Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

**MASTER EN SCIENCES MATHÉMATIQUES**

# Prédiction de faillite d'entreprise par discrimination bayesienne non paramétrique

Rasir, F.

*Award date:*
1996

Link to publication

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

**Take down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 29. Apr. 2024

Facultés Universitaires Notre-Dame De La Paix
Namur
Faculté des Sciences

# Prédiction de faillite d'entreprise par discrimination bayesienne non paramétrique

Mémoire présenté pour l'obtention du grade
de Licencié en Sciences
mathématiques
par

Promoteur: J.-P. RASSON

Fabrice RASIR

Année académique: 1995-1996

La réalisation d'un travail de fin d'étude est une tâche ardue. C'est pourquoi je tiens spécialement à exprimer mes remerciements à toutes les personnes qui m'ont aidé de près ou de loin durant l'accomplissement de ce mémoire.

Je désire tout d'abord remercier Monsieur Jean-Paul Rasson pour sa présence constante et pour toutes les explications et les conseils fournis tout au long de cette année.

Je tiens également à exprimer ma reconnaissance à Monsieur Charles Van Wymeersch tant pour sa disponibilité que pour mon initiation au monde de l'analyse financière.

De même, je veux remercier les assurances du crédit "NAMUR" qui ont rendu possible cette étude en nous transmettant de précieux échantillons de travail, et plus particulièrement Monsieur Benjamin Decoene pour l'aide qu'il nous a apportée.

Je remercie aussi Mademoiselle Fabienne Montaigne, Madame Sandrine Baudart-Lissoir et Monsieur Vincent Bertholet pour leur soutien, leurs encouragements ainsi que leurs conseils.

Par ailleurs, je tiens à témoigner ma gratitude envers Mademoiselle Françoise Gérard et Monsieur Nicolas Jeannée autant pour l'aide prodiguée lors de la correction des notes que pour l'appui qu'ils m'ont apporté.

Enfin, je remercie vivement mes parents et amis, et en particulier Yannick, pour m'avoir encouragé et soutenu tout au long de cette année.

# Résumé

Ce mémoire a pour objectif la construction d'un outil de détection précoce de défaillance d'entreprise. La méthode mise au point est basée sur l'utilisation de ratios financiers et de techniques de discrimination bayesienne non paramétriques sous l'hypothèse d'un Processus de Poisson non homogène.

Nous étudierons dans un premier temps les fondements économiques nécessaires à la compréhension de ce problème. Ensuite, nous nous attacherons à exposer les principes statistiques qui ont permis l'élaboration de nos méthodes de discrimination. Enfin, nous présenterons les résultats obtenus par l'application de nos méthodes et des procédés de détection classiques à des échantillons de données. Les comparaisons effectuées nous permettront de mesurer la performance de notre modèle.



# Abstract

The purpose of this work is to build a tool for the detection of firm's failure. The method elaborated is based on the use of financial ratios and bayesian non parametric discrimination techniques under the hypothesis of a non homogeneous Poisson process.

First, we will give the economic explanations needful for the comprehension of the problem. After this, we will expose the statistic principles subjacents to our discriminations methods. Then, we will present the results obtained by the application of our methods and classic operating processes of detection on data samples. So, we will be able to measure the model's performance by comparison.

# Table des matières

# Introduction

L'évolution constante et rapide des moyens informatiques durant ces dernières années s'est avérée très profitable pour de nombreux secteurs scientifiques, notamment celui de l'analyse financière. Parallèlement, la publication régulière d'une information comptable normalisée a permis aux spécialistes de ce domaine de développer une série d'outils performants en vue de plusieurs applications.

Nous nous intéresserons ici plus particulièrement à l'étude du *diagnostic financier* des entreprises. Celui-ci doit permettre de mettre en évidence les symptômes de difficultés d'une entreprise dans un but de détection et de prévention. Ce genre d'information intéresse non seulement les créanciers qui verront leurs risques diminuer mais elle concerne aussi d'autres intervenants tels que les actionnaires, les travailleurs, les pouvoirs publics, les tribunaux du commerce, etc.

Pour réaliser cet objectif, les analystes financiers ont recours à la théorie de l'analyse discriminante qui consiste à *classer* parmi plusieurs *groupes* ou *catégories* un individu en fonction de plusieurs observations le concernant, et ce suivant un modèle statistique. Cependant, ce dernier se base sur des hypothèses de normalité sujettes à de nombreuses controverses.

En 1982, J-P. Rasson a proposé une nouvelle méthode de classification utilisant un modèle statistique basé sur un *processus de Poisson non homogène* [4]. Durant plusieurs années, le laboratoire GEOSATEL a appliqué cette technique au problème de télédétection consistant à obtenir à partir d'images *brutes* de régions captées par satellite (ou par voie aérienne) une classification complète des différents éléments au sol.

Les résultats qui ont été obtenus étant supérieurs aux méthodes plus traditionnelles, l'idée de l'application de cette technique au diagnostic financier des entreprises a été mise en avant par Benoît Rasson [1]. Ce mémoire a pour point de mire la continuation de cette expérience.

Les échantillons de travail sur lesquels sont basés la majeure partie des recherches nous ont été confiés par les assurances du crédit "NAMUR".

# Partie I

# Position du problème

# Motivations

La construction d'un modèle permettant d'établir un diagnostic financier des entreprises ne peut s'entreprendre sans une certaine connaissance des tenants et aboutissants du problème étudié. En règle générale, le mathématicien n'est guère familiarisé avec le sens précis des termes "entreprise", "faillite", "ratio", "valeur ajoutée", etc.

L'objectif de cette première partie est de pallier cet inconvénient en effectuant un rapide survol de ces notions qui nous paraissent étrangères, en évitant toutefois de nous perdre dans des détails.

Le premier chapitre propose une introduction sommaire à la notion d'entreprise, son fonctionnement, ses formes et ses règles.

Le second chapitre aborde la comptabilité et en décrit les mécanismes de base.

Le troisième chapitre introduit le concept de faillite d'une entreprise et en explique les signes avant-coureurs, à savoir les ratios.

# Chapitre 1

# La notion d'entreprise

## 1.1 Introduction

Le concept d'entreprise est difficile à synthétiser en une définition unique et universelle. Aussi, à titre d'introduction, nous en proposons trois définitions.

1. "*Association d'hommes* essayant de fabriquer un produit ou de fournir un service afin d'en *tirer un bénéfice*".
2. "*Organisation* génératrice de valeur ajoutée qu'elle répartit entre ses facteurs de production".
3. "*Organisation* où des facteurs de production s'associent pour transformer des *consommations intermédiaires* (*input* de biens et services) en *produits finis* (*output* de biens et services). La valeur ajoutée n'est toutefois *réalisée* que par la *vente* de ces produits finis dans un marché, à un prix déterminé par ce dernier." [6]

### 1.1.1 Facteurs de production et valeur ajoutée

La première définition sous-entend un acte de *production* dont la réalisation n'est possible qu'avec des matériaux et des moyens adéquats. Ces éléments prennent la dénomination de *facteurs de production.* Parmi ces derniers, on distingue le *travail* et le *capital.*

- *Le travail*

Ce premier facteur inclut la contribution directe dans la production par des personnes travaillant avec leurs mains ou leurs idées. On discerne trois aspects du travail :

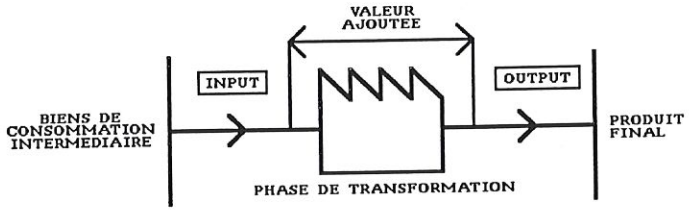
1. *le travail manuel* ou la fabrication proprement dite du produit ou des outils nécessaires à la production,
2. *le travail de direction* ou la coordination des tâches réalisées par les équipes de travailleurs manuels,
3. *le travail d'invention* ou l'élaboration de nouvelles techniques et idées devant permettre l'amélioration de la production.

- *Le capital*

  Ce facteur se compose de *capitaux physiques* et de *capitaux financiers*. Les capitaux physiques sont les équipements de production, les bâtiments, les biens de consommation intermédiaire (matières premières, etc.), les ressources naturelles, etc. Les capitaux financiers désignent les avoirs financiers de l'entreprise.

- *La valeur ajoutée*

  Le schéma suivant illustre la troisième définition.




  La *valeur ajoutée* peut se définir comme

  valeur ajoutée = valeurs de la production − valeurs des biens et services consommés.

  Par biens et services consommés, on entend généralement les *produits finis* d'autres entreprises. Ainsi, l'économie peut être perçue comme étant un gigantesque *réseau d'entreprises.*

## 1.2 Le fonctionnement interne de l'entreprise

L'entreprise, comme nous venons de le voir, est une *association d'individus* qui effectue un acte de production.

### 1.2.1 Les associés

Les fondateurs d'une telle organisation sont appellés *associés.* Il s'agit de personnes investissant de l'argent pour le démarrage et le fonctionnement de l'entreprise. Cet argent, dans sa globalité, prend la dénomination de *capitaux propres de l'entreprise.* Il est à noter que ces associés prennent un risque total sur les sommes qu'ils engagent, celles-ci pouvant totalement disparaître si l'entreprise tombe en état de faillite. A l'inverse, si l'entreprise prospère, le gain peut se révéler très élevé.

Les associés investissent de l'argent dans l'entreprise par l'achat d'*actions.* Il s'agit de documents émis par l'entreprise elle-même et dont le prix va déterminer le montant des capitaux propres de l'entreprise.

A cela s'ajoutent les *capitaux empruntés* provenant de prêts octroyés par des organismes financiers tels que les banques. La somme de tous ces avoirs forme les *capitaux financiers* de l'entreprise.

**Exemple** Les capitaux propres d'une entreprise s'élèvent à vingt millions de francs représentés par 10.000 actions. Chacune d'entre elles vaut 2.000 francs.

### 1.2.2 L'assemblée des associés et le conseil d'administration

Les grandes décisions sont prises par *l'assemblée générale des associés.* Tout actionnaire peut participer aux réunions de cette assemblée et possède un vote dont le poids est déterminé par le pourcentage des capitaux propres détenu par celui-ci. On dira que *le pouvoir de décision est déterminé par le capital.*[1]

Une telle assemblée se déroule au moins une fois par an : le but de cette réunion consiste notamment à élire le *conseil d'administration.* Un tel conseil sera composé *d'administra-*

[1]En pratique, seuls les gros porteurs participent. Ce qui signifie qu'il n'est pas nécessaire de posséder 51% des parts pour avoir la majorité des voix car le capital détenu par les petits porteurs n'est généralement pas représenté.

*teurs* chargés de gérer l'entreprise tout au long de l'année. Ces derniers sont donc directement responsables devant l'assemblée des associés et indirectement devant tous les actionnaires.

### 1.2.3 Actions ordinaires, actions privilégiées et dividendes

Si nous relisons la troisième définition que nous avons faite de l'entreprise au paragraphe 1.1, nous voyons que la valeur ajoutée créée par l'entreprise est répartie entre les différents facteurs de productions.

**Exemple** Une entreprise fabriquant un produit quelconque crée pour vingt millions de francs de valeur ajoutée. La répartion possible d'une telle somme serait :

| Facteur de production | Rémunération |
|---|---|
| Personnel | 12 millions |
| Capitaux physiques | 2 millions |
| Capitaux financiers empruntés | 1 million |
| Capitaux financiers propres | 3 millions |
| Pouvoirs publics | 2 millions |

La somme attribuée aux capitaux propres est divisée par le nombre d'actions que l'entreprise a émises. Le montant obtenu, qui a pour nom *dividende* est distribuéaux actionnaires au pro-rata du nombre d'actions dont ils disposent, moyennant un impôt nommé *précompte.*[2]

Cependant, les administrateurs d'une entreprise en difficulté peuvent décider de diminuer, voire supprimer la rémunération allouée aux capitaux propres. Cela peut se produire pour plusieurs raisons. Par exemple, si l'entreprise traverse des difficultés financières, elle doit avant tout subvenir à des frais d'une plus grande importance comme le paiement du personnel ou le remboursement d'intérêts. Mais il est également possible qu'elle diminue les dividendes en vue de se constituer des réserves ou afin d'effectuer des investissements à court ou à long terme.

[2]En Belgique, le précompte s'élève à environ 28%.

Enfin, il convient de souligner l'existence de deux types d'actions :

1. *action ordinaire* : action donnant droit de vote au propriétaire ;
2. *action privilégiée* : le propriétaire de telles actions est en première ligne pour la distribution des dividendes car rien n'est payé aux porteurs d'actions ordinaires tant que les porteurs d'actions privilégiées n'ont pas reçu un dividende.

## 1.3 Les différentes formes de l'entreprise

Le statut juridique d'une entreprise définit les responsabilités et devoirs de ses propriétaires et est essentiellement fonction du capital engagé et des responsabilités légales que les associés s'apprêtent à assumer. Etant donné que les noms et modalités de ces statuts diffèrent d'un pays à l'autre, nous nous limiterons à évoquer les formes juridiques les plus usitées en Belgique, ainsi que leurs caractéristiques [7]. On distingue notamment deux types d'entreprise : les *entreprises individuelles* et les *sociétés commerciales*.

### 1.3.1 Les entreprises individuelles

Dans l'entreprise individuelle, la personnalité juridique de l'entreprise se confond avec celle de l'entrepreneur qui en est le propriétaire : c'est pourquoi on parlera d'*entreprise personne physique*. L'entrepreneur apporte la plus grande partie des capitaux propres et assume les responsabilités de la direction. Ce status est généralement adopté par les petits commerçants, les artisans, les professions libérales, etc.

Les avantages d'une telle structure sont les suivants.

- L'entrepreneur ne doit partager son bénéfice avec personne.
- Le travail de direction ne subit pas les transactions et marchandages auxquels sont sujettes les directions multiples.
- Une entreprise individuelle possède une grande faculté d'adaptation par rapport au marché.

Les inconvénients sont les suivants.

- La responsabilité financière de l'entrepreneur n'est pas limitée : ses avoirs personnels peuvent servir à *éponger les dettes.*
- L'entreprise cesse ses activités temporairement en cas de maladie de l'entrepreneur et définitivement si ce dernier vient à décéder.
- Les possibilités de crédit de la part d'organismes fincanciers sont relativement restreintes.
- Assumer seul un travail de gestion et de direction n'est pas simple.
- Les bénéfices de l'entreprise sont imposés comme étant le revenu salarial de l'entrepreneur.

### 1.3.2 Les sociétés commerciales

La *société* se définit comme étant une association de deux ou plusieurs personnes mettant quelque chose en commun, en vue de partager le bénéfice pouvant en résulter. Une société constitue une *personne morale*, ce qui signifie qu'elle possède un patrimoine, une nationalité et qu'elle peut intenter des actions en justice.

Nous allons maintenant décrire trois formes de société.

#### La société en nom collectif

On appellera société en nom collectif une association de deux ou plusieurs personnes ayant pour objectif de faire commerce. Le nom de la société identifie celle-ci aux associés. (la "société en nom collectif Dupont et fils"). Une telle structure corrige certains inconvénients de l'entreprise individuelle.

- La société peut survivre, moyennant accord préalable, au décès d'un des associés.
- L'apport en capital est généralement plus élevé.
- Les qualités et le travail des associés permettent de faciliter les tâches d'administration et de gestion.

Cependant, cette organisation engendre des contraintes.

- Les associés sont responsables solidairement.
- Le contrôle de l'entreprise peut être perturbé par des divergences d'opinion au sein des associés.
- La capacité d'adaptation au marché de ce type d'entreprise est inférieure à celle de l'entreprise individuelle.
- Les bases d'imposition et les responsabilités financières sont identiques aux entreprises individuelles.

### La S.P.R.L.

La société privée à responsabilité limitée (S.P.R.L.) est celle où les associés n'engagent que leurs apports. De cette manière, ils protègent leur patrimoine des dettes éventuelles contractées par la société.

Un autre avantage réside dans l'imposition des revenus de la société. Outre un impôt sur les bénéfices, les gains des associés sont perçus sous la forme de dividendes (comme nous l'avons décrit au paragraphe 1.2.3) et ne sont sujets qu'au précompte.

Cependant, la mise en oeuvre d'une telle forme d'entreprise nécessite un apport en capitaux propres d'un montant minimum de 750.000 francs.

### La S.A.

La société anonyme (S.A.) est analogue à la S.P.R.L. Ces deux formes d'entreprises diffèrent sur des points tels que le montant des capitaux propres engagés, le type d'actions autorisé, les contrôles fiscaux, etc. Nous ne nous étendons pas sur ces détails [7].

Cette forme d'exploitation est idéale lorsqu'il s'agit de rassembler un grand nombre d'associés en vue de dégager d'importants capitaux.

**Remarque** Pour se fixer une idée, il faut savoir que la Belgique compte environ 20.000 S.A. et 100.000 S.P.R.L.

### 1.3.3 Franchise et société coopérative

Pour achever ce tour d'horizon, il convient de jeter un coup d'oeil sur des formes hybrides d'entreprise : la franchise et la société coopérative.

#### La franchise

Une société commerciale que l'on appelle le *franchiseur* concède à un commercant ou *franchisé*, moyennant une *redevance de franchise*, l'exploitation d'une marque ou d'un brevet en s'engageant à lui fournir assistance. Cette dernière peut revêtir diverses formes (cours de gestion, location de bâtiment et/ou d'équipement, etc.). Les restaurants *Quick* constituent un excellent exemple de franchise.

#### La société coopérative

Il s'agit de la forme d'entreprise idéale pour tout ce qui concerne l'organisation de la coopération entre producteurs, acheteurs, vendeurs et consommateurs. Le droit de vote est attribué suivant la règle "un associé, une voix". La participation aux bénéfices de la société se fait généralement sous forme de ristournes, non pas sur base du capital détenu mais plutôt sur la participation du coopérateur aux activités et opérations de la société.

A titre d'exemple, citons le cas des fermes qui s'associent au sein d'une coopérative agricole afin d'obtenir un réseau de commercialisation ou encore un équipement à prix réduit.

# Chapitre 2

# La comptabilité

Ainsi que nous l'avons vu au chapitre précédent, l'économie peut être considérée comme un gigantesque réseau d'entreprises. En d'autres termes, toute entreprise entretient des relations de marché avec ses fournisseurs, ses clients et ses facteurs de production.

