d'étapes : construction des données, spécification et estimation du modèle, validation et estimation du taux de défaillance.

Devant le manque de données internes, nous avons recours aux données nationales issues de l'INSEE. Après avoir recalibré les données de l'INSEE, nous avons analysé la cointégration, qui permet d'identifier la relation entre le taux de défaillance et les variables macroéconomiques. Lorsque les variables sont cointégrées, nous avons estimé leurs relations au travers d'un modèle à correction d'erreurs en deux étapes. Nous avons estimé la relation de long terme et la relation de court terme par la méthode des moindres carrés. Le modèle à correction d'erreurs exploite les propriétés de non stationnarité des séries macroéconomiques et sont relativement simple à mettre en oeuvre. Cependant, lorsque le nombre de variables explicatives est supérieur à deux, l'approche de Engle et Granger pourrait être insuffisante car elle ne prend pas compte la relation de cointégation entre ces variables.

Néanmoins, la vérification de la performance du modèle à travers un back-testing assure la qualité des prévisions de la défaillance. Les prévisions du taux de défaillance du portefeuille d'entreprises, dans les scénarii Baseline et Adverse, sont dès lors évaluées par le modèle à correction d'erreurs.

Afin de mieux comprendre l'impact de la conjoncture économique et financière sur ce taux, nous avons proposé un modèle alternatif. Ce dernier consiste, dans un premier temps, à évaluer le nombre des entreprises faisant faillite à l'aide du modèle

linéaire généralisé puis, dans un second temps, à estimer le nombre total d'entreprises par la procédure ARMA. Nous avons obtenu le taux de défaillance des entreprises françaises en combinant les deux variables estimées. L'inconvénient principal de ce modèle est d'effectuer deux estimations, ce qui risquent de biaiser le résultat de l'estimation du taux de défaillance.
La comparaison des modèles proposés nous a montré que celui à correction d'erreurs était le plus adéquat. Cette conclusion nous permet de sélectionner l'approche de Engle et Granger afin d'estimer le taux de défaillance pour les différents groupes du périmètre étudié. Les résultats des prévisions du taux de défaillance semblent alors être cohérents et refléter la réalité économique. Malgré tout, nous ne pouvons pas affirmer qu'elles sont correctes car nous n'avons pas examiné toutes les résultats obtenus avec précision. Finalement, un des principaux enseignements de cette étude est d'avoir démontré qu'il est possible d'aborder la question de la défaillance d'entreprises en s'appuyant sur l'économétrie.

# ANNEXE

## A.1. Valeurs critiques des tests de racine unitaire de Dickey-Fuller

| seuil de risque | | 1% | |
|---|---|---|---|
| Modèles | 1 | 2 | 3 |
| T=50 | -4,15 | -3,58 | -2,62 |
| T=100 | -4,04 | -3,51 | -2,6 |
| T=250 | -3,99 | -3,46 | -2,58 |
| seuil de risque | | 5% | |
| Modèles | 1 | 2 | 3 |
| T=50 | -3,5 | -2,93 | -1,95 |
| T=100 | -3,45 | -2,89 | -1,95 |
| T=250 | -3,43 | -2,88 | -1,95 |
| seuil de risque | | 10% | |
| Modèles | 1 | 2 | 3 |
| T=50 | -3,18 | -2,6 | -1,61 |
| T=100 | -3,15 | -2,58 | -1,61 |
| T=250 | -3,13 | -2,57 | -1,62 |

Modèles concernés :

1. $\Delta Y_t = \rho Y_{t-1} + \alpha + \beta t + \sum_{j=1}^{p} \varphi_j \Delta Y_{t-j} + \varepsilon_t$

2. $\Delta Y_t = \rho Y_{t-1} + \alpha + \sum_{j=1}^{p} \varphi_j \Delta Y_{t-j} + \varepsilon_t$

3. $\Delta Y_t = \rho Y_{t-1} + \sum_{j=1}^{p} \varphi_j \Delta Y_{t-j} + \varepsilon_t$

## A.2. Le tableau de MacKinnon en fonction du nombre d'observations

| Formule | | | |
|---|---|---|---|
| n = Nombre d'observations | | | |
| Valeur Critique (5%,n)=B3+B1/n+B2/n$^2$ | | | |
| Seuil de risque | 5% | | |
| Modèles | B3 | B2 | B1 |
| 3: *Zero Mean* | -1,9393 | 0 | -0,398 |
| 2: *Single Mean* | -2,8621 | -8,36 | -2,738 |
| 1: *Trend* | -3,4126 | -17,83 | -4,309 |

## A.3. Le principe de la méthode de Newton - Raphson

La méthode de Newton - Raphson s'applique à des équations du type $f(x) = 0$, pour lesquelles nous pouvons calculer la dérivée de $f$ : $f'(x)$. Soit $x_1$ une valeur

approchée de la racine *a*. Posons : $x_2 = x_1 + h$, et nous cherchons l'accroissement qu'il faut donner à $x_1$, de façon à ce que : $f(x_2) = f(x_1 + h) = 0$
Le principe peut se traduire par les formules suivantes grâce aux développements de Taylor :

$$f(x_1 + h) \approx f(x_1) + hf'(x_1) + \frac{h^2}{2} f''(x_1 + \theta h)....$$

Supposons qu'on néglige les termes d'ordres supérieurs à 1, écrire $f(x_1 + h) = 0$ revient à écrire $f(x_1) + hf'(x_1) \approx 0$, d'où :

$$h = -\frac{f(x_1)}{f'(x_1)}$$

De manière générale, La solution : $x_{n+1} - x_n = h$, soit

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

- **Convergence de la méthode**

Pour $\varepsilon$ petit

$$f(x_1 + \varepsilon) \approx f(x_1) + \varepsilon f'(x_1) + \frac{\varepsilon^2}{2} f''(x_1 + \theta\varepsilon)....$$

D'après la formule de Newton-Raphson

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \Leftrightarrow \varepsilon_{i+1} = \varepsilon_i - \frac{f(x_i)}{f'(x_i)}$$

Quand une solution $x_i$ diffère de la vraie solution $a$ par une quantité $\varepsilon_i = a - x_i$, alors

$$f(a) = f(x_i + \varepsilon_i) = f(x_i) + \varepsilon_i f'(x_i) + \varepsilon_i^2 \frac{f''(x_i)}{2} + ...$$

$$f(a) = 0 \Leftrightarrow f(x_i + \varepsilon_i) = f(x_i) + (a - x_i) f'(x_i) + \varepsilon_i^2 \frac{f''(x_i)}{2} + ... = 0$$