Dès lors, il est très fréquent qu'une entreprise établisse un *contrat* avec une autre, et ce pour diverses raisons : convention de prêt avec un organisme financier, commande de matières premières avec les fournisseurs, vente de produits finis à des grossistes, etc. On parlera de *relations contractuelles*. Les droits et devoirs de chacun des signataires sont régis par le code civil. Ce qui n'est pas défini dans ce code l'est par le contrat.[1]

Cependant, plusieurs questions restent posées.

- Comment vérifier si un partenaire est réellement capable de respecter ses engagements ? Quels sont ses moyens ?
- De quelle manière un organisme financier peut-il s'assurer que son débiteur lui remboursera son prêt ?
- Comment choisir un fournisseur ? Quel est le plus fiable ?
- etc.

Une solution à toutes ces interrogations réside dans l'*information comptable*. Grâce à celle-ci, l'entreprise se voit dotée d'un puissant outil de surveillance et de décision, pour autant que l'information soit régulière et de nature comparable.

[1] Certains pays comme les pays Anglo-Saxons ne disposent pas de code civil. De ce fait, les contrats adoptent une forme volumineuse.

## 2.1 Le processus de comptabilité

La *comptabilité* est un processus qui se déroule en trois étapes :

1. *La phase de capture des données*

   Cette phase consiste à sélectionner parmi une série illimitée de données une quantité finie d'informations pertinentes sur la situation de l'entreprise.

2. *La phase de traitement des données*

   Les données saisies sont dites *brutes.* Elles doivent subir un traitement avant de devenir une information utile.

3. *La communication de l'information résultante*

   L'information financière obtenue est transformée en information utilisable pour deux groupes distincts d'utilisateurs :

   - Les gestionnaires de l'entreprise (administrateurs, etc.),
   - Les partenaires non-gestionnaires de l'entreprise (pouvoirs publics, créanciers, entreprises concurrentes, etc.).

La comptabilité se scinde donc en deux approches :

1. *La comptabilité analytique*

   Celle-ci est destinée au premier groupe d'utilisateurs. L'information communiquée concerne essentiellement des détails de l'ordre du prix de revient des produits fabriqués, des bénéfices réalisés, etc. Il s'agit d'un outil destiné à la gestion de l'entreprise. Cette information prend donc un caractère *confidentiel.*

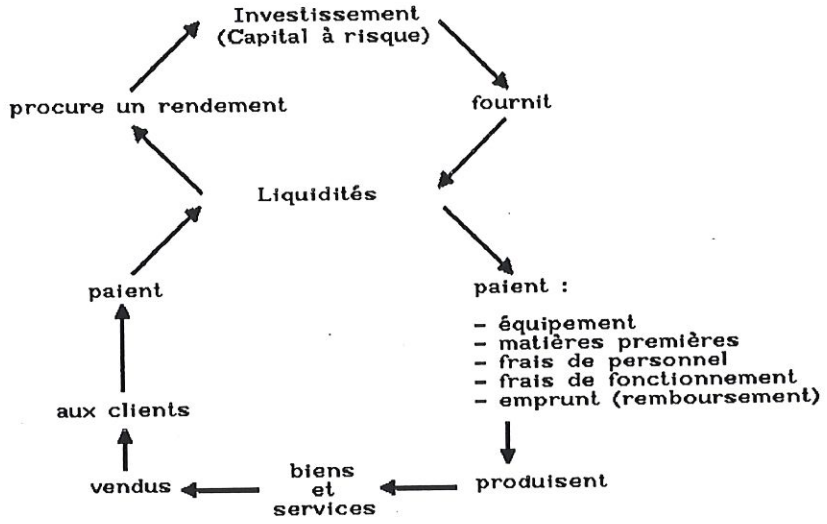
2. *La comptabilité financière*

   Celle-ci est communiquée aux partenaires. Elle rend compte essentiellement de son patrimoine, de ses engagements financiers envers d'autres agents économiques et contient les bénéfices réalisés ou les pertes contractées.

Nous ne développerons ici que l'approche financière.

## 2.2 Le cycle financier de l'entreprise

Le cycle d'exploitation de l'entreprise est largement résumé par le schéma ci-dessous [5].




La comptabilité financière enregistre les états financiers provoqués par ce cycle.

## 2.3 Les mécanismes de base de la comptabilité financière

### 2.3.1 Origine du système comptable

Les prémices de la comptabilité sont apparus à Venise au début du XVIème siècle, dans un ouvrage intitulé *"Summa de arithmetica, geometria, proportioni et proportionalita"*. L'activité économique de Venise était alors en pleine expansion. Cette situation engendrait une série de problèmes de type comptable, qui ne pouvaient être résolus que par une élite de spécialistes.

Les premiers instituts de formation comptable se sont créés à partir du milieu du XVIème siècle. Ces établissements portaient le nom d' *"Institution du Collège des comptables"*.

L'essor économique et intellectuel de la Renaissance propagea ces connaissances dans de nombreuses régions d'Europe. Depuis, le système s'est adapté à travers l'Histoire. On l'utilise de nos jours dans de nombreuses nations.

### 2.3.2 Fonctionnement

La comptabilité s'effectue sur un *livre comptable* ou *bilan* suivant la convention *"à partie double"*. Le bilan est en fait un tableau *à deux colonnes* portant les mentions suivantes :

| Actif | Passif |
|---|---|
| ⋮ | ⋮ |

- *L'actif* précise l'emploi des fonds (immeuble, équipement, stocks disponibles, etc.).
- *Le passif* définit la source des fonds de l'entreprise (capital propre, bénéfices non distribués, dettes envers les banques et fournisseurs, etc.).

Le total des colonnes *est toujours identique.*

La fonction de ce bilan est d'indiquer à l'utilisateur la situation de la provenance et de l'emploi des fonds d'une entreprise *à un moment précis du temps.* Il est à noter que toute opération de l'entreprise en question provoque un changement de cette situation.

### 2.3.3 Exemple introductif

Soit une nouvelle entreprise disposant de 10 millions de capitaux propres. Les gestionnaires décident d'acheter un bâtiment muni d'équipements adéquats pour une somme de 5 millions. A ce moment, le bilan s'organise comme suit :

| Actif | | Passif | |
|---|---|---|---|
| Immeuble et équipements | 5.000.000 | Capitaux propres | 10.000.000 |
| Caisse | 5.000.000 | | |
| Total | 10.000.000 | | 10.000.000 |

L'entreprise décide alors d'acheter pour 2 millions de francs de marchandises (biens de consommation intermédiaire) afin d'effectuer sa première production. Le bilan prend alors la disposition suivante :

| Actif | | Passif | |
|---|---|---|---|
| Immeuble et équipements | 5.000.000 | Capitaux propres | 10.000.000 |
| Marchandises (stock) | 2.000.000 | | |
| Caisse | 3.000.000 | | |
| Total | 10.000.000 | | 10.000.000 |

L'entreprise effectue ensuite une première vente de 500.000 francs de marchandises qu'elle cède au prix de 650.000 francs. Le bénéfice réalisé est donc de 150.000 francs. Ce bénéfice augmente la richesse de l'entreprise et est considéré à juste titre comme une nouvelle source de fonds. Cela donne le bilan :

| Actif | | Passif | |
|---|---|---|---|
| Immeuble et équipements | 5.000.000 | Capitaux propres | 10.000.000 |
| Marchandises (stock) | 1.500.000 | Bénéfice | 150.000 |
| Caisse | 3.650.000 | | |
| Total | 10.150.000 | | 10.150.000 |

Après une certaine période, l'entreprise doit payer *l'impôt des sociétés* qui consiste à imposer le bénéfice réalisé sur une certaine période d'un pourcentage de 33.3%. Le bilan affiche alors :

| Actif | | Passif | |
|---|---|---|---|
| Immeuble et équipements | 5.000.000 | Capitaux propres | 10.000.000 |
| Marchandises (stock) | 1.500.000 | Bénéfice | 100.000 |
| Caisse | 3.650.000 | Administration fiscale | 50.000 |
| Total | 10.150.000 | | 10.150.000 |

Voilà donc une illustration simple du système comptable. L'exemple est bien entendu très incomplet. En effet, une multitude d'opérations peut intervenir dans ce schéma: vente à crédit, investissements en tous genres (matériel, des marchandises ou l'achat actions d'autres sociétés), paiement des dividendes, des charges de personnel et des intérêts des capitaux empruntés, T.V.A., etc. Un extrait d'un véritable bilan se trouve en annexe.

### 2.3.4 Le fichier comptable

Etant donné le nombre élevé des opérations pouvant intervenir dans un schéma financier, le bilan n'est réalisé qu'à des moments précis (en général, au terme d'une année, et ce suivant les dispositions légales en cours). En fait, les opérations que l'entreprise effectue sont enregistrées régulièrement, selon les rubriques auxquelles elles se rapportent sur ce que l'on nomme un *compte.*

A l'instar du bilan, ce compte se conforme à la convention *"double entrée"*. Il s'agit donc également d'un tableau à deux colonnes.

| Débit | Crédit |
|---|---|
| ⋮ | ⋮ |

Une écriture au *débit* d'un compte signifie un *emploi de fond* ( ou *la réduction d'une source*) tandis qu'une écriture au *crédit* constitue *une source de fond* (ou *la réduction d'un emploi*).

### 2.3.5 Exemple de comptes

Reprenons l'exemple du paragraphe 2.3.3 et examinons successivement le compte de caisse (actif) et le compte du bénéfice (passif).

Le compte de caisse, qui est une rubrique de l'actif, s'écrit comme:

| Débit | | Crédit | |
|---|---|---|---|
| Montant initial | 5.000.000 | Achat marchandises | 2.000.000 |
| Vente | 650.000 | | |
| | | Solde | **3.650.000** |
| Total | 5.650.000 | Total | 5.650.000 |

Le compte de bénéfice, qui figure au passif, se note comme :

| Débit | | Crédit | |
|---|---|---|---|
| Administration fiscale | 50.000 | Bénéfice brut | 150.000 |
| Solde | **100.000** | | |
| Total | 150.000 | Total | 150.000 |

On remarque dans cet exemple que les emplois sont inscrits à gauche et signifient une augmentation de l'actif ou une diminution du passif. A l'inverse, les sources sont notées à droite et mentionnent soit une augmentation du passif, soit une réduction de l'actif.

Le *solde* du compte est la valeur *définitive* de la rubrique en question telle qu'elle apparaîtra au bilan[2] : ce solde s'inscrit dans la colonne où le total est le plus faible. Le total des deux colonnes (solde compris) doit être identique.

## 2.4 Normalisation de l'information financière

Le besoin de *normalisation* de l'information financière provient du nombre croissant d'intervenants économiques utilisant cette information. La normalisation a deux objectifs : proposer une information plus compréhensible et rendre l'information plus comparable.

La normalisation précède généralement l'obligation faite aux entreprises de communiquer l'information financière par le biais d'un *journal officiel* ou d'autres techniques plus élaborées apparentées principalement à l'informatique [5].

La normalisation de l'information financière est souvent réalisée par des instances nationales. En Belgique, cette fonction est remplie par l'Etat[3]. Soulignons le fait qu'une uniformisation européenne est en cours et se concrétise au moyen de *directives* à partir

[2]Cfr. exemple du paragraphe 2.3.3

[3]Dans les pays anglo-saxons, ce rôle est joué par des associations professionnelles.

desquelles chaque état-membre édicte ses propres lois et règlementations afin de satisfaire à celles-ci.

La loi belge du 17 juillet 1975 définit un certain nombre de principes de base en matière de comptabilité des entreprises et des comptes annuels. Elle prévoit notamment les obligations suivantes :

- tenir une comptabilité complète (celle-ci couvre les opérations, droits, obligations, avoirs, dettes et engagements de toute nature),
- utiliser le Plan Comptable Mininum Normalisé (P.C.M.N.) en ce qui concerne la présentation et la teneur de l'information comptable.

## Le plan comptable minimum normalisé

Le P.C.M.N. est un plan de travail constituant un guide à suivre par les comptables pour faciliter leur tâche. Il est dit *minimum* parce qu'il peut être étendu et complété selon les besoins de l'entreprise.

Ce plan comprend huit divisions ou classes regroupant les comptes de même nature.

- *Classe I*: Comptes de fonds propres et dettes à plus d'un an (réserves, capital, etc.).
- *Classe II*: Actifs immobilisés et frais d'établissement (bâtiments, terrain, outillage, etc.).
- *Classe III*: Stocks et commandes (matières premières, produits finis, etc.).
- *Classe IV*: Créances et dettes à un an au plus (acompte reçu sur commandes, etc.).
- *Classe V*: Valeurs disponibles et placements de trésorerie (titres à revenu fixe, etc.).
- *Classe VI*: Comptes de charges (rémunérations, charges sociales et pensions, impôts, etc.).
- *Classe VII*: Comptes de produits (ventes de marchandises, etc.).
- *Classe 0*: Comptes et engagements hors bilan. L'objectif de cette classe est de rappeler à l'entreprise les engagements qu'elle a contractés envers des tiers et réciproquement.

# Chapitre 3

# Le concept de faillite

A l'instar de la notion d'entreprise, le terme *faillite* se condense difficilement en une proposition unique. C'est pourquoi nous allons dans un premier temps dresser une esquisse juridique et économique de ce phénomène. Ensuite, nous établirons les premières bases de notre diagnostic financier en introduisant le concept de *ratio*.

## 3.1 La faillite d'un point de vue juridique

Il convient de souligner la nuance entre les termes *faillite* et *banqueroute*.

La *faillite* est un ensemble de mesures prises à l'encontre d'un commerçant cessant d'honorer ses engagements. Celles-ci ont pour objectif de protéger les créanciers.

La *banqueroute* est un terme désignant la situation d'un commerçant en faillite qui a commis des fautes. La banqueroute est dite *simple* si le commerçant s'est rendu responsable de négligence. Par contre, si le commerçant s'est rendu coupable de fraude ou de tromperie, on parlera de *banqueroute frauduleuse*.

### 3.1.1 Définition juridique de la faillite

La loi belge du 18 avril 1851 déclare que *"Tout commerçant qui cesse ses paiements et dont le crédit se trouve ébranlé est en état de faillite"* [6]. Ainsi, la déclaration de faillite d'un débiteur ne peut se réaliser sans l'accomplissement de trois conditions de fond.

1. *Etre commerçant*

   Seuls les commerçants et les sociétés commerciales peuvent être déclarés en faillite.

2. *Avoir cessé ses paiements*

   Cette condition concerne le concept de *liquidité* de l'entreprise ou sa capacité à faire face à ses engagements à court terme. Cependant, cette condition est nécessaire mais non suffisante: un problème temporaire de trésorerie peut être résolu à l'aide de crédits extérieurs.

3. *Avoir son crédit ébranlé*

   Cette condition signifie que tous les créanciers ont perdu confiance envers l'entreprise débitrice. Ainsi, les fournisseurs ne livrent plus et les banques n'avancent plus le moindre franc. En d'autres termes, cette condition touche la *solvabilité* de l'entreprise (i.e. sa capacité de faire face à ses engagements à long terme).

### 3.1.2 Forme de la déclaration de faillite

La déclaration de faillite se fait d'une des trois manières suivantes.

1. Le commerçant dépose son bilan. Autrement dit, il va faire l'aveu de sa situation au tribunal de commerce.
2. Le commerçant est assigné en faillite devant le tribunal du commerce par un ou plusieurs créanciers.
3. Le tribunal du commerce prononce la faillite d'office du commerçant.

### 3.1.3 Le concordat judiciaire

Un commerçant de bonne foi ne sachant plus faire face à ses obligations peut solliciter un arrangement à l'amiable ou *concordat*: il demande à ses créanciers un plus grand étalement des échéances de remboursement. Ce concordat doit bien entendu être demandé avant la déclaration de faillite. Si le tribunal du commerce marque son accord, les créanciers sont convoqués en assemblée afin de voter l'application du concordat.

### 3.1.4 Le concordat après faillite

Un failli peut solliciter le droit d'être remis à la direction de son entreprise dans l'objectif d'apurer ses dettes au terme d'un délai donné. La décision d'accorder ou non ce concordat est soumise à l'assemblée des créanciers.

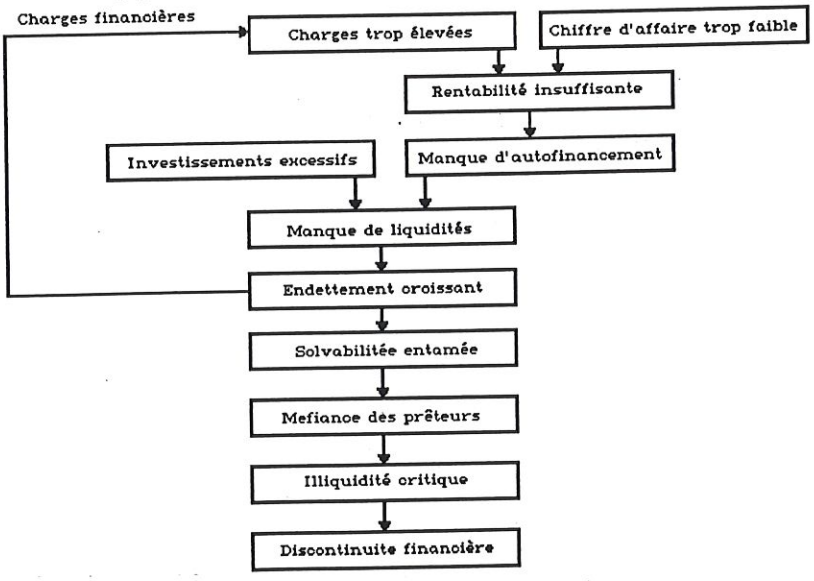
#### Remarque

La législation belge classe les entreprises en difficulté en deux catégories : celles bénéficiant d'un concordat et celles qui ont été déclarées en failite.

## 3.2 La faillite d'un point de vue économique

"D'un point de vue économique, une entreprise en difficulté peut être définie comme une entreprise qui ne *parvient pas à réaliser de manière continue ses objectifs économiques* (maximisation de la *valeur de l'entreprise* aux actionnaires), compte tenu des contraintes sociales et d'environnement (emploi, fiscalité, contribution au développement économique de la région, etc.)." [6].

Le schéma suivant illustre le *failure path* ou l'enchaînement économique menant à la discontinuité financière [1].