Supposons que $f'(x_i) \neq 0$, on divise la formule par $f'(x_i)$

$$\underbrace{\frac{f(x_i)}{f'(x_i)}}_{= -x_{i+1}} - x_i + a + \varepsilon_i^2 \frac{f''(x_i)}{2f'(x_i)} \approx 0$$

D'où

$$\varepsilon_{i+1} = a - x_{i+1} \approx -\varepsilon_i^2 \frac{f''(x_i)}{2f'(x_i)}$$

L'algorithme de Newton - Raphson approxime le logarithme de la fonction de vraisemblance (i.e. $Vraisemblance = L = \prod_i f(x_i, \theta), i = 1,2,...,n$), dans un voisinage du paramètre initial, par une fonction polynôme qui a la forme d'une parabole concave. Elle a la même pente et la même courbure dans les conditions initiales que la log-fonction de vraisemblance. Il est facile de déterminer le maximum de ce polynôme d'approximation. Ce maximum fournit la seconde étape du processus d'estimation comme la procédure précédente.
Les approximations successives convergent rapidement vers les estimations au sens du maximum de vraisemblance.

**A.4. Modèle Binomial Négative**

Dans cette partie, nous allons définir le modèle BN, ainsi que les liens avec le modèle de Poisson.
La loi binomiale négative dépende de deux paramètres. La loi de probabilité d'une variable aléatoire distribuée selon une binomiale négative de paramètres n et p, noté BN(n,p), prend la forme suivante :

$$f(k, r, p) = p^r (1-p)^{(k)} \frac{\Gamma(r+k)}{\Gamma(r)k!}$$

avec :

$k$ : est la valeur de contage, k=0, 1, 2, …
$p$ : est la probabilité de succès
$r$ : est le paramètre d'une loi de gamma

Maintenant, nous allons introduire l'espérance de cette loi, notons $\lambda = r\frac{1-p}{p}$.
La loi de probabilité devient :

$$g_{\lambda,r}(k) = \frac{\lambda^k}{k!} \cdot \frac{P(r+k-1,k)}{(r+\lambda)^k} \cdot \frac{1}{(1+\frac{\lambda}{r})^r}$$

avec :

$\lambda$ : est un réel positif
r : est un entier naturel non nul

Sur cette paramétrisation, nous avons

$$\lim_{r\to\infty} g_{\lambda,r}(k) = \frac{\lambda^k}{k!} \cdot 1 \cdot \frac{1}{esp(\lambda)}$$

qui est la loi de Poisson avec un paramètre $\lambda$. En d'autres termes, cette loi binomiale négative converge vers une loi de Poisson, et $r$ régit la déviation vis-à-vis de la Poisson. La loi binomiale négative peut donc être vue comme une alternative à la loi de Poisson en cas de sur-dispersion.

## A.5. Les résultats des tests de Dickey-Fuller des variables

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **taux d'endettements entreprises au sens large** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** ***Zéro Mean*** | 2 | 0,22 | 0,730 | 1,23 | **0,943** | | |
| | **Modèle (2):** ***Single Mean*** | 2 | -0,10 | 0,950 | -0,03 | 0,952 | 0,75 | 0,879 |
| | **Modèle (1):** ***Trend*** | 2 | -9,75 | 0,420 | -1,89 | 0,648 | 2,78 | 0,627 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **taux de 3 mois** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 1 | -2,29 | 0,296 | -1,75 | **0,076** | | |
| | **Modèle (2):** *Single Mean* | 1 | -5,65 | 0,363 | -1,84 | 0,358 | 2,28 | 0,497 |
| | **Modèle (1):** *Trend* | 1 | -9,55 | 0,439 | -2,28 | 0,439 | 2,63 | 0,656 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **taux de 10 ans** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -0,74 | 0,517 | -1,49 | **0,127** | | |
| | **Modèle (2):** *Single Mean* | 4 | -2,04 | 0,769 | -0,95 | 0,766 | 1,28 | 0,748 |
| | **Modèle (1):** *Trend* | 5 | -13,45 | 0,211 | -2,16 | 0,503 | 2,33 | 0,714 |

| | Tests De Dickey-Fuller Augment? | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **diff閳ヨence entre tx 3 m et tx 10 ans** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Mod鍒 (3):** *Z閳ヨo Mean* | 1 | -4,64 | 0,135 | -1,39 | **0,151** | | |
| | **Mod鍒 (2):** *Single Mean* | 1 | -14,05 | 0,041 | -2,84 | 0,057 | 4,16 | 0,082 |
| | **Mod鍒 (1):** *Trend* | 1 | -13,97 | 0,191 | -2,81 | 0,198 | 4 | 0,387 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **taux d'inflation sur 12 mois** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -1,20 | 0,436 | -88 | **0,332** | | |
| | **Modèle (2):** *Single Mean* | 4 | -13,67 | 0,044 | -2,1 | 0,246 | 2,25 | 0,504 |
| | **Modèle (1):** *Trend* | 4 | -13,74 | 0,197 | -2,07 | 0,554 | 1,16 | 0,747 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **taux de chômage** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 2 | -0,30 | 0,612 | -0,62 | **0,445** | | |
| | **Modèle (2):** *Single Mean* | 2 | -11,11 | 0,090 | -2,35 | 0,160 | 2,83 | 0,359 |
| | **Modèle (1):** *Trend* | 2 | -23,05 | 0,023 | -2,85 | 0,187 | 4,21 | 0,347 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC de PIB** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -2,80 | 0,247 | -1,22 | **0,202** | | |
| | **Modèle (2):** *Single Mean* | 4 | -7,43 | 0,233 | -1,6 | 0,479 | 1,35 | 0,730 |
| | **Modèle (1):** *Trend* | 4 | -14,05 | 0,184 | -2,21 | 0,476 | 2,53 | 0,675 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC d'investissements d'entreprises** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 2 | -20,13 | 0,001 | -3,14 | 0,002 | | |
| | **Modèle (2):** *Single Mean* | 2 | -31,08 | 0,001 | -3,65 | 0,007 | 6,7 | 0,001 |
| | **Modèle (1):** *Trend* | 2 | -32,14 | 0,002 | -3,59 | **0,039** | 6,66 | **0,046** |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC d'exports** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -5,02 | 0,119 | -1,61 | **0,100** | | |
| | **Modèle (2):** *Single Mean* | 4 | -8,79 | 0,164 | -1,84 | 0,359 | 1,77 | 0,623 |
| | **Modèle (1):** *Trend* | 4 | -35,80 | 0,001 | -2,81 | 0,199 | 3,97 | 0,394 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC d'imports** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -2,90 | 0,239 | -1,19 | **0,211** | | |
| | **Modèle (2):** *Single Mean* | 4 | -5,23 | 0,399 | -1,22 | 0,660 | 0,9 | 0,842 |
| | **Modèle (1):** *Trend* | 4 | -13,69 | 0,199 | -2,14 | 0,516 | 2,63 | 0,655 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC de production industrielle** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 5 | -33,55 | <0,0001 | -2,59 | 0,103 | | |
| | **Modèle (2):** *Single Mean* | 5 | -39,71 | 0,001 | -2,61 | 0,097 | 3,41 | 0,212 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC de crédit trésorie entreprises** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -8,41 | 0,041 | -1,74 | **0,078** | | |
| | **Modèle (2):** *Single Mean* | 4 | -10,31 | 0,110 | -1,82 | 0,368 | 1,67 | 0,649 |
| | **Modèle (1):** *Trend* | 4 | -9,45 | 0,442 | -1,73 | 0,727 | 1,84 | 0,810 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC de CAC 40** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -12,08 | 0,013 | -2,09 | **0,036** | | |
| | **Modèle (2):** *Single Mean* | 4 | -14,52 | 0,035 | -2,25 | 0,193 | 2,53 | 0,434 |
| | **Modèle (1):** *Trend* | 4 | -20,19 | 0,045 | -2,59 | 0,287 | 3,43 | 0,498 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **TC de production industrielle** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 5 | -33,55 | <0,0001 | -2,59 | **0,103** | | |
| | **Modèle (2):** *Single Mean* | 5 | -39,71 | 0,001 | -2,61 | 0,097 | 3,41 | 0,212 |
| | **Modèle (1):** *Trend* | 5 | 168,74 | 0,999 | -3,5 | 0,048 | 6,18 | 0,065 |