Le point d'entrée de ce cercle vicieux réside en une rentabilité insuffisante. Celle-ci est généralement provoquée par un chiffre d'affaire trop faible auquel s'ajoutent de lourdes

charges financières.

De même, le manque de rentabilité occasionne une diminution de l'*autofinancement* (i.e. la capacité de l'entreprise à financer ses dépenses par prélèvement sur ses ressources propres) de l'entreprise.

En conséquence, l'endettement de l'entreprise s'accroît. Pour s'y opposer, elle se tourne vers de nouveaux emprunts. Cependant, ceux-ci sont de plus en plus lourds à supporter, car les créanciers, informés de la situation de l'entreprise, exigent des taux d'intérêt plus élevés afin de couvrir leurs propres risques. Les nouvelles charges financières, alourdies par de nombreux emprunts à haut taux finissent par déboucher sur l'impossibilité de recourir au crédit. Dès lors, l'entreprise cesse ses paiements et remplis les conditions juridiques de faillite que nous avons décrit au paragraphe 3.1.1.

## 3.3 La technique des ratios

Le diagnostic financier que nous allons mettre au point s'appuie sur une technique fondamentale de l'analyse financière : les *ratios*.

Le terme *ratio* se traduit littéralement par *rapport*. Ceux-ci représentent en fait *des rapports entre des rubriques-clés du bilan* [6]. Il s'agit donc de valeurs *relatives* à l'aide desquelles l'analyste financier va pouvoir effectuer des comparaisons entre des entreprises appartenant au même secteur d'activité économique.

Comme le sous-entend la définition que nous venons d'en faire, il est possible de construire un grand nombre de ratio comparant chacun des éléments du bilan. Néanmoins, la plupart de ces valeurs n'auront aucune signification économique. L'art de l'analyste financier consiste donc à n'employer que les ratios les plus judicieux et informatifs.

Remarquons enfin qu'un bon diagnostic financier s'établit de préférence sur un ensemble de ratios et non sur un rapport isolé. Cependant, un trop grand nombre de ratio créerait de la confusion.

La théorie de l'analyse financière répartit les ratios en quatres grandes catégories selon leur signification économique que nous exposons à présent.

### 3.3.1 Ratios de liquidité

Comme nous l'avons vu au paragraphe 3.1.1, le concept de *liquidité* concerne la capacité de l'entreprise *à faire face à ses dettes à court terme.* Rappelons que l'*illiquidité* constitue la deuxième condition de fond à la définition juridique de faillite d'une entreprise. Citons, à titre d'exemple, quelques ratios de liquidité les plus usités.

**Ratio de liquidité immédiate**

$$\text{liquidité immédiate} = \frac{\text{valeurs disponibles}}{\text{dettes à court terme}}$$

Ce rapport indique si l'entreprise est capable de faire face à des besoins de trésorerie à très court terme. Autrement dit, cette valeur indique la part des dettes à court terme pouvant être remboursée dans l'immédiat. Ce ratio constitue un excellent indicateur de l'illiquidité d'une entreprise proche de la faillite.

**Ratio de rotation des stocks**

$$\text{rotation des stocks} = \frac{\text{prix de revient des ventes}}{\text{stocks et commandes en cours d'exécution}}$$

Ce ratio classique concerne la liquidité des stocks. Sa fonction est de comparer le montant des stocks par rapport aux ventes annuelles. Une valeur élevée indique que les stocks se vendent bien, ce qui signifie un apport de liquidité. A l'inverse, une valeur faible montre que l'entreprise a du mal à écouler sa production, ce qui provoque une diminution de sa liquidité.

### 3.3.2 Ratios de solvabilité

La *solvabilité* de l'entreprise mesure son aptitude à remplir l'ensemble de ses engagements financiers. Ceux-ci comprennent le remboursement des dettes et le paiement des intérêts. Comme nous l'avons déjà spécifié, l'insolvabilité d'une entreprise constitue la troisième condition à la définition de la faillite.

**Le ratio d'endettement**

$$\text{ratio d'endettement} = \frac{\text{total des dettes}}{\text{total de l'actif}}.100\ \%$$

Ce ratio indique simplement le *degré d'endettement de l'entreprise.* Il constitue un excellent indicateur du *risque financier de l'entreprise.* Plus celui-ci est élevé, plus les créanciers

voient leur risque augmenter. De cette manière, une entreprise en difficulté finit tôt ou tard à ne plus être en position de recourir à un emprunt.

### 3.3.3 Ratios de rentabilité

La *rentabilité* de l'entreprise mesure sa capacité *à réaliser du bénéfice.* Ainsi que nous l'avons souligné en 3.2, une rentabilité faible constitue le point de départ du *failure path.*

A l'instar des catégories précédentes, il existe de nombreux ratios permettant d'évaluer la rentabilité d'une entreprise.

**Ratio de rentabilité des ventes**

$$\text{ratio de rentabilité des ventes} = \frac{\text{résultat d'exploitation}}{\text{ventes}}.100\%$$

Ce rapport mesure le bénéfice réalisé par rapport aux ventes.

**Ratio de rentabilité de l'actif total**

$$\text{ratio de rentabilité de l'actif total} = \frac{\text{résultat avant charges d'intérêts et impôts}}{\text{actif total}}.100\%$$

Ce ratio mesure le bénéfice réalisé par rapport à l'ensemble des actifs qui l'ont généré.

### 3.3.4 Ratios de valeur ajoutée

Les ratios de *valeur ajoutée* reflètent *la contribution de l'entreprise à la création globale de richesse.* Ceci constitue une mesure pertinente de la performance économique de l'entreprise.

**Taux de la valeur ajoutée**

$$\text{taux de la valeur ajoutée} = \frac{\text{valeur ajoutée}}{\text{valeur de la production}}.100\%$$

Ce ratio doit être considéré avec attention. Les entreprises du secteur de première transformation (le secteur de la sidérurgie, etc.) disposent généralement d'un taux assez faible car elles ne fabriquent que des produits semi-finis et sont souvent soumises à une concurrence internationale écrasante.

# Partie II

# Analyse discriminante

# Motivations

L'analyse discriminante constitue un fondement capital à l'élaboration d'un modèle de diagnostic financier. Il est donc essentiel de détailler les principes sur lesquels repose cette théorie. Par ailleurs, nous nous attacherons également au développement systématique des modèles statistiques qui y sont associés.

Le premier chapitre se veut être une introduction générale au concept de classification. Nous y effectuerons une première formalisation de notre problème.

Dans le second chapitre, nous définirons les principes mathématiques de l'analyse discriminante selon les approches paramétrique et non paramétrique. Nous commenterons également les méthodes d'estimation du taux d'erreur réalisé.

Le troisième chapitre fera l'objet d'une étude plus approfondie des processus de Poisson homogène et non homogène.

# Chapitre 1

# La classification

## 1.1 Introduction

L'homme passe une majeure partie de son temps à reconnaître et classer les éléments de son environnement. La théorie de la classification n'est en fait qu'une formalisation de ce processus de pensée. Par ailleurs, elle dispose d'un champ d'application assez étendu : les problèmes de reconnaissance de sons, de formes ou de caractères, la classification des éléments d'une image (télédétection, biologie), la détection de faux billets, etc. Les techniques de classification se répartissent suivant deux catégories, à savoir l'analyse de groupe et l'analyse discriminante.

### 1.1.1 L'analyse de groupe

L'analyse de groupe consiste à *partitionner* une population donnée en un certain nombre de *classes* ou *groupes*. Ce classement s'effectue sans qu'aucune information préalable ne soit disponible en ce qui concerne les propriétés et/ou caractéristiques des groupements résultants [1]. On parle aussi de classification *non supervisée*.

[1] Cette méthode est généralement employée dans une optique d'exploration préliminaire.

### 1.1.2 L'analyse discriminante

L'analyse discriminante est fondée sur l'existence d'un échantillon de données que l'on appelle *training set* ou *base d'entraînement.* Celle-ci est composée d'individus dont on connaît la classe d'appartenance. Le but de l'analyse discriminante est de mettre au point une *règle de classification* pour des individus dont la classe n'est pas encore connue, en s'appuyant sur l'information contenue dans la base d'entraînement. La classification est alors dite *supervisée.*

## 1.2 Formulation du problème étudié

Dans le cadre de la détection précoce de faillite d'entreprise, le problème de classification peut se formuler de la façon suivante.

*Trouver, sur base des échantillons disponibles, un critère de classification permettant de déterminer la classe d'appartenance d'une entreprise donnée.*

Les entreprises économiquement saines constituent une classe, l'autre classe étant formée par les entreprises présentant une défaillance à court terme.

#### Conventions

Chacun des ensembles de données de départ est formé de $n_1$ entreprises économiquement saines et de $n_2$ entreprises en faillite. De plus, nous disposons de $k$ ratios pour chaque entreprise.

Une entreprise est notée

$$x_{ji} = \begin{pmatrix} x_{ji1} \\ \vdots \\ x_{jik} \end{pmatrix}.$$

où $j$ indique la classe d'appartenance de l'entreprise et $i$ son numéro ($1 \leq j \leq n_j$).

# Chapitre 2

# L'analyse discriminante

## 2.1 Critère de ressemblance

L'affectation d'une nouvelle entreprise à une classe s'effectue suivant un *critère de ressemblance* entre son vecteur d'observation et l'ensemble des vecteurs d'observation donnés par la base d'entraînement. La nature de ce critère est déterminée par la méthode employée.

### 2.1.1 Méthodes déterministes

Chaque entreprise $x_{ji}$ peut être représentée par un point dans l'espace usuel à $k$ dimensions $\mathbb{R}^k$. La base d'entraînement sera alors constituée par un nuage de points.

Le critère de ressemblance entre deux entreprises peut se concevoir comme la *distance euclidienne* entre les points de $\mathbb{R}^k$ correspondants. Plus celle-ci est petite, plus leur ressemblance est grande.

### 2.1.2 Méthodes statistiques

Les méthodes statistiques supposent que le vecteur d'observation est un *vecteur aléatoire* possédant une certaine *densité de probabilité*. Le critère de ressemblance est choisi de sorte qu'il minimise le *risque de mauvais classement* de l'entreprise. Néanmoins, le calcul de ce risque nécessite la connaissance de la densité de probabilité en question.

## 2.2 Règles de discrimination

Nous nous consacrerons, dans le cadre de ce mémoire, à des méthodes statistiques. Nous supposons que les vecteurs d'observations $x_{ji}$ sont des vecteurs aléatoires multivariés continus.

Notons

$$X_1 = \{x_{ji} \dagger j = 1\,, i = 1, \ldots, n_1\}$$

et

$$X_2 = \{x_{ji} \dagger j = 2\,, i = 1, \ldots, n_2\}$$

Nous admettons également que l'échantillon $X_1$ des entreprises saines suit une fonction de densité $f_1$ tandis que l'échantillon $X_2$ des entreprises en faillite possède une fonction de densité $f_2$.

Toute nouvelle entreprise $y$ sera affectée à la première classe si

$$f_1(y) \geq f_2(y)\,.$$

Cette règle de discrimination suit une approche basée sur le maximum de vraisemblance.

D'autre part, on peut employer une approche *bayesienne*. Celle-ci nécessite la connaissance des probabilités *a priori* qu'une nouvelle entreprise $y$ appartienne à une classe donnée. Le calcul des probabilités *a posteriori* s'effectue alors de la manière suivante

$$P_{\text{post}}(y \in \text{Classe 1}) = \frac{f_1(y)}{f_2(y)} P_{\text{prior}}(y \in \text{Classe 1}).$$

La règle de discrimination découlant de cette approche consiste à affecter $y$ à la classe qui maximise la probabilité a posteriori.

Cependant, le problème est loin d'être résolu car ces deux approches recquièrent la connaissance des fonctions de densités $f_1$ et $f_2$. Etant donné que nous ne disposons d'aucune information les concernant, nous nous trouvons dans l'obligation de les estimer à partir de la base d'entraînement.

## 2.3 Estimation paramétrique

L'estimation paramétrique se fonde sur l'hypothèse que les fonctions de densité sont issues d'une famille paramétrique spécifique. Le problème se réduit donc à déterminer les paramètres en question. Nous allons aborder deux méthodes de classement utilisant une telle approche.

### 2.3.1 Règle linéaire de Fisher

La règle linéaire de Fisher[1] a pour hypothèse sous-jacente que les fonctions de densité $f_1$ et $f_2$ sont des réalisations de lois normales multivariées de moyennes respectives $\mu_1$ et $\mu_2$ et de matrice de variance-covariance $\Sigma$ commune.

Les entreprises formant la base d'entraînement étant représentées par un nuage de points dans $\mathbb{R}^k$, la règle de Fisher consiste à séparer ce nuage par un *hyperplan discriminant*. Celui-ci doit permettre de distinguer les sous-nuages de points correspondant à chaque classe. Cet hyperplan est déterminé à l'aide d'une fonction dite *fonction de score* définie par

$$f(y) = (\bar{x}_1 - \bar{x}_2)' S^{-1} \left( y - \frac{\bar{x}_1 + \bar{x}_2}{2} \right) ,$$

où

- $\bar{x}_1$ est le vecteur moyenne de $\mathbb{R}^k$ des $n_1$ entreprises non défaillantes,
- $\bar{x}_2$ est le vecteur moyenne de $\mathbb{R}^k$ des $n_2$ entreprises défaillantes,
- $S$ est la matrice de variance-covariance empirique.

L'équation de l'hyperplan est donnée par $f(x) = 0$. La position d'un point par rapport à cet hyperplan détermine la classe de l'entreprise correspondant. La règle de discrimination devient

$$\text{Affecter une nouvelle entreprise } y \text{ à la classe } \begin{cases} 1 & \text{si } f(y) \geq 0 , \\ 0 & \text{si } f(y) < 0 . \end{cases}$$

[1] Cette méthode est couramment employée par les analystes financiers. On la retrouve notamment dans des logiciels de traitement statistique de données ou dans des librairies de routines mathématiques (SPSS, SAS, IMSL, ...).

### 2.3.2 Règle quadratique

Le principe de la règle quadratique est similaire à celui de la règle linéaire de Fisher [3]. Cependant, les fonctions de densité sont considérées comme étant des réalisations de lois normales multivariées de matrices de variance-covariances distinctes.

La fonction de score est ici

$$f(y) = \frac{(y - \bar{x}_1)' S_1^{-1} (y - \bar{x}_1) - (y - \bar{x}_2)' S_2^{-1} (y - \bar{x}_2)}{\ln\left(\frac{p_1}{p_2}\right) \ln\left(\frac{S_1}{S_2}\right)},$$

où

- $S_j$ est la matrice de variance-covariance empirique de la classe $j$,
- $p_j$ est la probabilité a priori qu'une entreprise appartienne à la classe $j$.

La surface déterminée par l'équation $f(x) = 0$ est ici une hypersurface quadratique.

Cette technique est moins utilisée que la règle de Fisher. Ceci est dû en partie aux facilités de calcul offertes par celle-ci.

## 2.4 Estimation non paramétrique

Comme son nom le sous-entend, l'estimation non paramétrique ne pose aucune hypothèse sur les fonctions de densités $f_1$ et $f_2$. Celle-ci seront estimées à l'aide de la base d'entraînement grâce à des *estimateurs non paramétriques.*

Nous allons passer en revue quelques estimateurs non paramétriques les plus employés. Dans un souci de compréhension, nous illustrerons chacun d'entre eux à l'aide de l'exemple suivant, extrait de [2].

| *Durée des éruptions (en minutes) de 107 éruptions du geyser "Old Faithful"* | | | | | | |
|---|---|---|---|---|---|---|
| 4.37 | 3.87 | 4.00 | 4.03 | 3.50 | 4.08 | 2.25 |
| 4.70 | 1.73 | 4.93 | 1.73 | 4.62 | 3.43 | 4.25 |
| 1.68 | 3.92 | 3.68 | 3.10 | 4.03 | 1.77 | 4.08 |
| 1.75 | 3.20 | 1.85 | 4.62 | 1.97 | 4.50 | 3.92 |
| 4.35 | 2.33 | 3.83 | 1.88 | 4.60 | 1.80 | 4.73 |
| 1.77 | 4.57 | 1.85 | 3.52 | 4.00 | 3.70 | 3.72 |
| 4.25 | 3.58 | 3.80 | 3.77 | 3.75 | 2.50 | 4.50 |
| 4.10 | 3.70 | 3.80 | 3.43 | 4.00 | 2.27 | 4.40 |
| 4.05 | 4.25 | 3.33 | 2.00 | 4.33 | 2.93 | 4.58 |
| 1.90 | 3.58 | 3.73 | 3.73 | 1.82 | 4.63 | 3.50 |
| 4.00 | 3.67 | 1.67 | 4.60 | 1.67 | 4.00 | 1.80 |
| 4.42 | 1.90 | 4.63 | 2.93 | 3.50 | 1.97 | 4.28 |
| 1.83 | 4.13 | 1.83 | 4.65 | 4.20 | 3.93 | 4.33 |
| 1.83 | 4.53 | 2.03 | 4.18 | 4.43 | 4.07 | 4.13 |
| 3.95 | 4.10 | 2.72 | 4.58 | 1.90 | 4.50 | 1.95 |
| 4.83 | 4.12 | | | | | |

Nous noterons ces observations par $x_1, \ldots, x_n$ ($n = 107$). L'estimation de la fonction de densité sous-jacente à ces données sera notée par $\hat{f}$.

### 2.4.1 L'histogramme

L'*histogramme* est un estimateur de densité classique fréquemment utilisé. Etant donné une origine $x_0$ et une largeur de fenêtre $h$, nous définissons les fenêtres de l'histogramme comme étant les intervalles

$$[x_0 + mh, x_0 + (m+1)h) \ , \ \forall m \in \mathbb{Z}.$$

Remarquons que les intervalles sont par convention fermés à gauche et ouverts à droite.