| | Tests De Dickey-Fuller Augmenté | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **taux de défaillance INSEE grande entreprise** | **Type** | **Retards** | **Rho** | **P-Value** | **Tau** | **P-Value** | **F** | **P-Value** |
| | **Modèle (3):** *Zéro Mean* | 4 | -0,89 | 0,490 | -0,99 | **0,285** | | |
| | **Modèle (2):** *Single Mean* | 4 | -7,09 | 0,253 | -1,93 | 0,319 | 1,98 | 0,570 |
| | **Modèle (1):** *Trend* | 4 | -18,42 | 0,677 | -2,23 | 0,462 | 2,72 | 0,639 |

## A.6. Résultats des tests statistiques sur les erreurs et les variables correspondantes des modèles 2,3,4

| Variable A Expliquer | Variables Explicatives Modèle 2 | VIF |
|---|---|---|
| *taux de défaillance INSEE grande entreprise* | *taux de chômage* | 4,692 |
| | *taux de 3 mois* | 1,345 |
| | *TC de PIB* | 1,310 |
| | *taux d'inflation sur 12 mois* | 1,801 |
| | *TC de crédit investissement entreprises* | 3,589 |
| Nombre d'observations | 63 | |
| R2 | 0,850 | |
| Durbin-Watson | 1,552 | |
| Test d'hétéroscédastivité: P-Value | 0,419 | |

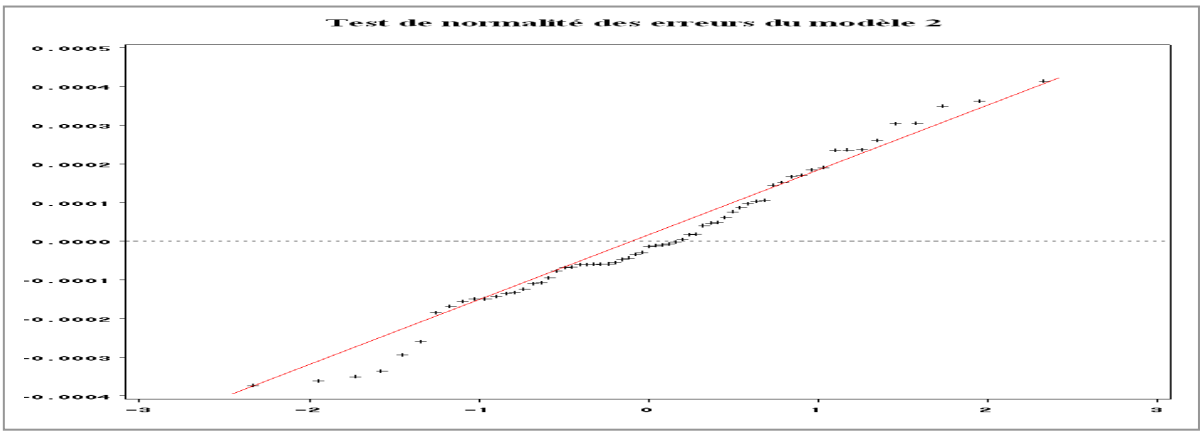
| Variable A Expliquer | Variables Explicatives Modèle 3 | VIF |
|---|---|---|
| *taux de défaillance* | *taux de chômage* | 2,042 |
| *INSEE grande* | *taux de 10 ans* | 1,585 |
| | *TC de PIB* | 2,817 |
| *entreprise* | *taux de 10 ans* | 2,224 |
| Nombre d'observations | 61 | |
| R2 | 0,853 | |
| Durbin-Watson | 1,362 | |
| Test d'hétéroscédastivité: P-Value | 0,799 | |