L'histogramme est définit par

$$\hat{f}(x) = \frac{1}{nh}(\text{nombre de } x_i \text{ dans la même fenêtre que } x) .$$

Si nous appliquons cet estimateur à notre exemple, nous obtenons

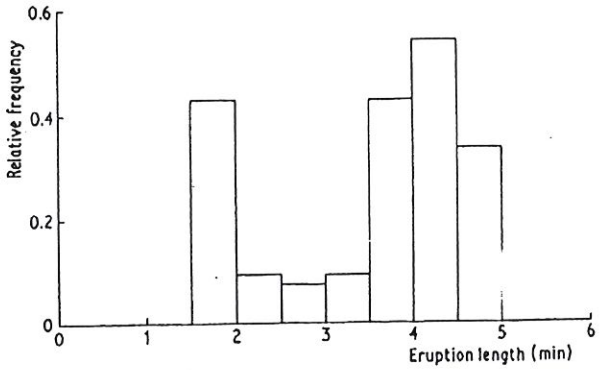



Fig. 2.1 *Histograms of eruption lengths of Old Faithful geyser.*

L'histogramme peut facilement être généralisé à des fenêtres de largeur variable. L'estimation résultante est

$$\hat{f}(x) = \frac{1}{n}\frac{(\text{nombre de } x_i \text{ dans la même fenêtre que } x)}{(\text{largeur de la fenêtre contenant } x)} .$$

### 2.4.2 L'estimateur naïf

Soit $f$ la densité de variable aléatoire $X$. De par la définition de la densité de probabilité, nous avons que

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P\left(x - h < X < x + h\right) .$$

Si nous connaissons $h$, nous pouvons estimer $P(x - h < X < x + h)$ en calculant la proportion de l'échantillon se situant dans l'intervalle $(x - h, x + h)$. Dès lors, l'estimation de la densité s'effectue comme

$$\hat{f}(x) = \frac{1}{2nh}(\text{nombre de } x_i \in (x - h, x + h)) ,$$

où $h$ est un nombre proche de 0. Cet estimateur est appellé *estimateur naïf.*

Nous pouvons en simplifier l'écriture en définissant une *fonction de poids* $w$ telle que

$$w(v) = \begin{cases} \frac{1}{2} & \text{si } \mid v \mid < 1 , \\ 0 & \text{sinon.} \end{cases}$$

L'estimateur naïf devient

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} w\left(\frac{x - x_i}{h}\right) .$$

D'un point de vue intuitif, l'estimateur naïf place une "boîte" de largeur $2h$ et de poids $\frac{1}{2nh}$ sur chaque observation. L'estimateur naïf permet ainsi de construire un histogramme où chaque point est le centre d'un intervalle dont la largeur de fenêtre est $2h$. Plus le paramètre $h$ est petit, plus on donne du "poids" à chaque observation.

Si nous appliquons cet estimateur de densité à notre exemple, nous obtenons

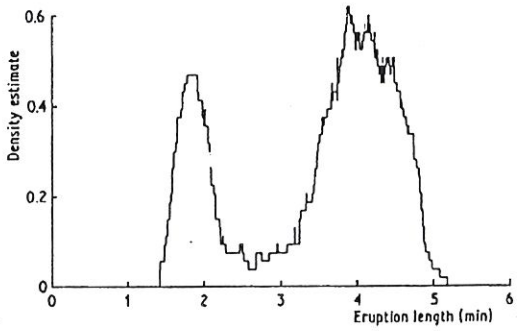



Fig. 2.3 *Naive estimate constructed from Old Faithful geyser data, $h = 0.25$.*

### 2.4.3 Estimateur du noyau

L'*estimateur du noyau* est une généralisation de l'estimateur naïf. La fonction de poids $w(v)$ est remplacée par une fonction noyau $K(v)$. Cette fonction est souvent symétrique, à valeurs positives et vérifie

$$\int_{-\infty}^{\infty} K(x)dx = 1 .$$

La forme de l'estimateur du noyau est donc

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) .$$

Le rôle de l'estimateur naïf était une somme de "boîtes" centrées en chaque observation. Par analogie, l'estimateur du noyau somme des "bosses" dont la forme est déterminéee par la nature de la fonction noyau K tandis que la largeur de ces dernières est définie par le paramètre $h$.

Le rôle du paramètre $h$, appellé *largeur de fenêtre* mais également *paramètre de lissage* ou *bandwith*, est déterminant. Plus cette valeur tend vers 0, plus le poids des bosses se concentre sur les observations et plus la courbe est déchiquetée. En d'autres termes, tous les détails de la distribution apparaissent. A l'inverse, un $h$ élevé engendre une courbe plus lisse et unimodale.

Pour notre exemple, nous obtenons pour différentes valeurs de $h$

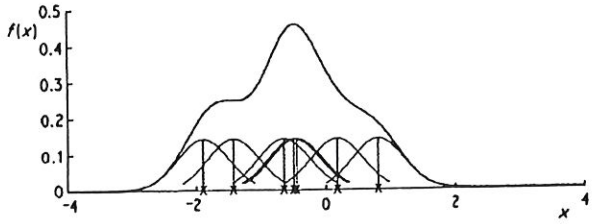



Fig. 2.4 *Kernel estimate showing individual kernels. Window width 0.4.*

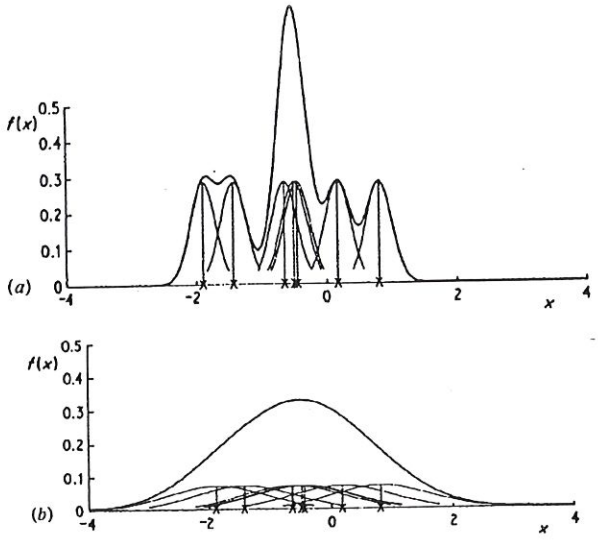



Fig. 2.5 *Kernel estimates showing individual kernels. Window widths: (a) 0.2; (b) 0.8.*

Grâce aux propriétés de la fonction noyau $K$, nous avons que l'estimateur du noyau est bel et bien une fonction de densité. En effet,

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) dx &= \frac{1}{nh} \sum_{i=1}^{n} \int_{-\infty}^{\infty} K\left(\frac{x - x_i}{h}\right) dx \\
&= \frac{1}{nh} \sum_{i=1}^{n} \int_{-\infty}^{\infty} hK(y)\, dy \\
&= \sum_{i=1}^{n} \frac{1}{n} = 1 .
\end{aligned}$$

Remarquons enfin que l'estimateur du noyau hérite des propriétés de continuité et de différentiabilité de la fonction noyau $K$.

### 2.4.4 Estimateur du noyau multivarié

La généralisation au cas multidimensionnel se réalise facilement grâce à l'*estimateur du noyau multivarié*. Si on note par $k$ la dimension de l'espace des observations, cet estimateur prend la forme

$$\hat{f}(x) = \frac{1}{nh^k} \sum_{i=1}^{n} \left( \frac{x - x_i}{h} \right) ,$$

où $x$ et $x_i$ représentent des vecteurs à $k$ composantes.

La fonction noyau que nous emploierons est dite *rectangulaire*. Elle se définit par

$$K(v) = \begin{cases} \frac{1}{2^k} & \text{si } \|v\|_\infty < 1 \\ 0 & \text{sinon.} \end{cases}$$

Notons que $K$ jouit des mêmes propriétés que dans le cas univarié.

L'estimateur du noyau multivarié présente plusieurs avantages. Plusieurs auteurs la considèrent commme étant une des méthodes les plus performantes, surtout quand la taille des échantillons de données est relativement petite. De plus, elle est à la fois simple et intuitive.

Néanmoins, deux inconvénients sont inhérents à cette technique. D'une part, l'utilisation d'un paramètre de lissage unique peut se révéler être source de mauvais résultats (nous reviendrons plus tard sur ce point). D'autre part, les coûts de stockage réalisés sont relativement élevés. Cependant, le matériel informatique performant dont nous disposons permet en général de contourner ce handicap.

## 2.5 Estimation du taux d'erreur réalisé

La performance d'une règle de classement se mesure grâce au taux d'erreur réalisé. Le calcul de ce dernier mérite toute notre attention.

D'un point de vue intuitif, le taux d'erreur se mesure en comparant les résultats de la classification par rapport à la situation réelle. Cependant, cette dernière n'est généralement pas disponible: dans notre cas, nous ne disposons que de quelques échantillons. Nous nous retrouvons une fois de plus contraints à recourir à une *estimation* fondée sur la base d'entraînement.

Plusieurs techniques sont envisageables.

### 2.5.1 La méthode de resubstition

Une manière naturelle de procéder consiste à estimer le taux d'erreur engendré par le classificateur en évaluant le rapport formé par le nombre d'individus mal classés sur le nombre total d'individus de l'échantillon d'entraînement.

Cette démarche, appellée *méthode de resubstitution*, présente un inconvénient fondamental : elle se révèle *sur-optimiste* car elle favorise essentiellement les données d'entraînement. Il est donc dangereux de tenir compte de cette estimation biaisée. Malgré tout, ce procédé est quelques fois employé lorsque la taille de la base d'entraînement est très élevée. Le biais se trouve en effet réduit si le nombre d'individus traité est important.

### 2.5.2 La méthode "Leaving-One-Out"

Le biais résultant de la méthode par resubstitution peut être évité en divisant la base d'entraînement en deux sous-ensembles dont le premier servirait à construire le classificateur. Le second permettrait d'estimer le taux d'erreur engendré. Ainsi, il y a indépendance entre l'échantillon d'entraînement et l'échantillon sur lequel est réalisée l'estimation.

Cependant, ce procédé "sacrifie" la moitié des données disponibles. Cette diminution de l'information risque d'altérer la qualité du classificateur. On peut malgré tout conserver *le classificateur final* (c'est-à-dire celui basé sur la totalité des données) en considérant que le taux d'erreur engendré par ce dernier a été approximé par le classificateur fondé sur la partition de la base d'entraînement en deux sous-ensembles.

Par ailleurs, l'idée de division peut être améliorée en augmentant le nombre de sous-ensembles. Ainsi, en scindant la base d'entraînement en $p$ sous-ensembles égaux, il est possible de construire, pour chaque sous-ensemble, un classificateur à partir des $p-1$ sous-ensembles restants. Le taux d'erreur se conçoit alors comme la moyenne des $p$ estimations résultantes. De cette manière, les estimations se réalisent toujours sur des sous-échantillons indépendants des points d'entraînement. De plus, la qualité d'approximation du taux d'erreur du classificateur final est beaucoup plus élevée car la taille du sous-échantillon d'entraînement est proche de la base d'entraînement initiale. Ce processus peut être étendu à la limite. En d'autres termes, on considérera $n$ sous-ensembles qui sont en fait les $n$ singletons formant la base d'entraînement. Ainsi, chaque individu peut être classé sur base des $n-1$ individus restants. La proportion de point mal classés nous donnera une excellente approximation du taux d'erreur réalisé par le classificateur final.

# Chapitre 3

# Règle de classification basée sur un processus de Poisson non homogène

## 3.1 Introduction

Les recherches effectuées par le laboratoire GEOSATEL en matière de télédétection consistent à produire une classification complète des éléments contenus dans une photographie d'un terrain, prise par satellite ou par avion et renvoyée sous forme d'image digitalisée. Les *pixels* [1] constituant cette image sont munis de coordonnées spatiales (déterminant leur position dans l'image) et de trois valeurs spectrales qui fixent leur couleur suivant le système de codification RGB *(Reg, Green, Blue)*.

L'objectif étant d'assigner à chaque pixel une classe d'appartenance (champs de blé, forêts, habitation, route), J-P. RASSON a proposé une méthode de classification basée sur l'hypothèse que les données à classer sont la réalisation d'un *processus de Poisson non homogène* [4].

Par analogie, nous supposerons que les données relatives aux entreprises dont nous disposons sont également une réalisation d'un processus de Poisson non homogène.

Dans ce chapitre, nous décrirons le processus de Poisson non homogène et la règle de classification résultante. Mais pour cela, il apparaît nécessaire de développer dans un premier temps le *processus de Poisson homogène.*

[1] Terme anglo-saxon provenant de la contraction de *PICture* et *ELement.*

## 3.2 Processus de Poisson homogène

Notons par $x_1, \ldots, x_n$ les vecteurs représentant les $n$ entreprises contenues dans la base d'entraînement. Ces $n$ vecteurs possèdent $k$ composantes qui sont les $k$ ratios provenant de l'information comptable.

Un échantillon est dit être une réalisation d'un *processus de Poisson homogène* (ou *stationnaire*) sur un domaine $D$ si les hypothèses suivantes sont vérifiées :

1. La variable aléatoire qui compte le nombre de points dans le domaine $D$ suit une distribution de *Poisson* dont le paramètre est la mesure de Lebesgue[2] du domaine $D$ (notée $\mu(D)$). En d'autres termes, on suppose que le nombre moyen de points contenu dans $D$ est proportionnel à $\mu(D)$.
2. Les variables aléatoires comptant le nombre de points dans des domaines disjoints sont indépendantes.

On déduit de ces hypothèses que, conditionnellement au fait que $n$ points aléatoires engendrés par ce processus appartiennent au domaine $D$, ceux-ci sont distribués *uniformément* et *indépendamment*.

Nous considérerons dans un premier temps que les observations relatives aux $n$ entreprises sont générées par un processus de Poisson homogène et distribuées uniformément et indépendamment dans un domaine $D$, lequel est la réunion de $p$ domaines disjoints $(D_j)_{1 \leq j \leq p}$ que nous désirons déterminer.

### 3.2.1 Solution du maximum de vraisemblance

Soit $\mathbf{x}$ le vecteur échantillon $(x_1, \ldots, x_n)$. Rappellons que $\mathbb{1}_D(y)$ est la fonction indicatrice du domaine $D$.

$$\mathbb{1}_D(y) = \begin{cases} 1 & \text{si } y \in D \\ 0 & \text{sinon.} \end{cases}$$

[2]La mesure de Lebesgue représente la distance usuelle sur la droite, la surface sur le plan, le volume dans l'espace à 3 dimensions, l'hypervolume dans les espaces de dimension supérieure.

La fonction de vraisemblance s'écrit alors

$$F_D(\mathbf{x}) = \frac{1}{(\mu(D))^n} \prod_{i=1}^{n} \mathbb{1}_D(x_i) ,$$

où $\mu(D)$ est la somme des mesure des $p$ sous-ensembles $(D_j)_{1 \leq j \leq p}$.

L'inconnue cherchée est le domaine $D$, contenant les $n$ points, pour lequel la vraisemblance est maximale : cela revient à chercher le domaine $D$ de mesure de Lebesgue $\mu(D)$ minimale.

Cependant, le problème n'est pas bien posé car de nombreuses solutions triviales existent. Il suffit de déterminer $p$ ensembles $D_j$ contenant tous les points et dont la mesure est nulle. Considérons par exemple que les $p$ sous-ensembles sont des segments reliant les $n$ points. Puisque la mesure d'un segment est nulle, nous avons déterminé un domaine $D$ répondant à nos exigences. Pour remédier à cela, il nous faut imposer une condition supplémentaire, à savoir la convexité des $p$ sous-ensembles $D_j$.

Considérons les partitions de l'ensemble des points en $p$ groupes auxquels nous associons $p$ ensembles convexes disjoints les contenant. Pour chaque partition, la vraisemblance a un maximum local : les *enveloppes convexes*[3] des $p$ groupes. Le maximum global sera atteint avec la partition pour laquelle la somme des mesures de Lebesgue des $p$ enveloppes convexes correspondantes est minimale.

A titre d'exemple, si nous travaillons dans $\mathbb{R}^2$, nous devons déterminer les $p$ groupes de points tels que la somme des aires des enveloppes convexes disjointes associées est minimale.

### 3.2.2 Règle de discrimination résultante

Etant donné $p$ groupes $G_1, \ldots, G_p$ formant l'échantillon, le problème consiste à affecter une nouvelle observation $x$ à l'un de ces groupes. Selon la règle bayesienne que nous avons décrite au paragraphe 2.2, l'individu $x$ sera assigné à la classe $j$ qui maximise la probabilité a posteriori que $x$ appartienne au groupe $G_j$. Le calcul de ces probabilités est donné par

$$p_j f(x \mid G_j) ,$$

où $p_j$ est la probabilité a priori d'appartenir au groupe $G_j$ et $f(\cdot \mid G_j)$ représente la fonction de densité des points appartennant au groupe $G_j$.

[3] L'envellope convexe de $l$ points est le plus petit convexe contenant ces $l$ points.

L'hypothèse que l'échantillon est une réalisation d'un processus de Poisson homogène implique que

$$f(x \mid G_j) = \frac{\mathbb{1}_{D_j}(x)}{\mu(D_j)}$$

où $\mu(\cdot)$ est la mesure de Lesbesgue dans l'espace usuel $\mathbb{R}^k$ et $D_j$ désigne le support convexe du $j^{\text{ème}}$ groupe. Rappellons que $D_j$ est estimé par l'enveloppe convexe $H_j$ du $j^{\text{ème}}$ groupe. D'un autre côté, les probabilités a priori $p_j$ se définissent comme

$$p_j = \frac{\mu(D_j)}{\sum_{i=1}^{p} \mu(D_i)} .$$

Si on assigne une nouvelle observation $x$ au $j^{\text{ème}}$ groupe, l'estimation du support convexe $D_j$ est réalisée par l'enveloppe convexe du groupe $G_j$ auquel on adjoint l'observation $x$. En d'autres termes, $D_j$ est estimé par

$$H(G_j \cup \{x\}) \overset{\text{not}}{=} H_j(x) .$$

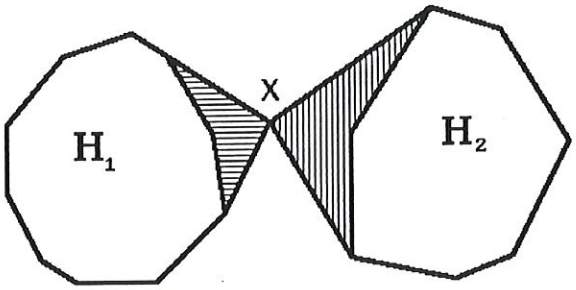
On définit alors par $S_j(x)$ l'*hypervolume ajouté à l'enveloppe convexe du $j^{\text{ème}}$ groupe augmenté de l'observation $x$*. Celui-ci se calcule comme

$$S_j(x) = \mu(H_j(x) \setminus H_j) = \int_{H_j(x)\setminus H_j} \mathrm{dy} .$$

La règle de classification résultant de toutes ces considérations est

*"Affecter une nouvelle observation $x$ à la classe $j$ minimisant $S_k(x)$."*

Le schéma suivant illustre l'affectation d'un point $x$ entre deux groupes donnés suivant la règle que nous venons d'énoncer.