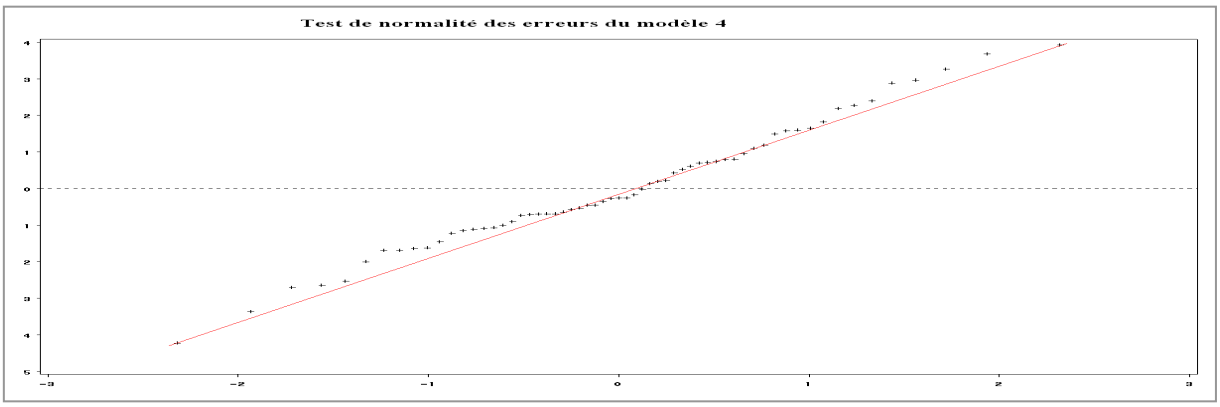
| Variable A Expliquer | Variables Explicatives Modèle 4 | VIF |
|---|---|---|
| *taux de défaillance* | *TC de crédit investissement entreprises* | 2,836 |
| | *taux de 3 mois* | 2,121 |
| *INSEE grande* | *TC de PIB* | 2,250 |
| | *différence entre tx 3 m et tx 10 ans* | 2,957 |
| *entreprise* | *aux d'endettement entreprises au sens large* | 2,819 |
| Nombre d'observations | 61 | |
| R2 | 0,854 | |
| Durbin-Watson | 1,357 | |
| Test d'hétéroscédastivité: P-Value | 0,663 | |

| Proportion De Variation Modèle 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Nombre** | **Valeur Propre** | **Index de condition** | ***TC de crédit investissement entreprises*** | ***taux de 10 ans*** | ***taux d'endettement entreprises au sens large*** | ***TC de PIB*** | |
| **1** | 2,277 | 1,000 | 0,047 | 0,070 | 0,053 | 0,038 | |
| **2** | 1,037 | 1,482 | 0,162 | 0,073 | 0,021 | 0,205 | |
| **3** | 0,496 | 2,144 | 0,210 | 0,763 | 0,119 | 0,019 | |
| **4** | 0,191 | 3,457 | 0,581 | 0,094 | **0,807** | **0,737** | |

| Proportion De Variation Modèle 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Nombre** | **Valeur Propre** | **Index de condition** | ***taux de chômage*** | ***taux de 3 mois*** | ***TC de PIB*** | ***taux d'inflation sur 12 mois*** | ***TC de crédit investissement entreprises*** |
| **1** | 2,009 | 1,000 | 0,049 | 0,010 | 0,021 | 0,023 | 0,052 |
| **2** | 1,533 | 1,144 | 0,001 | 0,212 | 0,117 | 0,115 | 0,000 |
| **3** | 0,831 | 1,555 | 0,000 | 0,015 | 0,509 | 0,153 | 0,067 |
| **4** | 0,508 | 1,990 | 0,001 | 0,762 | 0,200 | 0,311 | 0,033 |
| **5** | 0,119 | 4,112 | **0,949** | 0,001 | 0,153 | 0,398 | **0,848** |

| Proportion De Variation Modèle 4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Nombre** | **Valeur Propre** | **Index de condition** | ***TC de crédit investissement entreprises*** | ***taux de 3 mois*** | ***TC de PIB*** | ***différence entre tx 3 m et tx 10 ans*** | ***taux d'endettement entreprises au sens large*** |
| **1** | 2,048 | 1,000 | 0,031 | 0,027 | 0,062 | 0,001 | 0,071 |
| **2** | 1,803 | 1,066 | 0,042 | 0,060 | 0,016 | 0,091 | 0,001 |
| **3** | 0,751 | 1,652 | 0,119 | 0,251 | 0,178 | 0,003 | 0,019 |
| **4** | 0,244 | 2,896 | 0,002 | 0,206 | 0,460 | 0,320 | 0,593 |
| **5** | 0,154 | 3,653 | **0,807** | 0,456 | 0,285 | 0,584 | **0,316** |



**A.7. Résultats de l'estimation du taux de défaillance des entreprises françaises sur le périmètre corporate de 2010 à 2012 dans les scénarii Adverse et Baseline**

| Trimestre | Td_Insee_BASELINE | Td_Insee_ADVERSE |
|---|---|---|
| 10-T1 | 0,140% | 0,140% |
| 10-T2 | 0,146% | 0,146% |
| 10-T3 | 0,147% | 0,147% |
| 10-T4 | 0,144% | 0,144% |
| 11-T1 | 0,141% | 0,143% |
| 11-T2 | 0,143% | 0,146% |
| 11-T3 | 0,145% | 0,150% |
| 11-T4 | 0,148% | 0,155% |
| 12-T1 | 0,151% | 0,166% |
| 12-T2 | 0,151% | 0,175% |
| 12-T3 | 0,148% | 0,184% |
| 12-T4 | 0,146% | 0,193% |
| 13-T1 | 0,143% | 0,200% |

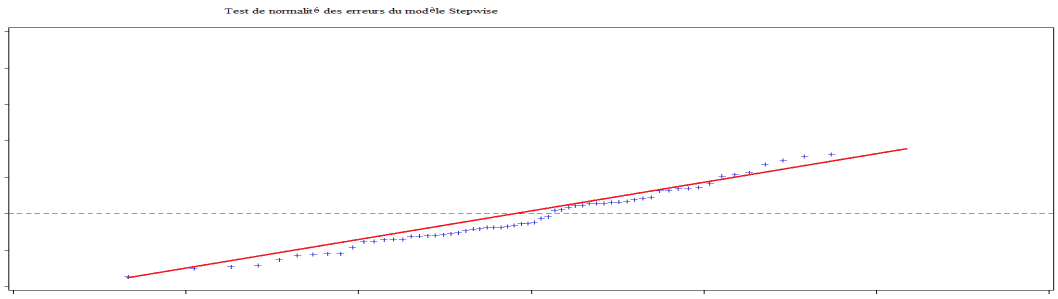
**A.8. Résultats des tests statistiques sur les erreurs et les variables correspondantes des modèles Stepwise et SAS/INSIGHT**

| **Variable A Expliquer** | **Variables Explicatives Modèle STEPWISE** | **VIF** |
|---|---|---|
| *Nombre d'entreprises en défaillance INSEE* | *CAC 40* | 1,347 |
| | *Taux de 10 ans* | 1,347 |
| Nombre d'observations | 64 | |
| R2 | 0,832 | |
| Durbin-Watson | 1,039 | |
| Test d'hétéroscédastivité: P-Value | 0,069 | |

| **Variable A Expliquer** | **Variables Explicatives Modèle INSIGHT** | **VIF** |
|---|---|---|
| *Nombre d'entreprises en défaillance INSEE* | *CAC 40* | 1,846 |
| | *Investissement d'entreprise* | 1,846 |
| Nombre d'observations | 64 | |
| R2 | 0,822 | |
| Durbin-Watson | 0,852 | |
| Test d'hétéroscédastivité: P-Value | 0,242 | |

| Proportion De Variation Modèle Stepwise | | | | |
|---|---|---|---|---|
| **Nombre** | **Valeur Propre** | **Index de condition** | ***CAC 40*** | ***Taux de 10 ans*** |
| **1** | 1,508 | 1,000 | 0,246 | 0,246 |
| **2** | 0,492 | 1,750 | 0,754 | 0,754 |

| Proportion De Variation Modèle INSIGHT | | | | |
|---|---|---|---|---|
| **Nombre** | **Valeur Propre** | **Index de condition** | ***CAC 40*** | ***Investissement d'entreprise*** |
| **1** | 1,677 | 1,000 | 0,162 | 0,162 |
| **2** | 0,323 | 2,278 | 0,838 | 0,838 |




**A.9. Résultat détaillé du test de SAS-GENMOD pour la loi Binomiale Négative avec la fonction de lien Logit pour les variables macroéconomiques : Taux de 10 ans, CAC 40 et CAC 40, Investissement des investissements**

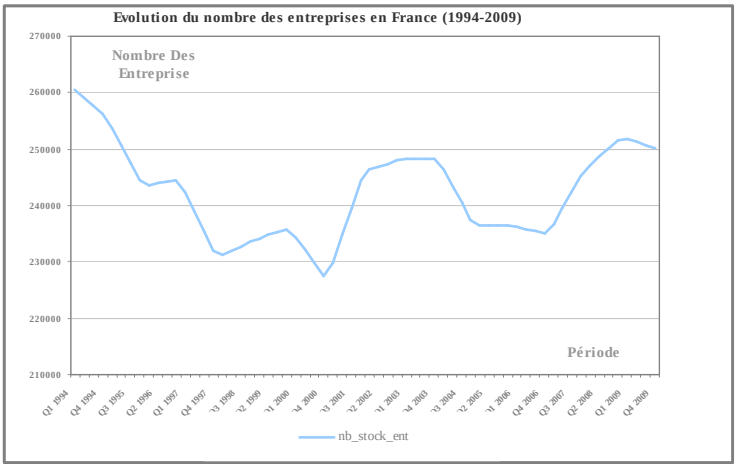
| SAS GENMOD Procedure | | | |
|---|---|---|---|
| **Distribution** | Negative Binomial | | |
| **Link Function** | Logit | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | |
| **Macro Variable** | *Taux de 10 ans* *CAC 40* | | |
| **Critère d'évaluation de l'adquation** | | | |
| **Critère** | **DDL** | **Valeur** | **Valeur/DDL** |
| **Deviance** | 61 | 126,39 | **2,07** |
| **Scaled Deviance** | 61 | 0,28 | 0,005 |
| **Pearson Chi-Square** | 61 | 27638,39 | 453,09 |
| **Log Likelihood** | | 275,87 | |
| **AIC** | | 1657,57 | |
| **BIC** | | 1666,21 | |

| SAS GENMOD Procedure | | | |
|---|---|---|---|
| **Distribution** | Negative Binomial | | |
| **Link Function** | Logit | | |
| **Depend Variable** | Nombre d'entreprises défaillantes | | |
| **Macro Variable** | *Investissement d'entreprise* *CAC 40* | | |
| **Critère d'évaluation de l'adquation** | | | |
| **Critère** | **DDL** | **Valeur** | **Valeur/DDL** |
| **Deviance** | 61 | 126,39 | **2,07** |
| **Scaled Deviance** | 61 | 0,28 | 0,005 |
| **Pearson Chi-Square** | 61 | 27638,39 | 453,09 |
| **Log Likelihood** | | 275,87 | |
| **AIC** | | 1657,57 | |
| **BIC** | | 1666,21 | |

## A.10. Etude de la stationnarité du nombre des entreprises défaillantes en France

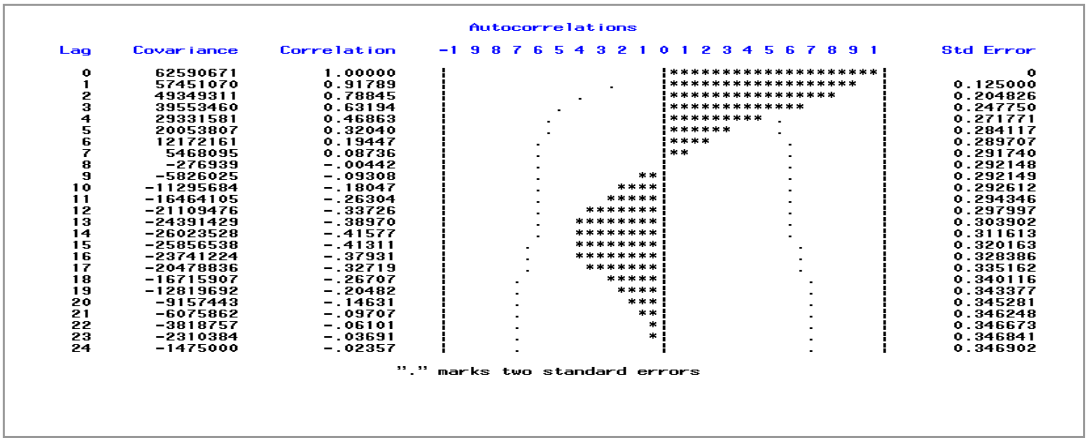
Etape 1 : Etudier la non stationnarité de la série initiale

-Le graphe (*cf.* **A.10.- Fig.1**) de la série nous permet de visualiser l'évolution du nombre des entreprises en France depuis 1994. Nous constatons qu'il y a la présence d'une tendance au cours d'année.




**A.10. - Fig.2 Evolution du nombre des entreprises défaillantes en France (1994-2009)**

- Ainsi, le graphique (*cf.***A.10. - Fig.2**) de la fonction d'autocorrélation empirique de la série nous indique la présence de la non stationnarité, étant donné que la présence d'une tendance décroissante lente de la fonction d'autocorrélation empirique peut montrer la non stationnarité de la série étudiée.




**A.10. - Fig.2 Fonction d'autocorrélation empirique du nombre des entreprises totales**

Enfin, le test de bruit blanc nous permet de constater que la série initiale n'est qu'un bruit blanc, ce qui met en terme à l'analyse. D'après le tableau (cf.**A.10. - Tab.1**), l'hypothèse que la série est un bruit blanc (H0 : la série est un bruit blanc) est rejeté avec une probabilité de 95%.