Ce modèle présente néanmoins un inconvénient majeur : que se passe-t-il si l'observation $x$ appartient à l'intersection de plusieurs enveloppes convexes ? Pour traiter cette situation, nous devons envisager une nouvelle hypothèse, à savoir le *processus de Poisson non homogène.*

### 3.2.3 Processus de Poisson non homogène

Le *processus de Poisson non homogène* ou *non stationnaire* est une généralisation du processus de Poisson homogène, consistant en *l'abandon de l'hypothèse de distribution uniforme*. Ainsi, nous considérerons que les distributions conditionnelles au fait que $n$ points aléatoires engendrés par ce processus appartiennent à un domaine $D$ sont absolument *quelconques*. Donc, les fonctions de densités $f(\cdot \mid G_j)$ $(j = 1, \ldots, p)$ sont quelconques sur les domaines convexes $D_j$ de sorte que les données puissent provenir de n'importe quelle distribution non paramétrique.

Le critère de classement ne dépend plus de la mesure de Lebesgue d'un domaine, mais de *l'intensité* integrée sur ce domaine.

### 3.2.4 Règle de classification résultante

Soit $q_j(\cdot)$ l'intensité inconnue associée au groupe $G_j$ satisfaisant l'équivalence

$$q_j(x) > 0 \iff x \in D_j(x)\,.$$

Notons par $f_j(\cdot)$ la fonction de densité $f(\cdot \mid G_j)$.

Nous supposons que chaque individu $x$ est distribué sur $D = \bigcup_{j=1}^{p} D_j$ avec la densité

$$f(x) = \sum_{j=1}^{p} p_j f_j(x)\,,$$

où

$$\begin{aligned} p_j &= \frac{\int_{D_j} q_j(y)dy}{\sum_{i=1}^{p} \int_{D_i} q_i(y)dy}\,, \\ f_j(x) &= \frac{q_j(x)}{\int_{D_j} q_j(y)dy}\,. \end{aligned}$$

Comme auparavant, les supports convexes $D_j$ sont estimés par les enveloppes convexes $H_j$ de la base d'entraînement.

En notant

$$\begin{aligned} S &= \sum_{i=1}^{p} \int_{H_i} q_i(y)dy \\ S_j(x) &= \int_{H_j(x)\setminus H_j} q_j(y)dy\,, \end{aligned}$$

la règle bayesienne de classification devient

*Assigner la nouvelle observation $x$ à la classe $j$ minimisant*

$$p_j f_j(x) = \frac{q_j(x)}{S + S_j(x)}$$

Ainsi, lorsque les intensités $q_j(x)$ ne dépendent pas de $j$, la règle de classification traite les trois cas de figure suivants.

1. *Le point $x$ appartient à une seule enveloppe convexe $H_j$.*

   Dans ce cas, $x$ est affecté à la classe $j$.

2. *Le point $x$ n'est contenu dans aucune enveloppe convexe.*

   $x$ est assigné à la classe $j$ pour laquelle l'intensité ajoutée $S_k(x)$ est minimale.

3. *Le point $x$ se situe dans l'intersection de plusieurs enveloppes convexes.*

   $x$ est affecté à la classe $j$ pour laquelle l'intensité $q_j(x)$ est maximale.

### 3.2.5 Estimateurs d'intensités

Chaque intensité $q_j(\cdot)$ est directement liée à la densité $f_j(\cdot)$ par la relation

$$f_j(x) = \frac{q_j(x)}{n_j},$$

où $n_j$ indique le nombre de points d'entraînement dont la classe d'appartenance est $j$.

Nous pouvons en déduire que l'estimation de l'intensité revient à un problème d'estimation de densité. Or, dans le chapitre précédent, nous avons passé en revue différents estimateurs de densité non paramétriques, dont l'estimateur du noyau qui semblait assez performant. C'est pourquoi l'emploi de ce dernier nous semble approprié.

L'estimation de $f_j(\cdot)$ se réalise au moyen de

$$\hat{f}_j(x) = \frac{1}{n_j(2h_j)^k} \sum_{i=1}^{n_j} K\left(\frac{x - x_{ji}}{h_j}\right),$$

où

- $k$ représente la dimension de l'espace de travail,
- $x_{j1}, \ldots, x_{jn_j}$ sont les données d'entraînement relatives à la classe $j$,
- $h_j$ est le paramètre de lissage associé à la classe $j$,
- $K(\cdot)$ est le *noyau uniforme* qui se définit comme

$$K(v) = \begin{cases} 1 & \text{si } \|v\|_\infty < 1 \\ 0 & \text{sinon.} \end{cases}$$

Nous déduisons de cette expression la forme de l'estimateur d'intensité $q_j(\cdot)$ :

$$\hat{q}_j(x) = \frac{1}{(2h_j)^k} \sum_{i=1}^{n_j} K\left(\frac{x - x_{ji}}{h_j}\right).$$

De même, nous pouvons estimer les probabilités a priori. En effet, nous avons que

$$p_j = \frac{\int_{D_j} q_j(y)dy}{\sum_{i=1}^{p} \int_{D_i} q_i(y)dy}.$$

En remplaçant les $q_k$ par leur estimation, nous obtenons

$$\hat{p}_j = \frac{\int_{D_j} \hat{q}_j(y)dy}{\sum_{i=1}^{p} \int_{D_i} \hat{q}_i(y)dy}.$$

Or, il est facile de prouver que

$$\int_{D_j} \hat{q}_j(y)dy = n_j.$$

Nous avons ainsi que

$$\hat{p}_j = \frac{n_k}{\sum_{i=1}^{p} n_i}.$$

Ainsi, le classement d'une nouvelle observation $x$ s'effectue en appliquant la règle de discrimination bayesienne et en remplaçant les intensités inconnues par leurs estimations. Notons toutefois que le choix du noyau n'est pas un facteur prépondérant en ce qui concerne la qualité des résultats.

### 3.2.6 Recherche des paramètres de lissage

La qualité de la classification que nous allons produire dépend essentiellement du choix des paramètres de lissage $h_j$ relatifs à chaque classe $j$. Ceux-ci doivent être déterminés de manière à minimiser au maximum le taux de mauvaise classification engendré. Ce dernier peut facilement être estimé à l'aide de la méthode du Leaving-One-Out que nous avons détaillée au cours du chapitre précédent.

De manière plus formelle, il s'agit de déterminer le paramètre $h$ qui minimise

$$\frac{1}{n}\sum_{i=1}^{n} L\left(c_i, \hat{c}_i(h)\right),$$

où

- $c_i$ représente la classe connue du $i^{\text{ème}}$ point de la base d'entraînement,
- $\hat{c}_i(h)$ est la classe prédite par le classificateur basé sur l'échantillon d'entraînement duquel on a retiré le point $i$,
- $L(a, b)$ est une fonction *de perte* quelconque.

Plusieurs fonctions de pertes sont envisageables. Nous choisirons la fonction de perte la plus commune, nommée *fonction de perte 0/1*. Elle se définit par

$$L(a,b) = \begin{cases} 1 & \text{si } a \neq b \\ 0 & \text{sinon.} \end{cases}$$

Cette fonction considère tout mauvais classement comme équi-indésirable et tout bon classement comme équi-désirable. Bien entendu, il est très simple de modifier une telle fonction en vue d'accorder plus de poids à un type spécifique de bon ou mauvais classement.

# Partie III

# Recherches et résultats

# Chapitre 1

# Présentation des échantillons de travail

Avant d'aborder la présentation des recherches effectuées au cours de cette année, nous allons passer en revue les différents échantillons d'entraînement auxquels nous avons appliqué nos méthodes.

## 1.1 Description

Nous disposons de cinq échantillons de données d'entraînement sous forme de fichiers informatiques.

| Nom | Contenu | Délai avant la faillite |
|---|---|---|
| *commerce.xls* | 3571 entreprises du secteur commercial français. | Entre 1 et 3 ans |
| *industri.xls* | 4018 entreprises du secteur industriel français. | Entre 1 et 3 ans |
| *ct.xls* | 1342 entreprises du secteur commercial belge. | 1 an |
| *mt.xls* | 5107 entreprises du secteur commercial belge. | 3 ans |
| *datab.xls* | 464 entreprises du secteur commercial belge. | Entre 1 et 3 ans |

Le délai avant faillite indique le temps écoulé entre la remise du dernier bilan et la déclaration de faillite ou la demande de concordat.

Les quatres premiers fichiers nous ont été transmis par les assurances du crédit "NAMUR". Le cinquième provient des recherches effectuées par B. RASSON.

Les entreprises contenues dans ces échantillons d'entraînement sont encodées selon les critères suivants.

1. Le numéro de T.V.A.
2. Le nom de l'entreprise.
3. L'*indicateur de faillite.*
4. $k$ ratios financiers déterminés sur base de l'information comptable.

L'indicateur de faillite est une variable binaire valant 1 si l'entreprise est en situation de faillite ou de concordat et est égal à 0.

Le tableau suivant nous donne la répartition des entreprises de chaque fichier selon les deux catégories décrites ci-dessus.

| Nom | Nombre de ratios | Nombre d'entreprise | Faillite | Saine |
|---|---|---|---|---|
| *commerce.xls* | 81 | 3571 | 920 | 2657 |
| *industri.xls* | 75 | 4018 | 1871 | 2147 |
| *ct.xls* | 8 | 1342 | 545 | 797 |
| *mt.xls* | 6 | 5107 | 2288 | 2819 |
| *modif.xls* | 17 | 464 | 273 | 191 |

La liste des noms des $k$ ratios financiers pour chacun des échantillons se trouve dans les annexes.

## 1.2 Orientation des recherches

Les recherches que nous avons réalisées se décomposent en deux parties.

La première concerne deux fichiers de travail nommés *comtrans.xls* et *indtrans.xls*, provenant respectivement des fichiers *commerce.xls* et *industri.xls* et ne contenant que quelques ratios sélectionnés par les assurances du crédit "NAMUR" comme étant les plus discriminants.

Ainsi, le fichier *comtrans.xls* dispose des six ratios suivants.

1. *Age*
2. *Rentaeco*
3. *Collperi*
4. *R4*
5. *Rc_Dct*
6. *Dct_Bil*

De même, le fichier *indtrans.xls* possède huit ratios.

1. *Performn*
2. *Stocktrno*
3. *Cstemplt*
4. *Cap_Dtt*
5. *R4prc*
6. *Dct_Bil*
7. *Cvcafrn*
8. *Va_Bx*

La deuxième partie comprend les recherches effectuées sur les cinq échantillons initiaux. Elles concernent davantage des méthodes de sélection de variables discriminantes en fonction des techniques employées.

## 1.3 Autres indications

Toutes les méthodes que nous avons mises au point ont été programmées en langage *Fortran 77* sur deux stations :

- une station *DECterm 5000/125* munie d'un processeur unique et d'une capacité de 5 méga-octets de mémoire vive,
- une station *SUNSparc 20* possédant deux processeurs en parallèle et 256 méga-octets de mémoire vive. Ce puissant outil a été mis à la disposition de J-P. Rasson par le Ministère de l'Equipement et des Transports (M.E.T.).

L'utilisation de la deuxième station a rendu possible un certain nombre de techniques extrêment coûteuses du point de vue du stockage des donnés et du temps d'exécution.

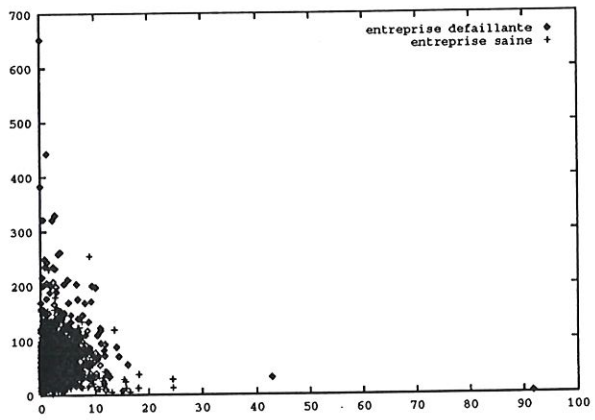
# Chapitre 2

# Recherches réalisées à partir des fichiers indtrans.xls et comtrans.xls

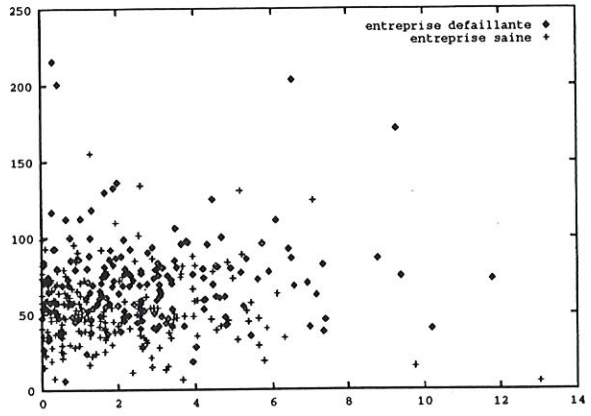
## 2.1 Introduction

Avant toute chose, il peut être intéressant de visualiser les deux échantillons sous forme de nuages de points dans $I\!R^2$ afin de se faire une idée de la complexité du problème. Pour cela, nous avons selectionné deux ratios pour chaque fichier.

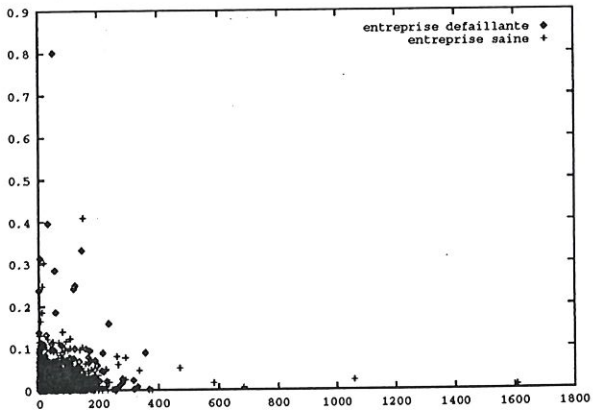
Le graphique suivant montre la répartition des 4018 entreprises du fichier *indtrans.xls* selon les ratios *R4Prc* et *Dct_Bil*.



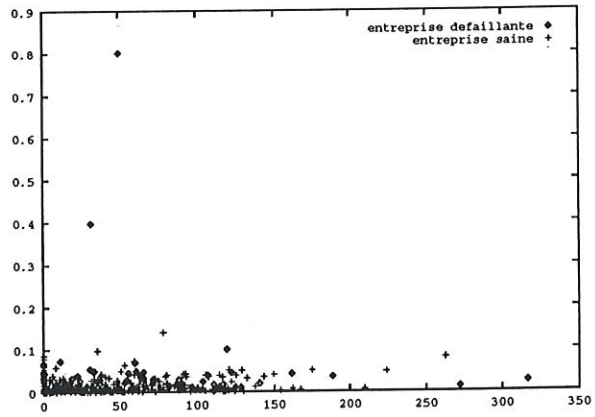

Si nous prenons une entreprise sur dix (401 au total), nous obtenons la représentation moins "encombrée" suivante.




Nous procédons de même avec le fichier *comtrans.xls*. Les ratios représentés sont *Coll-peri* et *R4*.




En prenant une entreprise sur dix (357 au total), nous avons le graphe suivant.




Par ailleurs, le lecteur trouvera en annexe une série de graphes illustrant la répartition des entreprises par classes pour chacun des ratios que nous utilisons.

Parmi les techniques expérimentées, nous avons retenu celles dont les résultats étaient dignes d'intérêt. Nous nous sommes attachés dans un premier temps à appliquer les méthodes paramétriques de discrimination habituellement utilisées par les analystes financiers, et ce dans une optique de comparaison. Par ailleurs, notons que toutes les estimations des taux d'erreur engendrés ont été effectuées par la méthode du *Leaving-One-Out.*

## 2.2 Méthodes de discrimination paramétriques

Dans la partie précédente, nous avons présenté deux méthodes paramétriques de discrimination : la méthode linéaire de Fisher et la méthode quadratique. Nous avons appliqué ces techniques à nos échantillons à l'aide de la librairie de routines mathématiques *IMSL.*

### 2.2.1 Méthode linéaire de Fisher

L'emploi de la méthode de Fisher sur le fichier *indtrans.xls* a donné la *matrice de confusion*

$$\begin{pmatrix} 1100 & 771 \\ 402 & 1745 \end{pmatrix}$$

Ce qui correspond à un taux de bon reclassement de 70.81%.

Pour le fichier *comtrans.xls*, nous obtenons la matrice de confusion

$$\begin{pmatrix} 202 & 749 \\ 100 & 2520 \end{pmatrix}$$

et le taux de bon de bon reclassement vaut 76.23%.

**Remarque**

Une *matrice de confusion* se lit de la manière suivante :

- La ligne donne la classe d'appartenance d'origine de l'entreprise.
- La colonne indique la classe d'appartenance de l'entreprise, déterminée par l'algorithme utilisé.

Précisons que

- la classe 1 représente le groupe des entreprises défaillantes,
- la classe 2 représente le groupe des entreprises saines.

Cette convention d'écriture nous permet de mieux nous rendre compte des proportions de bon et mauvais reclassements produits lors de la discrimination.

### 2.2.2 Méthode quadratique

L'emploi d'une fonction discriminante quadratique sur le fichier *indtrans.xls* donne

$$\begin{pmatrix} 1179 & 692 \\ 575 & 1572 \end{pmatrix}$$

et le taux de bon reclassement vaut 68.47%.

Par alleurs, nous avons pour le fichier *comtrans.xls* la matrice de confusion

$$\begin{pmatrix} 186 & 765 \\ 146 & 2474 \end{pmatrix}$$

Par conséquent, le taux de bon reclassement est ici égal à 74.48%.

### 2.2.3 Inconvénient

Comme nous l'avons déjà expliqué, ces deux approches paramétriques posent une hypothèse de normalité sur la distribution des données. Celle-ci a pour avantage de fournir au mathématicien des expressions analytiques simples. Cependant, les statisticiens s'accordent pour dire qu'une telle supposition est souvent peu réaliste.

## 2.3 Méthodes de discrimination non paramétriques basées sur une transformation de l'échantillon

Les recherches qui suivent utilisent la règle de discrimination bayesienne basée sur le processus de Poisson non homogène. Pour rappel, toute nouvelle observation $x$ est assignée à la classe $j$ qui maximise

$$p_j f_j(x) .$$

Or, nous avons vu que les $f_j(x)$ sont estimés par

$$\hat{f}_j(x) = \frac{1}{n_j(2h_j)^k} \sum_{i=1}^{n_j} K\left(\frac{x - x_{ji}}{h_j}\right) .$$

L'utilisation d'un paramètre de lissage $h_k$ unique (toute dimension confondue) pose un problème sérieux : le noyau placé sur chaque donnée est de même grandeur dans toutes les dimensions. Si l'étendue des données est beaucoup plus grande dans une dimension spécifique, cela nous conduira à de mauvais résultats.

A titre d'exemple, reprenons le graphe concernant le fichier *comtrans.xls* de la page 62. Nous voyons que le ratio *Collperi* prend pour ces entreprises des valeurs comprises entre 0 et 1600, tandis que les données relatives au ratio *R4* se répartissent sur un intervalle compris entre 0 et 1. Supposons que nous fixons à 5 les valeurs des paramètres de lissage relatifs aux deux classes. Dans ce cas, la définition du noyau uniforme $K$ donnée par

$$K(v) = \begin{cases} 1 & \text{si } \|v\| < 1 , \\ 0 & \text{sinon,} \end{cases}$$

permet de déduire que le ratio *R4* ne sera pas pris en compte car l'écart entre les valeurs du ratio pour n'importe quelle paire d'entreprises est toujours inférieur ou égal à 1.

Pour résoudre ce problème, nous avons le choix entre

- soit transformer les données de manière à ce que chaque ratio prenne des valeurs sur un intervalle commun,
- soit déterminer une nouvelle fonction noyau adaptée à cette situation.

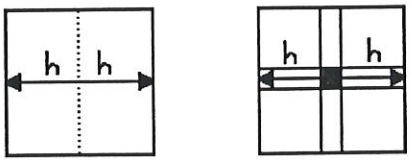
Nous allons décrire dans cette section les procédés inspirés par la première alternative.

### 2.3.1 Discrétisation des données

Les logiciels qui ont été mis au point par le laboratoire GEOSATEL dans le cadre de la classification d'images traitent des données tri-dimensionnelles (les valeurs spectrales Red Green Blue) appartenant à l'espace $[0, 255]^3$. Nous avons donc *discrétisé* les valeurs prises par les entreprises pour chacun des ratios traités sur un intervalle $[0, 255]$.

Puisque nous travaillons dans l'espace discret $[0, 255]^k$, nous devons adapter nos estimateurs de densité $\hat{f}_j(\cdot)$. En effet, l'estimateur du noyau multivarié que nous utilisons

ne s'applique qu'au cas continu. Dans le cas discret, le volume de la fenêtre n'est plus représenté par une longueur mais par un nombre de parcelles.




Nous voyons sur ce schéma que la longueur de la fenêtre passe de la valeur $2h$ à $2h+1$. L'estimateur de densité devient

$$\hat{f}_j(x) = \frac{1}{n_j(2h_j+1)^k} \sum_{i=1}^{n_j} K\left(\frac{x - x_{ji}}{h_j}\right)$$

Après avoir effectué la digitalisation, il faut appliquer la règle de discrimination en retenant les paramètres de lissage par classe $h_j$ fournissant le taux maximal de bon reclassement.

Ce procédé comprend quelques avantages. Premièrement, nous contournons le problème provoqué par l'utilisation d'un paramètre de lissage commun à toutes les dimensions car les données se situent à présent sur une même échelle. Deuxièment, nous ne traitons pas des données *réelles* mais *entières*, ce qui signifie une réduction considérable des coûts de stockage et des temps d'exécution des programmes.

Néanmoins, nous constatons que l'application de cette transformation modifie de manière considérable la forme de la distribution des données. En effet, nous observons un phénomène de *dilatation* (resp. une *de contraction*) si l'intervalle contenant les valeurs initiales que prennent les entreprises pour un ratio donné est de longueur *inférieure* (resp. *supérieure*) à 255.

L'application de cette technique au fichier *indtrans.xls* et la recherche de paramètres de lissage optimaux par classe donnent la matrice de confusion

$$\begin{pmatrix} 0 & 0 & 0 \\ 84 & 1138 & 649 \\ 52 & 555 & 1540 \end{pmatrix}$$

Nous avons donc un taux de bon reclassement de 66.65%.

Les résultats pour le fichier *comtrans.xls* sont

$$\begin{pmatrix} 0 & 0 & 0 \\ 6 & 149 & 796 \\ 3 & 103 & 2514 \end{pmatrix}$$

ce qui correspond à un taux de bon reclassement de 74.58%.

**Remarque**

Dans les méthodes quadratique et linéaire, nous avons vu que le calcul du score $f(\cdot)$ de l'entreprise détermine sa classe d'appartenance par la règle de discrimination paramétrique

Affecter toute nouvelle observation à la classe des entreprises saines si

$$f(x) \geq 0$$

La règle de discrimination bayesienne que nous utilisons classe quant à elle toute nouvelle observation $x$ à la classe maximisant les probabilités a posteriori $p_k f_k(x)$.

Cependant, en cas d'égalité, l'observation est considérée comme *non classée*. Comme il ne s'agit pas d'un mauvais ou bon reclassement, nous traiterons ce cas de figure comme s'il s'agissait d'une classe à part. C'est pourquoi les matrices de confusion que nous venons de montrer sont des matrices carrées d'ordre 3. Nous considérerons donc

- la classe 1 : les entreprises non classées,
- la classe 2 : les entreprises défaillantes,
- la classe 3 : les entreprises saines.

De par cette convention, nous avons que la première ligne est toujours nulle.

### 2.3.2 Réajustement des données

Nous savons que toutes les valeurs prises par les entreprises pour les $k$ ratios se répartissent sur $k$ intervalles différents. L'idée de cette méthode est de réajuster les données relatives à $k-1$ ratios par rapport au ratio dont l'intervalle de répartition est le plus grand. En d'autres termes, si note par $l$ le numéro du ratio possèdant le plus grand intervalle d'échantillon, toutes les données sont transformées de manière à ce que leur intervalle de répartition soit identique à l'intervalle du ratio $l$. Nous procédons de ce fait à une *dilatation* des données, exception faite de celles concernant le ratio $l$.

Grâce à ce procédé, nous évitons toujours le problème dû à l'emploi de l'estimateur du noyau uniforme multivarié. De plus, les formes de distribution des $k-1$ ratios sont altérées de la même manière. Néanmoins, les données que nous traitons sont réelles, ce qui signifie une augmentation des coûts de stockage et des temps d'exécution des programmes.

L'emploi de cette méthode sur le fichier *indtrans.xls* et la recherche de paramètres de lissage optimaux par classe donne la matrice de confusion

$$\begin{pmatrix} 0 & 0 & 0 \\ 53 & 1057 & 761 \\ 40 & 490 & 1617 \end{pmatrix}$$

Les données ont été réajustées sur le ratio *Stocktrno* dont l'intervalle de répartion est $[0, 1607]$. Le taux de bon reclassement vaut 66.57%.

Pour le fichier *comtrans.xls*, nous obtenons

$$\begin{pmatrix} 0 & 0 & 0 \\ 13 & 157 & 781 \\ 15 & 91 & 2514 \end{pmatrix}$$

Les données ont été réajustées sur le ratio *Collperi* dont l'intervalle de répartion est $[0, 3147]$. Le taux de bon reclassement est égal 74.89%.

### 2.3.3 Normalisation des données

La méthode suivante se base la transformation

$$x'_{ijd} = \frac{x_{ijd} - \bar{x}_{jd}}{\sigma_{jd}}$$

où

- $j$ représente la classe d'appartenance de l'entreprise $i$ $(1 \leq i \leq n_j)$,
- $d$ indique le ratio utilisé $(1 \leq d \leq k)$,
- $x_{ijd}$ est le vecteur à $k$ dimensions relatif à l'entreprise $i$ avant la transformation,
- $\bar{x}_{jd}$ est la moyenne des valeurs prises pour le ratio $d$ par les entreprises dont la classe d'appartenance est $j$,
- $\sigma_{jd}$ est l'écart-type empirique des valeurs prises pour le ratio $d$ par les entreprises dont la classe d'appartenance est $j$.

Cette transformation, correspondant à une homogénéisation des donnée, vise à ramener les valeurs prises par les entreprises d'une classe donnée pour un ratio donné à une moyenne nulle et une variance unité. En d'autres termes, on effectue un *centrage* et une *réduction* des données. De cette manière, les $k$ intervalles de répartition sont plus ou moins identiques.

L'application de la règle de discrimination bayesienne sur le fichier *indtrans.xls* et la recherche de paramètres de lissage optimaux par classe donne la matrice de confusion

$$\begin{pmatrix} 0 & 0 & 0 \\ 15 & 1111 & 745 \\ 11 & 569 & 1567 \end{pmatrix}$$

Le taux de bon reclassement est 66.65%.

Pour le fichier *comtrans.xls*, nous avons

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 165 & 785 \\ 1 & 100 & 2519 \end{pmatrix}$$

Le taux de bon reclassement est 75.16%.

## 2.4 Produit de fonctions noyaux

La seconde alternative au problème décrit en 2.3 consistait à trouver une fonction noyau adaptée à la situation. C'est pourquoi nous proposons l'utilisation d'un produit de fonctions noyaux ou *kernel product*. Il s'agit en fait d'une fonction noyau multivariée, qui n'est rien d'autre que le produit de $k$ fonctions noyaux uniformes.

Le produit de noyau se définit par

$$K(v) = \prod_{i=1}^{k} K_i(v_i)$$

où

- $v$ est un vecteur à $k$ composantes,
- $K_i$ est une fonction noyau uniforme appliquée à la $i^{\text{ème}}$ composante du vecteur $v$.

Nous pouvons en déduire que

$$K(v) = \begin{cases} 1 & \text{si } v_j < 1 \ (j = 1, \ldots, k), \\ 0 & \text{sinon.} \end{cases}$$

L'estimateur de densité $\hat{f}_j(\cdot)$ devient

$$\hat{f}_j(x) = \frac{1}{n_j} \frac{1}{2^k \prod_{d=1}^{k} h_{jd}} \sum_{i=1}^{n_j} K\left(\frac{x - x_{ji}}{h_j}\right)$$

où

- $h_j$ est un vecteur à $k$ composantes $h_{jd}$ $(d = 1, \ldots, k)$,
- $K$ est une produit de fonctions noyaux uniformes.

L'estimateur de noyau multivarié que nous utilisions auparavant plaçait des hypercubes de dimension $k$ et d'une largeur $2h$ sur chaque observation. L'estimateur que nous venons de définir remplace les hypervolumes par des parallélotopes dont la longueur des arêtes est fixée par $k$ données, à savoir un vecteur $h$ à $k$ composantes.

Ainsi, nous ne cherchons plus deux paramètres de lissage optimaux mais bien $2k$ paramètres optimaux, c'est-à-dire un par classe et par dimension. Grâce à cette technique, nous ne sommes plus obligés de standardiser les données par une transformation quelconque.

Cependant, nous faisons face à un autre problème important d'ordre technique. Il nous est impossible d'effectuer une recherche simultanée des $2k$ paramètres de lissage optimaux. En effet, les coûts de stockage et les temps d'exécution exigés par cette méthode sont trop lourds pour les stations dont nous disposons. Nous avons donc appliqué la technique du produit de fonctions noyaux selon deux approches particulières.

### 2.4.1 Recherche des paramètres de lissage avec réajustement

L'idée de cette approche est de déterminer dans un premier temps les paramètres de lissages $h_{jd}(j = 1, 2$ et $d = 1, \ldots, k)$ optimaux par classe, et ce indépendamment pour chaque ratio. Nous ne retenons que les ratios les plus prometteurs. Ceux-ci sont soumis à un produit de deux fonctions noyaux. Cependant, si nous utilisons les paramètres $h_{jd}$ que nous avons déterminé séparément, le taux de bon reclassement résultant du produit de deux fonctions noyaux sera inférieur aux taux de bon reclassement engendrés par l'estimateur du noyau uniforme pour les deux ratios concernés. En effet, il se peut qu'une entreprise $x$ soit classée dans la catégorie des entreprises saines par le ratio 1 et répertoriée dans la classe des entreprises en faillite par le ratio 2. Afin d'éviter ce problème, nous réajusterons les $h_{jd}$ entre eux en leur imposant un certain nombre de variations. Les nouveaux paramètres de lissage ainsi déterminés permettront l'inclusion d'autres ratios suivant le même procédé.

Plus concrètement, les temps d'éxecution et les coûts de stockage augmentent de manière exponentielle avec le nombre de ratios utilisés. L'utilisation de trois ratios demande énormément de temps pour un faible nombre de variations. L'inclusion d'un quatrième ratio est techniquement irréalisable pour des raisons liées au langage *Fortran 77* et à la taille de la mémoire vive disponible sur la station *SUNSparc*.

### Résultats concernant le fichier *indtrans.xls*

La recherche des meilleurs paramètres de lissage par ratio et par classe a donné les résultats suivants.

| Ratio | Taux de bon reclassement |
|---|---|
| 1 | 66.22 % |
| 2 | 54.38 % |
| 3 | 58.78 % |
| 4 | 67.74 % |
| 5 | 62.31 % |
| 6 | 65.75 % |
| 7 | 62.89 % |
| 8 | 57.31 % |

Les quatre meilleurs ratios 1, 4, 6 et 7 ont été testés deux par deux. Les résultats de l'utilisation d'un produit de deux fonctions noyaux sont

| Ratios utilisés | Taux de bon reclassement |
|---|---|
| 1 et 4 | 67.5 % |
| 1 et 6 | 69.4 % |
| 1 et 7 | 64.6 % |
| 4 et 6 | 69.5 % |
| 4 et 7 | 63.6 % |
| 6 et 7 | 63.7 % |

Nous avons retenu les ratios 1, 4 et 6 pour l'utilisation d'un produit de trois fonctions noyaux. Le taux de bon reclassement est dans ce cas de 69.7%.

**Résultats concernant le fichier** *comtrans.xls*

En ce qui concerne la recherche par ratio des meilleurs paramètres de lissage par classe, nous obtenons

| Ratio | Taux de bon reclassement |
|---|---|
| 1 | 57.18 % |
| 2 | 74.99 % |
| 3 | 54.38 % |
| 4 | 73.42 % |
| 5 | 66.39 % |
| 6 | 75.24 % |

Seuls les ratios 2, 4, 5 et 6 ont été retenus. Les résultats d'un produit de deux fonctions noyaux sont

| Ratios utilisés | Taux de bon reclassement |
|---|---|
| 2 et 4 | 75.3 % |
| 2 et 5 | 75.4 % |
| 2 et 6 | 76.2 % |
| 4 et 5 | 74.8 % |
| 4 et 6 | 75.9 % |
| 5 et 6 | 76.9 % |

Nous avons utilisé les ratios 2,5 et 6 pour un produit de trois fonctions noyaux. Le taux de bon reclassement réalisé est égal à 76.5%.

### 2.4.2 Méthode "step by step"

Le principe "step by step" consiste, dans un premier temps, à appliquer une certaine technique de classification séparément sur les $k$ ratios. Nous retenons alors celui pour lequel le taux de bon reclassement est maximum. Ensuite, nous réappliquons la technique mais en deux dimensions. En d'autres termes, nous utilisons le ratio déterminé par la première étape tout en testant les $k-1$ ratios restants. Parmi ceux-ci, nous conserverons le ratio qui, couplé au ratio retenu précédemment, améliore au mieux le taux de bon reclassement. En réitérant ce procédé, nous pouvons déterminer les $p$ $(1 \leq p \leq k)$ variables les plus discriminantes selon la technique employée.

En ce qui concerne la technique du produit de noyaux, nous effectuerons d'abord une recherche par ratio des paramètres de lissage optimaux par classe selon une fonction noyau uniforme. Nous retiendrons donc le ratio noté $l$ pour lequel le taux de bon reclassement est maximum. Nous considérerons que les paramètres de lissages optimaux par classe de ce ratio sont fixés de manière définitive. Ensuite, nous réinjectons ces paramètres dans un produit de deux fonctions noyaux. Nous déterminerons ainsi pour les $k-1$ ratios restants le ratio pour lequel le taux de bon reclassement obtenu améliore au maximum le taux précédent. Nous considérerons à nouveau que les paramètres de lissage optimaux par classe obtenus pour ce second ratio sont fixés. En poursuivant selon cette idée, nous pourrons situer les $p$ ratios les plus discriminants pour la méthode du produit de fonctions noyaux.

L'avantage de cette approche réside dans le fait que nous ne déterminons que 2 paramètres de lissage à chaque étape. De cette manière, nous diminuons de manière significative les temps d'exécution et les coûts de stockage associés.

**Résultats**

L'utilisation de ce procédé pour le fichier *indtrans.xls* fournit un taux de bon reclassement égal à 71.06% avec les 8 ratios dont l'ordre d'apparation est 4, 6, 5, 3, 2, 1, 7, 8.

En ce qui concerne le fichier *comtrans.xls*, nous obtenons un taux de bon reclassement valant 76.53% avec les 6 ratios dont l'ordre d'apparition est 6, 5, 4, 2, 3, 1.

## 2.5 Technique des $m$ plus proches voisins

La dernière technique que nous avons employée est la règle des $m$ plus proches voisins [3]. Le principe de cette méthode est simple : toute observation $x$ est affectée au groupe qui compte le plus de représentants parmi les $m$ observations les plus proches de $x$ au sens de la distance euclidienne. Si nous travaillons avec $m = 1$, on retrouve la règle du plus proche voisin : l'observation $x$ est affectée à la classe d'appartenance de l'observation la plus proche de $x$.

Nous sommes en présence d'une méthode de discrimination non paramétrique. En effet, soit $x$ une nouvelle observation à classer. Le volume de la plus petite hypersphère, centrée en $x$, contenant $m$ observations est noté $v(x)$. Nous définissons les estimateurs de densité

$f_j(x)$ par

$$\hat{f}_j(x) = \frac{m_j}{n_j v(x)},$$

où $m_j$ représente le nombre d'observations présentes dans l'hypersphère dont la classe d'appartenance est $i$.