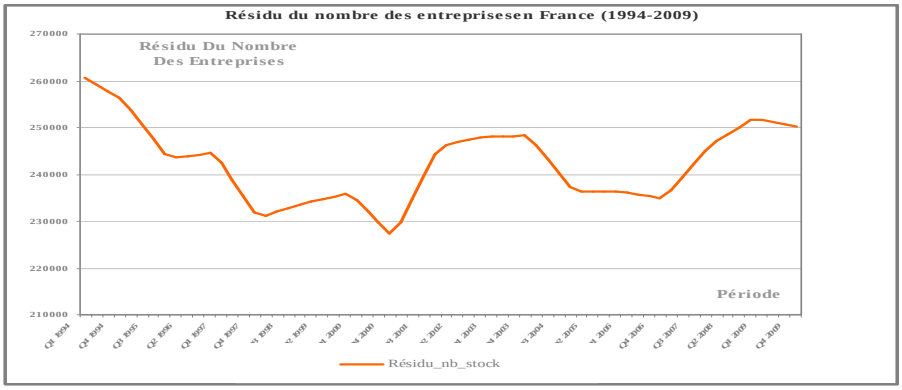
| Tests De Bruit Blanc | | | | | | |
|---|---|---|---|---|---|---|
| Lag | DF | Pr>Khi-2 | Autocorrélations | | | |
| 6 | 6 | <0,0001 | 0,918 | 0,788 | 0,632 | 0,469 |
| 12 | 12 | <0,0001 | 0,087 | -0,004 | -0,093 | -0,18 |

**A.10. - Tab.1 Test de bruit blanc**

Nous pouvons constater la présence de la non stationnarité à l'aide du graphique de l'évolution de la série et le test de bruit blanc.
Par la suite, nous nous intéressons à détecter le type de la non stationnarité afin de rendre la série stationnaire.

Etape 2 : Détecter la non stationnarité

Pour commencer, la non stationnarité du nombre des entreprises est supposée de type déterministe. Nous régressons la série sur une tendance et nous obtenons le résidu de la régression. Ensuite, nous vérifions le résidu de la série par la méthode indiquée dans l'étape 1. Selon le graphique de l'évolution du résidu (*cf.***A.10. - Fig.3**) et son diagramme de la fonction d'autocorrélation empirique (*cf.***A.10. - Fig.4**). Nous concluons donc que la série non stationnaire n'est pas de type déterministe.
Il convient de donc procéder à l'étape de la détection de la non stationnarité stochastique.




**A.10. - Fig.3 Evolution du résidu du nombre des entreprises défaillantes**

Autocorrelations

| Lag | Covariance | Correlation | -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1 | Std Error |
|---|---|---|---|---|
| 0 | 62538733 | 1.00000 | \|********************\| | 0 |
| 1 | 57461497 | 0.91881 | \|******************\ | 0.125000 |
| 2 | 49412073 | 0.79010 | \|****************\ | 0.204956 |
| 3 | 39658155 | 0.63414 | \|*************\ | 0.248022 |
| 4 | 29467442 | 0.47119 | \|*********\ | 0.272180 |
| 5 | 20206655 | 0.32311 | \|******\ | 0.284640 |
| 6 | 12332176 | 0.19719 | \|****\ | 0.290314 |
| 7 | 5625142 | 0.08995 | \|**\ | 0.292399 |
| 8 | -133306 | -.00213 | \| | 0.292831 |
| 9 | -5689342 | -.09097 | \| | 0.292831 |
| 10 | -11146532 | -.17823 | \| | 0.293273 |
| 11 | -16282454 | -.26036 | \| | 0.294960 |
| 12 | -20874688 | -.33379 | \| | 0.298530 |
| 13 | -24104862 | -.38544 | \| | 0.304305 |
| 14 | -25707963 | -.41107 | \| | 0.311840 |
| 15 | -25535499 | -.40831 | \| | 0.320195 |
| 16 | -23438981 | -.37479 | \| | 0.328230 |

"." marks two standard errors

**A.10. - Fig.4 Fonction d'autocorrélation empirique du résidu du nombre des entreprises**

Nous essayons différencier une fois la série pour voir si elle devient stationnaire. Selon le diagramme de la fonction d'autocorrélation (*cf.***A.10. - Fig.5)** et celui de la fonction d'autorrélation partielle (*cf.***A.10. - Fig.46),** nous remarquons la série semble stationnaire après une première différenciation. D'après la nature de la non stationnarité stochastique, la série devient stationnaire après l'avoir différenciée.

Autocorrelations

| Lag | Covariance | Correlation | -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1 | Std Error |
|---|---|---|---|---|
| 0 | 4044818 | 1.00000 | \|********************\| | 0 |
| 1 | 3325904 | 0.82226 | \|****************\ | 0.125988 |
| 2 | 2042288 | 0.50491 | \|**********\ | 0.193228 |
| 3 | 758671 | 0.18757 | \|****\ | 0.213144 |
| 4 | -203310 | -.05026 | \| | 0.215748 |
| 5 | -317855 | -.07858 | \| | 0.215934 |
| 6 | -92001.232 | -.02275 | \| | 0.216387 |
| 7 | 133853 | 0.03309 | \| | 0.216425 |
| 8 | 288418 | 0.07131 | \| | 0.216505 |
| 9 | 173348 | 0.04286 | \| | 0.216878 |
| 10 | -60612.068 | -.01499 | \| | 0.217012 |
| 11 | -294572 | -.07283 | \| | 0.217029 |
| 12 | -543175 | -.13429 | \| | 0.217416 |
| 13 | -726189 | -.17954 | \| | 0.218729 |
| 14 | -868141 | -.21463 | \| | 0.221056 |
| 15 | -1010092 | -.24973 | \| | 0.224339 |

"." marks two standard errors

**A.10. - Fig.5 Fonction d'autocorrélation après une différenciation**

Partial Autocorrelations

| Lag | Correlation | -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1 |
|---|---|---|
| 1 | 0.82226 | \|****************\ |
| 2 | -0.52859 | \|**********\ |
| 3 | -0.05323 | \| |
| 4 | -0.01659 | \| |
| 5 | 0.43579 | \|*********\ |
| 6 | -0.32121 | \|******\ |
| 7 | 0.00443 | \| |
| 8 | -0.00491 | \| |
| 9 | 0.10853 | \|**\ |
| 10 | -0.16760 | \|***\ |
| 11 | -0.04789 | \| |
| 12 | -0.07323 | \| |
| 13 | -0.01577 | \| |
| 14 | -0.13027 | \|***\ |
| 15 | -0.09525 | \|**\ |