En reprenant les probabilités a priori $p_j = \frac{n_j}{n}$, nous avons que

$$\begin{aligned} \hat{f}(x) &= \sum_{j=1}^{p} p_i \hat{f}_j(x) \\ &= \frac{m}{nv(x)}. \end{aligned}$$

Dès lors, l'estimation de la probabilité a posteriori d'appartenance de $x$ à la classe $j$ est égale à

$$\begin{aligned} \hat{p}_{j,\text{post}} &= \frac{n_j}{n} \frac{\frac{m_j}{n_j v(x)}}{\frac{m}{nv(x)}} \\ &= \frac{m_j}{m} \end{aligned}$$

La règle des $m$ plus proches voisins est donc bien une règle bayesienne.

Cette méthode présente de nombreux avantages. Nous ne sommes plus confrontés aux problèmes relatifs aux paramètres de lissage. De plus, les coûts de stockage et temps d'exécution ne sont pas très élevés. Nous avons appliqué cette technique à nos deux échantillons en utilisant l'approche "step by step" décrite ci-dessus. Le nombre de voisins a été fixé à 99.

**Résultats**

Pour le fichier *indtrans.xls*, nous obtenons un taux de bon reclassement égal à 71.06% avec les 8 ratios dans l'ordre d'apparition 4, 6, 5, 3, 2, 1, 7, 8.

L'application du procédé au fichier *comtrans.xls* fournit un taux de bon reclassement valant 75.78% avec 4 ratios dans l'ordre d'apparition 2, 1, 5, 4. Les ratios 3 et 6 n'apportent aucune amélioration.

## 2.6 Première conclusion

Comme on peut le constater, les performances des diverses méthodes non paramétriques que nous avons testées ne sont guères plus élevées que les performances des méthodes linéaire et quadratique. Remarquons néanmoins que la technique du produit de fonctions noyaux et la règle des $m$ plus proches voisins auxquelles nous avons appliqué l'approche "step by step" produisent des taux de bon reclassement supérieurs aux méthodes classiques. Cependant, le gain est peu significatif.

# Chapitre 3

# Comparaison entre les méthodes paramétriques et les méthodes non paramétriques basées sur une approche "step by step"

Nous allons maintenant présenter les résultats provenant de l'application des méthodes paramétriques (linéaire et quadratique) et des méthodes non paramétriques selon une approche "step by step" (technique du produit de noyau et $m$ plus proches voisins) sur cinq fichiers, à savoir

1. *industri.xls*,
2. *commerce.xls*,
3. *mt.xls*,
4. *ct.xls*,
5. *datab.xls*.

Ces résultats permettront de comparer les performances de chacunes des méthodes.

## 3.1 Données manquantes

Les fichiers *industri.xls* et *commerce.xls* contiennent des entreprises présentant un certain nombre de données manquantes. Nous avons mis au point un programme permettant dans un premier temps d'exclure les ratios dont le pourcentage de données manquantes est supérieur à un certain seuil fixé. Ensuite, ce programme supprime toutes les entreprises contenant des données manquantes parmi les ratios restant.

Nous avons ainsi créé deux nouveaux fichiers de travail.

1. Le fichier *indus10.xls*.

   Le seuil de tolérance est fixé à 10%. Ceci entraîne la suppression des ratios 38, 43 à 49, 53, 55 à 59, 68, 70, 72 et 73 et l'exclusion de 440 entreprises.

   En définitive, le fichier *indus10.xls* comprend ainsi 3578 entreprises avec 58 ratios.

2. Le fichier *commer05.xls*.

   Le seuil de tolérance est fixé à 5%. Les ratios 32,33,44,49 à 55,57,59,61 à 65,74,76,78 sont supprimés. De même, 359 entreprises sont éliminées.

   Le fichier *commer05.xls* contient donc 3212 entreprises avec 61 ratios.

## 3.2 Résultats

Le tableau suivant reprend les pourcentages de bon reclassement obtenus pour chacun des 5 fichiers avec chacune des 4 méthodes. Les ratios retenus par les méthodes non paramétriques sont indiqués entre parenthèses dans leur ordre d'apparition.

| Méthodes / Fichiers | Paramétriques | | Non paramétriques | |
|---|---|---|---|---|
| | Linéaire | Quadratique | Produit de fonctions noyaux | m plus proches voisins |
| *indus10.xls* | 69.73 % | 60.12 % | 71.90 %<br>(16-26-37-33-54-6-2-12-14-19) | 71.88 %<br>(42-7-39-15-60-62-36)<br>(m=99) |
| *commer05.xls* | 75.50 % | 74.16 % | 77.60 %<br>(18-68-4-25-37-11-66-29-23-5) | 77.15 %<br>(18-69-72-15-66-46-67-43)<br>(m=99) |
| *ct.xls* | 79.51 % | 76.53 % | 78.60 %<br>(2-4-3-7-5-6-1-8) | 79.58 %<br>(2-4-1-7-3-6)<br>(m=29) |
| *mt.xls* | 69.91 % | 68.10 % | 68.76 %<br>(2-4-3-7-6) | 70.04 %<br>(6-1-3-4-5-2)<br>(m=99) |
| *datab.xls* | 76.29 % | 73.49 % | 79.31 %<br>(3-1-9-11-15) | 77.80 %<br>(3-1-5-12-7-11-13-16-14-15)<br>(m=19) |

# Conclusion

Parmi les nombreuses techniques que nous avons expérimentées, nous retiendrons la règle des $m$ plus proches voisins et la méthode utilisant le produit de fonctions noyaux, selon l'approche "step by step". Ces méthodes sont facilement implémentables et ne recquièrent pas des coûts de stockage élevés. De plus, leur temps d'exécution est raisonnable.

Malheureusement, force est de constater que les performances réalisées par ces méthodes ne sont guère plus élevées que celles obtenues par les méthodes classiques. Le gain obtenu varie en effet de 1% à 2% selon le fichier utilisé.

Dans de nombreux domaines, la règle de discrimination fondée sur l'hypothèse d'un processus de Poisson non homogène permet d'obtenir des taux de bon classement généralement supérieurs à ceux des règles quadratique et linéaire. En ce qui concerne la télédétection, le gain obtenu est en moyenne d'environ 20%.

La question reste posée : comment obtenir de meilleurs taux de bon reclassement ? Nous pouvons émettre quelques hypothèses.

La construction d'un modèle financier passe en premier lieu par une sélection des ratios les plus pertinents et en second lieu par une méthode de classification efficace. Nous nous sommes essentiellement attachés à la seconde étape. Le problème mérite une plus ample réflexion propre à l'analyse financière.

Les travaux réalisés à l'Observatoire des entreprises de la Banque de France [8] ont mis en exergue l'importance de l'analyse économique du schéma de dégradation de la situation financière de l'entreprise en situant les entreprises traitées sur sept classes distinctes représentant différentes étapes caractéristiques du *failure path*. Ainsi, il est possible de définir des ratios plus spécifiques selon les classes prises en compte. Selon ce point de vue, il nous semble intéressant d'envisager l'application des méthodes que nous avons construites à la place des règles de discrimination classiques qui sont actuellement employées.

# Bibliographie


[1] B. RASSON, "La faillite en Belgique: description et prévision. Application de la technique de convexité.", Mémoire, 1995.

[2] B.W. SILVERMANN, "Density Estimation for Statistics and Data Analysis.", Chapmann and Hall, 1986.

[3] P. BEAUFAYS,"Une nouvelle règle en analyse discriminante", Thèse de doctorat, 1985.

[4] J.P. RASSON, F. ORBAN-FERAUGE, V. GRANVILLE,"From a natural to a behavioral classification rule", 1993.

[5] C. VAN WYMEERSCH, "Comptabilité financière", Syllabus, 1992.

[6] H. OOGHE, C. VAN WYMEERSCH, "Traité d'analyse financière", Presses Universitaires de Namur, 1985.

[7] P. DONCEL, "Notions élémentaires de droit", Collection "Gestion de l'entreprise", 1988.

[8] M. BARDOS, "Les défaillances d'entreprises dans l'industrie: ratios significatifs, processus de défaillance, détection précoce", Banque de France, Collection Entreprises, étude B 95/03.

# Annexe A

# Exemple de bilan d'une entreprise

Les pages suivantes reprennent un extrait de bilan non consolidé de l'Union Minière S.A. [5].

# BILANS AU 31 DECEMBRE
## NON CONSOLIDES

**ACTIF** (BEF milliers)

| | 1991 | 1990[1] |
|---|---|---|
| **ACTIFS IMMOBILISES** | **39.536.940** | **40.489.692** |
| **II. Immobilisations incorporelles** | **129.955** | **92.848** |
| **III. Immobilisations corporelles** | **12.847.820** | **13.448.608** |
| A. Terrains et constructions | 4.702.957 | 4 528.918 |
| B. Installations, machines et outillage | 6.319.497 | 6.190 934 |
| C. Mobilier et matériel roulant | 504.544 | 460 432 |
| D. Location-financement et droits similaires | 4.855 | 7.796 |
| E. Autres immobilisations corporelles | 408.547 | 465.460 |
| F. Immobilisations en cours et acomptes versés | 907.420 | 1.795.068 |
| **IV. Immobilisations financières** | **26.559.165** | **26.948.236** |
| A. Entreprises liées | | |
| 1. Participations | 25.268.715 | 24 970.901 |
| 2. Créances | 739.878 | 817.941 |
| B. Autres entreprises avec lesquelles il existe un lien de participation | | |
| 1. Participations | 323.457 | 675.580 |
| C. Autres immobilisations financières | | |
| 1. Actions et parts | 86.802 | 343.407 |
| 2. Créances et cautionnements en numéraire | 140.313 | 140.407 |
| **ACTIFS CIRCULANTS** | **30.038.212** | **27.498.029** |
| **V. Créances à plus d'un an** | **77.423** | **111.554** |
| A. Créances commerciales | 54.722 | 82 723 |
| B. Autres creances | 22.701 | 28.831 |
| **VI. Stocks et commandes en cours d'exécution** | **19.112.058** | **14.197.071** |
| A. Stocks | | |
| 1. Approvisionnements | 6.913.491 | 4 887.297 |
| 2. En-cours de fabrication | 1.546.634 | 1.678 493 |
| 3. Produits finis | 7.951.606 | 6.900 800 |
| 6. Acomptes verses | 2.534 448 | 730.481 |
| B. Commandes en cours d'execution | 165.879 | – |
| **VII. Créances à un an au plus** | **5.373.601** | **7.153.405** |
| A. Créances commerciales | 4 457 949 | 5 901 446 |
| B. Autres creances | 915.652 | 1.251.959 |
| **VIII. Placements de trésorerie** | | |
| B. Autres placements | **3.995.082** | **5.211.570** |
| **IX. Valeurs disponibles** | **754.802** | **452.290** |
| **X. Comptes de régularisation** | **725.246** | **372.139** |
| **TOTAL DE L'ACTIF** | **69.575.152** | **67.987.721** |

[1] Les chiffres de 1990 ont été retraités pour les rendre comparables avec ceux de 1991. Voir à ce sujet le commentaire de la rubrique "Stocks" en page 64 de ce rapport

# COMPTES DE RESULTATS
## NON CONSOLIDES

(BEF milliers)

| | | 1991 | 1990 |
|---|---|---|---|
| **I.** | **Ventes et prestations** | **67.895.495** | **80.027.133** |
| | A. Chiffre d'affaires | 65.337.271 | 78.000.288 |
| | B. Variation des en-cours de fabrication, des produits finis et des commandes en cours d'exécution (augmentation +, réduction −) | 1.030.199 | −541.673 |
| | C. Production immobilisée | 713.120 | 570.855 |
| | D. Autres produits d'exploitation | 814.905 | 1.997.663 |
| **II.** | **Coût des ventes et des prestations** | **67.355.818** | **78.116.749** |
| | A. Approvisionnements et marchandises | | |
| | 1. Achats | 48.677.349 | 56.812.432 |
| | 2. Variation des stocks (augmentation −, réduction +) | −1.630.148 | 821.341 |
| | B. Services et biens divers | 4.865.889 | 5.253.238 |
| | C. Rémunérations, charges sociales et pensions | 13.051.756 | 13.062.719 |
| | D. Amortissements et réductions de valeur sur frais d'établissement, sur immobilisations incorporelles et corporelles | 2.105.396 | 1.861.686 |
| | E. Réductions de valeur sur stocks, sur commandes en cours d'exécution et sur créances commerciales (dotations +, reprises −) | 202.798 | −69.571 |
| | F. Provisions pour risques et charges (dotations +, utilisations et reprises −) | −268.284 | −124.582 |
| | G. Autres charges d'exploitation | 351.062 | 499.486 |
| **III.** | **Résultat d'exploitation** | **539.677** | **1.910.384** |
| **IV.** | **Produits financiers** | **1.174.586** | **4.209.502** |
| | A. Produits des immobilisations financières | 313.711 | 2.727.568 |
| | B. Produits des actifs circulants | 532.080 | 919.332 |
| | C. Autres produits financiers | 328.795 | 562.602 |
| **V.** | **Charges financières** | **2.865.981** | **1.451.549** |
| | A. Charges des dettes | 1.855.162 | 975.594 |
| | B. Réductions de valeur sur actifs circulants autres que ceux visés sub II.E. (dotations +, reprises −) | −6.066 | 21.626 |
| | C. Autres charges financières | 1.016.885 | 454.329 |
| **VI.** | **Résultat courant avant impôts** | **−1.151.718** | **4.668.337** |

# Annexe B

# Liste des ratios utilisés

## B.1 Fichier *industri.xls*

| | | | | | |
|---|---|---|---|---|---|
| 1. | INDIC | 31. | CFLW_DTT | 61. | DCT_BIL |
| 2. | AGE | 32. | CAP_DTT | 62. | AC_CHEX |
| 3. | ENDET | 33. | @ACTC_T | 63. | AC_IMMO |
| 4. | AUTOFIN | 34. | BNF_TBL | 64. | DCT_IMMO |
| 5. | DAMORTIC | 35. | EH_DTT | 65. | AC_DT |
| 6. | LIQGE | 36. | R1 | 66. | DT_CA |
| 7. | LIQRED | 37. | R2 | 67. | EBE_CA |
| 8. | ENDETGL | 38. | R2ANN | 68. | FINACAN |
| 9. | CAPAREMB | 39. | R3 | 69. | FINAC |
| 10. | CAPAAUTO | 40. | R4PRC | 70. | CVCAFRNA |
| 11. | COUVCABF | 41. | R5 | 71. | CVCAFRN |
| 12. | EXPORT | 42. | SCORE_C_ | 72. | CVCABFRN |
| 13. | PRODPOPR | 43. | EQUILFAN | 73. | @AC_TRES |
| 14. | PRODCAPF | 44. | ENDETAN | 74. | INC_IMMO |
| 15. | PRODKINV | 45. | LIQGAN | 75. | CR_BX |
| 16. | RENTAECO | 46. | LIQRAN | 76. | VA_BX |
| 17. | PERFORMN | 47. | CAPRBAN | | |
| 18. | RENTANET | 48. | SHAREAN | | |
| 19. | TXVA | 49. | NETAN | | |
| 20. | PARTSALA | 50. | EQFIN | | |
| 21. | PARTETAT | 51. | SHARLIQR | | |
| 22. | PARTPRET | 52. | NTASSTUR | | |
| 23. | PARTAUTO | 53. | DFCT_DCT | | |
| 24. | RETRTTAS | 54. | EH_DCT | | |
| 25. | STOKTRNO | 55. | RC_DCTAN | | |
| 26. | LNSTOKTU | 56. | DCT_BILA | | |
| 27. | COLLPERI | 57. | AC_CHEXA | | |
| 28. | CRDTPERI | 58. | AC_IMAN | | |
| 29. | CSTEMPLT | 59. | DCT_IMAN | | |
| 30. | LNCSTEMP | 60. | RC_DCT | | |

## B.2 Fichier *commerce.xls*

| | | |
|---|---|---|
| 1. INDIC | 31. TXEXPOR | 61. RC_DCTAN |
| 2. AGE | 32. TXACHATS | 62. DCT_BILA |
| 3. ENDET | 33. TXCHPERS | 63. AC_CHEXA |
| 4. AUTOFIN | 34. TXTOTBIL | 64. AC_IMAN |
| 5. DAMORTIC | 35. TXACTFIM | 65. DCT_IMAN |
| 6. LIQGEN | 36. TXVA1 | 66. RC_DCT |
| 7. LIQRED | 37. CFLW_DTT | 67. DCT_BIL |
| 8. ENDETGL | 38. CAP_DTT | 68. AC_CHEX |
| 9. CAPAREMB | 39. @ACTC_T | 69. AC_IMMO |
| 10. CAPAAUTO | 40. BNFTBL | 70. DCT_IMMO |
| 11. CVCABFR | 41. EH_DTT | 71. AC_DT |
| 12. EXPORT | 42. R1 | 72. DT_CA |
| 13. PRODPOPR | 43. R2 | 73. EBE_VA |
| 14. PRODCAPF | 44. R2ANN | 74. FINACAN |
| 15. PRODKINV | 45. R3 | 75. FINAC |
| 16. RENTAECO | 46. R4 | 76. CVCAFRNA |
| 17. PERFORMN | 47. R5 | 77. CVCAFRN |
| 18. RENTANET | 48. SCORE_C_ | 78. CVCABFRN |
| 19. TXVA | 49. EQUILFIN | 79. @_AC_TRE |
| 20. PARTSALA | 50. ENDETAN | 80. INC_IMMO |
| 21. PARTETAT | 51. LIQGAN | 81. CR_BX |
| 22. PARTPRET | 52. LIQRAN | 82. VA_BX |
| 23. PARTAUTO | 53. CAPRBAN | |
| 24. PROFTMAR | 54. SHAREAN | |
| 25. RETRNTTA | 55. NETAN | |
| 26. STOKTRNO | 56. EQFIN | |
| 27. COLLPERI | 57. SHARLIQR | |
| 28. CRDTPERI | 58. NTASSTUR | |
| 29. CSTEMPLT | 59. DFCT_DCT | |
| 30. TXCANE | 60. EH_DCT | |