**A.10. - Fig.6 Fonction d'autocorrélation partielle après une différenciation**

Nous ne pouvons pas constatons que la nombre des entreprises défaillantes est un processus *I (1)*. Il est nécessaire de le confirmer en menant les tests de Dickey et Fuller. Au vu du diagramme de la fonction d'autocorrélation partielle du test, nous pouvons choisir au préalable le nombre de retards à introduire dans la régression. Selon la stratégie du test de racine unitaire (cf.**A.10. - Tab.2**), nous constatons que le nombre des entreprises est $I(1)$, ceci signifie que la série est non stationnaire de type stochastique (*TS*). Il suffit de différencier la série pour la rendre stationnaire**.**

| Tests De Dickey-Fuller Augumenté | | | | | | | |
|---|---|---|---|---|---|---|---|
| Type | Retards | Rho | P-Value | Tau | P-Value | F | P-Value |
| Modèle (3): | 6 | 0,20 | 0,684 | 0,17 | 0,733 | | |
| Modèle (2): | 6 | -19,95 | 0,007 | -2,14 | 0,229 | 2,32 | 0,485 |
| Modèle (1): | 6 | -14,35 | 0,169 | -2,27 | 0,440 | 3,17 | 0,549 |

**A.10. - Tab.2 Test de DF du nombre des entreprises avec un retard =6**

## A.11. Test de bruit blanc des deux modèles potentiels ARMA

| Tests De Bruit Blanc ARMA(6,2) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Lag | DF | Pr>Khi-2 | Autocorrélations | | | | | |
| 6 | 6 | <0,0001 | 0,822 | 0,505 | 0,188 | -0,05 | -0,079 | -0,023 |
| 12 | 12 | <0,0001 | 0,033 | 0,071 | 0,043 | -0,015 | -0,073 | -0,134 |

**A.11. - Tab.1 Test de bruit blanc ARMA (p=2, 4, 6 ; q=2)**

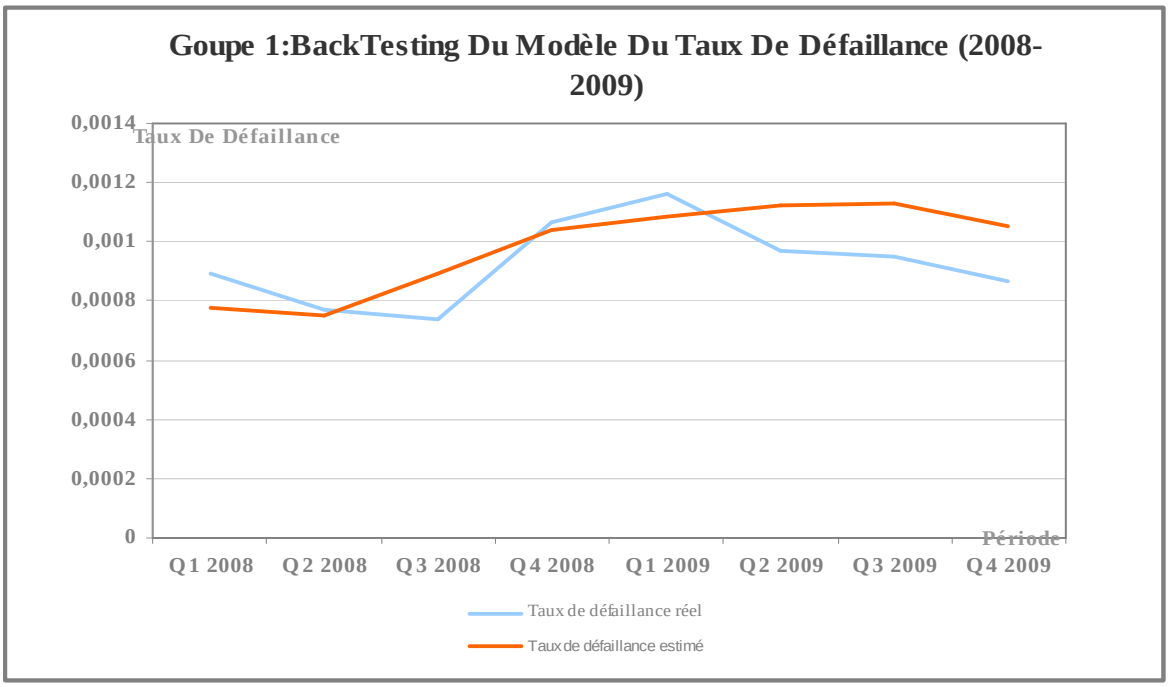
| Tests De Bruit Blanc ARMA(5,1) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Lag | DF | Pr>Khi-2 | Autocorrélations | | | | | |
| 6 | 6 | <0,0001 | 0,822 | 0,505 | 0,188 | -0,05 | -0,079 | -0,023 |
| 12 | 12 | <0,0001 | 0,033 | 0,071 | 0,043 | -0,015 | -0,073 | -0,134 |

**A.11. - Tab.2 Test de bruit blanc ARMA (p=1,4, 5 ; q=1)**

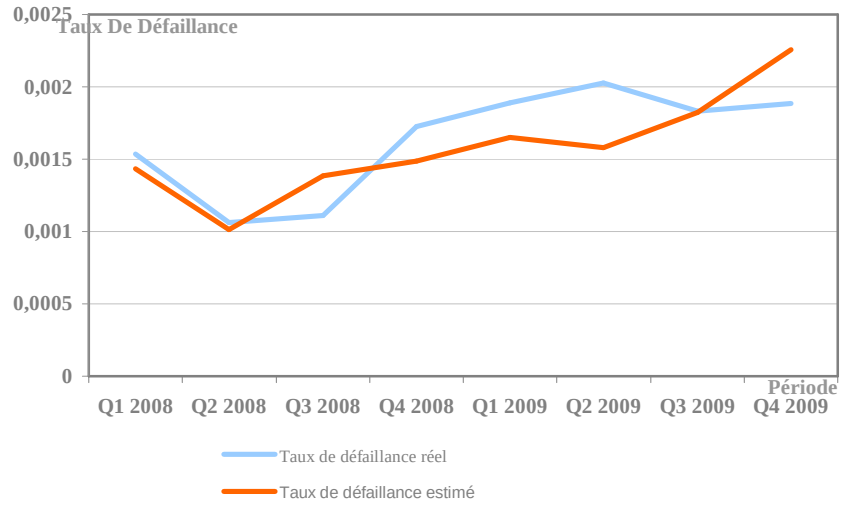
**A.12. Résultat de l'estimation du taux de défaillance des entreprises françaises estimé par le modèle GLM & ARIMA de 2010 à 2012 dans les scénarii Adverse et Baseline**

| Trimestre | Td_Insee_BASELINE | Td_Insee_ADVERSE |
|---|---|---|
| 10-T1 | 0,138% | 0,138% |
| 10-T2 | 0,135% | 0,135% |
| 10-T3 | 0,138% | 0,138% |
| 10-T4 | 0,133% | 0,133% |
| 11-T1 | 0,130% | 0,136% |
| 11-T2 | 0,127% | 0,139% |
| 11-T3 | 0,124% | 0,142% |
| 11-T4 | 0,121% | 0,145% |
| 12-T1 | 0,119% | 0,146% |
| 12-T2 | 0,116% | 0,146% |
| 12-T3 | 0,114% | 0,147% |
| 12-T4 | 0,112% | 0,147% |

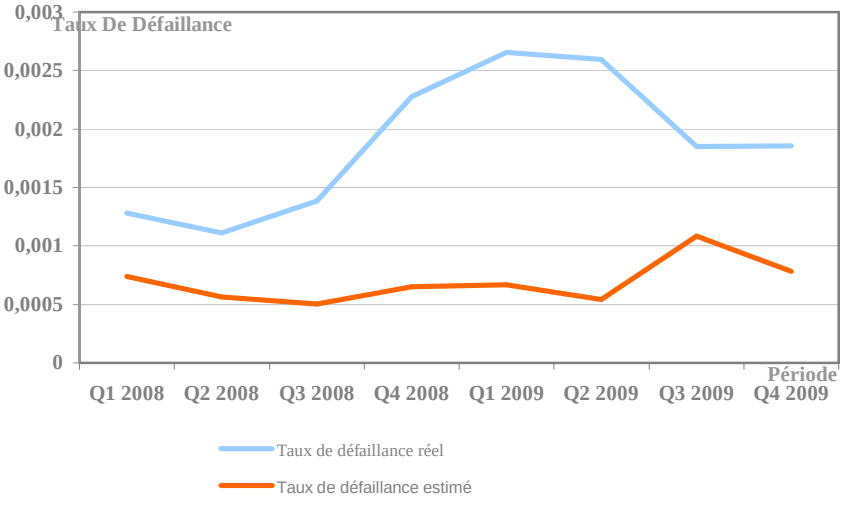
**A.13. Back-testing du modèle par groupe**



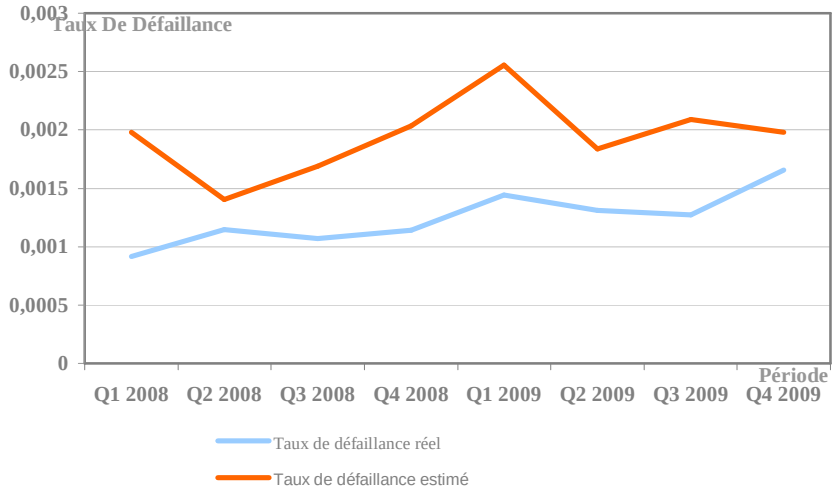
**Goupe 2:BackTesting Du Modèle Du Taux De Défaillance (2008-2009)**




**Goupe 3:BackTesting Du Modèle Du Taux De Défaillance (2008-2009)**



**Goupe 4:BackTesting Du Modèle Du Taux De Défaillance (2008-2009)**




Au vu des graphiques du *back-testing*, nous remarquons que les modèles sont plutôt satisfaisants pour le groupe agro alimentaire et le groupe BTP. Les écarts entre le taux de défaillance estimé et le taux de défaillance réel est assez faible. En plus, la courbe orange trace bien la tendance de l'évaluation du taux de défaillance réel. En revanche, l'estimation du taux de défaillance pour le groupe industriel a la tendance de sous estimer le taux réel. Or, dans le cas du groupe de service, le taux de défaillance réel est surestimé par le modèle. Ces différences peuvent être traduites par le faible nombre des échantillons présentés dans ces groupes.

# BIBLIOGRAPHIE


Jean-Jacques DORESBEKE, Michel LEJEUNE, Gilbert SAPORTA, *Modèles Statistiques Pour Données Qualitatives (Economica, 2005)*

Christian GOURIEROUX, Claude MONTMARQUETTE, *Econométrie Appliquée, (Economica, 1997)*

Stéphane TUFFERY, *Data Mining Et Statistique Décisonnelle, (Technip, 2007)*

Gilbert SAPORTA, *Probabilité Analyse Des Données Et Statistique, ( Technip, 2006)*

Régis BOURBONNAIS, *Econométrie, (DUNOD, 1998)*

Isabelle CADORET, Catherine BENJAMIN, Frank MARTIN, Nadine HERRARD, Steven TANGUY, *Econométrie Appliquée, (De Boeck & Larcier, 2004)*

Patrick ARTUS, Michel DELEAU, Pierre MALGRANGE, Astrid JOURDAN, Célestin C. KOKONENDJI, *Surdispersion et modèle binomial négatif généralisé, (Revue de statistique appliquée, tom 50, 2002)*

Sanvi AVOUYI-DOUVI, Mireille BARDOS, *Macro stress testing with a marcoeconomic credit risk model, (Document de travail de la Banque de France, 2009)*

Christian BORDES, Jaques MELITZ, *Endettement et défaillances d'entreprises en France (Anales d'économétrie et de statistique, 1992)*

Miroslav MISINA, David TESSIER, *La modélisation de l'évolution des taux de défaillance sectoriels en situation de crise : l'importance des non-linéarités (Revue du système financier)*

*Systèmes internes au Groupe Crédit Agricole (Document interne méthodologique, Groupe Crédit Agricole, 2009)*

*Rapport final de validation et calibration des méthodologies de notation « entreprise » du Groupe Crédit Agricole (Document interne, 2009)*