## B.3 Fichier *ct.xls*

| Ratios | Définition économique |
|---|---|
| R1 | Sens du levier financier |
| R2 | (Réserves + résultat reporté) / passif total hors comptes de régularisation |
| R3 | (Autres placements de trésorerie + valeurs disponibles) / actif total |
| R4 | Dettes échues envers le fisc et l'ONSS |
| R5 | (Stocks et commandes en cours d'exécution + créances à un an au plus - dettes commerciales à un an au plus - acomptes reçus sur commandes à un an au plus - dettes fiscales, salariales et sociales) / actif total |
| R6 | Résultat d'exploitation après amortissements / actifs d'exploitation hors comptes de régularisation |
| R7 | Dettes à un an au plus envers des établissements de crédit / dettes à un an au plus |
| R8 | Dettes garanties / dettes à plus d'un an et à un an au plus |

## B.4 Fichier *mt.xls*

| Ratios | Définition économique |
|---|---|
| R1 | (Réserves + résultat reporté) / passif total hors comptes de régularisation |
| R2 | Nombre de jours entre la date de clôture et la date du dépôt des comptes annuels |
| R3 | Dettes échues envers le fisc et l'ONSS |
| R4 | (Résultat brut - investissements en immobilisations corporelles et financières) / actif total |
| R5 | (Créances sur entreprises liées + garanties consenties en leur faveur + autres engagements financiers significatifs pris en leur faveur) / actif total |
| R6 | Dettes à plus d'un an et à un an au plus / passif total hors comptes de régularisation |

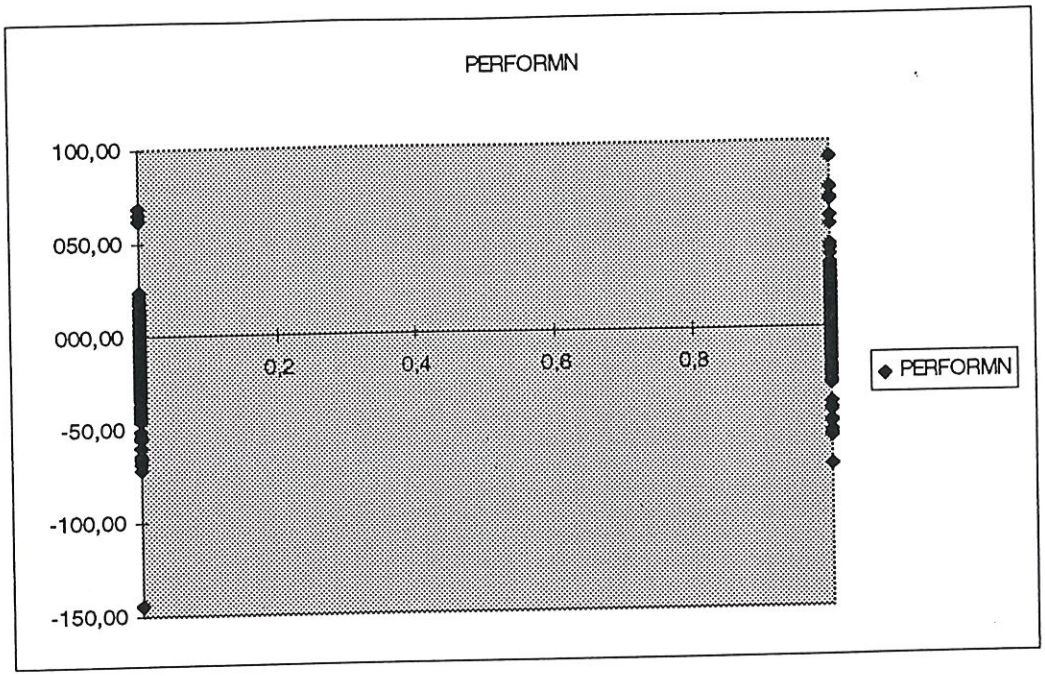
## B.5 Fichier *modif.xls*

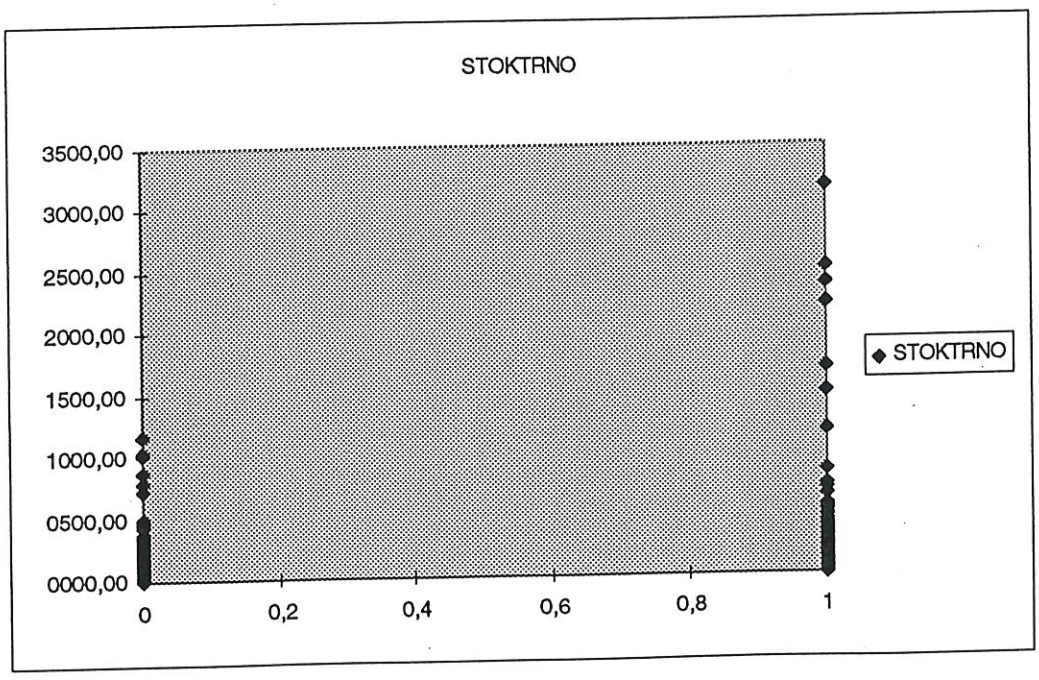
| Ratios | Définition économique |
|---|---|
| R1 | (résultat reporté + réserves) / actif total |
| R2 | dettes échues envers le fisc et l'ONSS / fonds de tiers à court terme |
| R3 | valeurs disponibles / actifs circulants restreints |
| R4 | (en-cours de fabrication + stocks de PF + commandes en cours d'exécution) / actifs circulants d'exploitation |
| R5 | dettes à un an au pus envers les établissements de crédit / fonds de tiers à court terme |
| R6 | actifs circulants restreints / fonds de tiers à court terme |
| R7 | (actifs circulants - stocks) / dettes a court terme |
| R8 | prix de revient des ventes / stocks et commandes en cours d'exécution |
| R9 | créances commerciales à un an au plus / chiffre d'affaires *365 |
| R10 | dettes commerciales à un an au plus / (achats de marchandises + achats de S&B divers + TVA) |
| R11 | fonds propres / passif total |
| R12 | fonds propres / capitaux permanents |
| R13 | exédent brut d'exploitation / (charges d'intérêt + remboursements) |
| R14 | résultats d'exploitation / charges d'intérêt |
| R15 | résultat net / actif total *100 |
| R16 | résultat net / fonds propres *100 |
| R17 | valeur ajoutée brute / valeurs de production *100 |

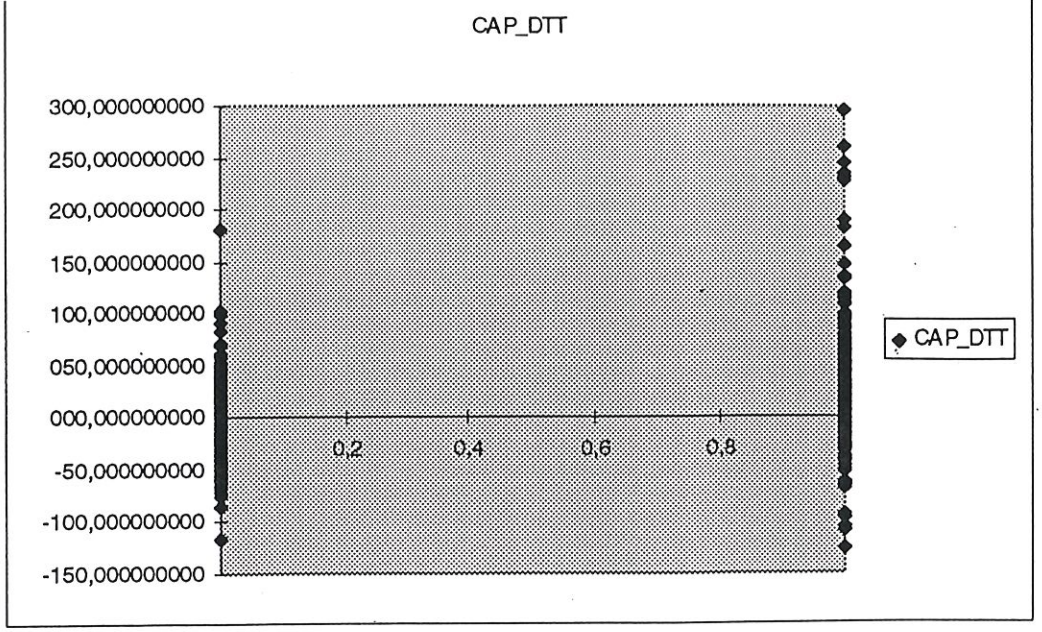
# Annexe C

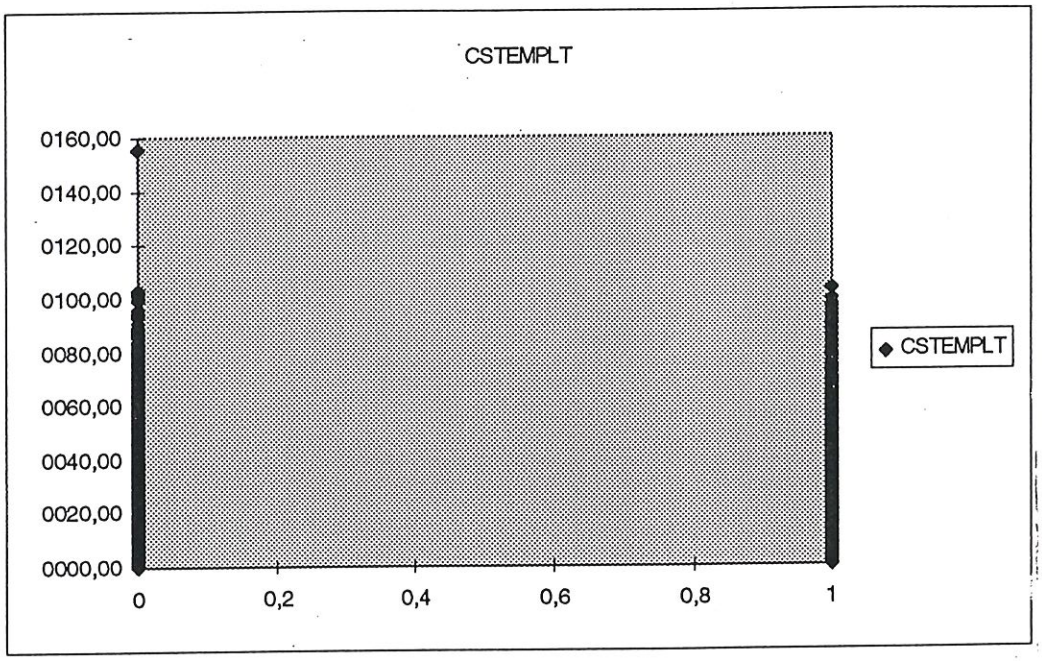
# Représentation des ratios de indtrans.xls et comtrans.xls

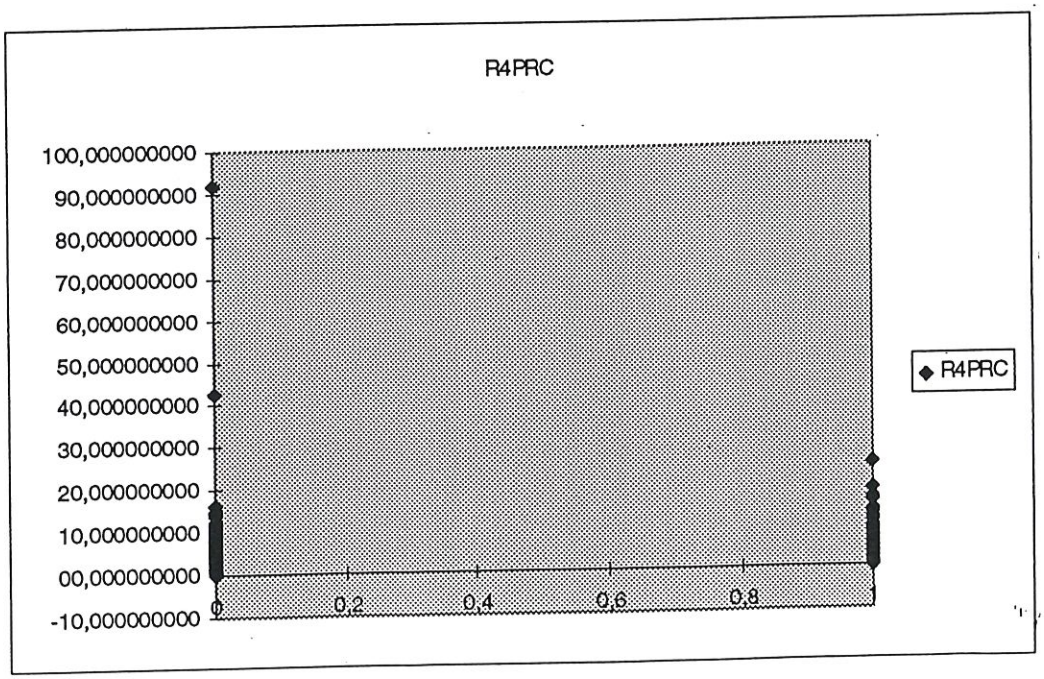
## C.1 *indtrans.xls*

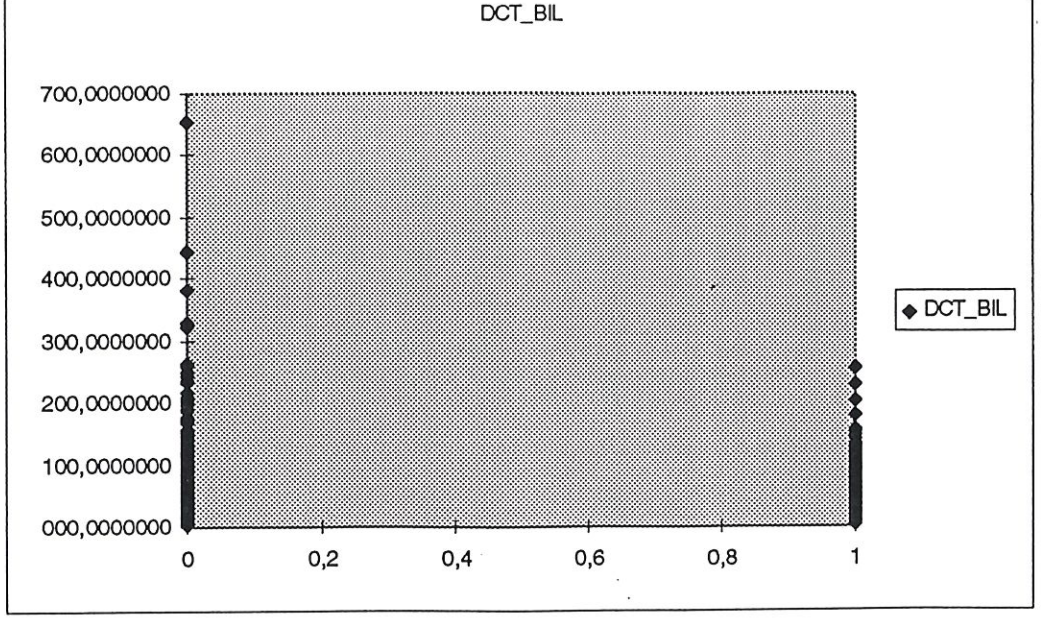


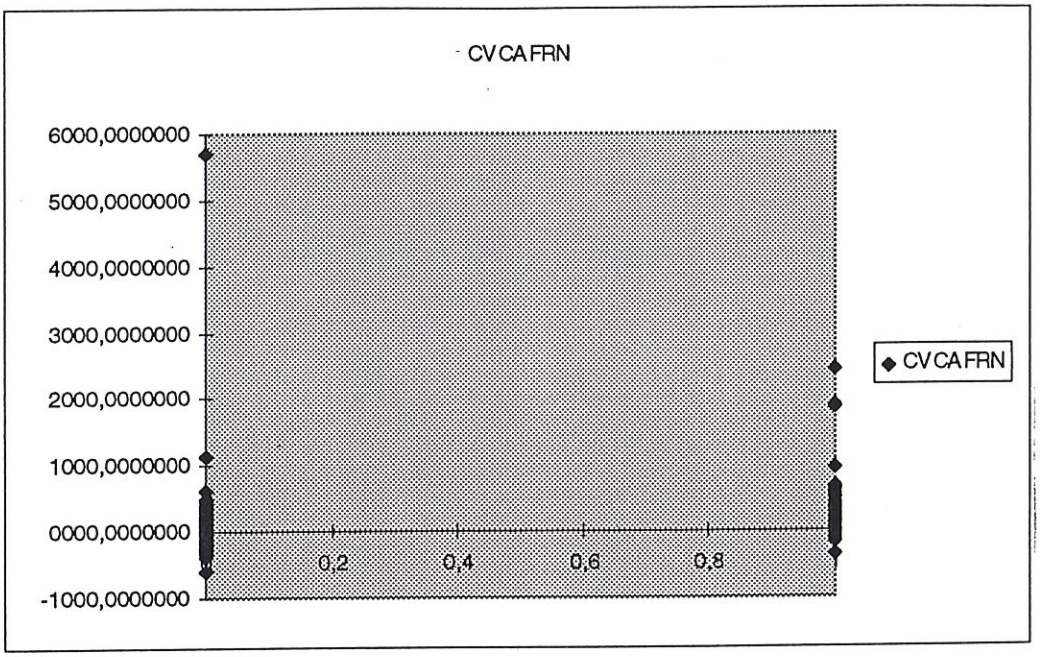



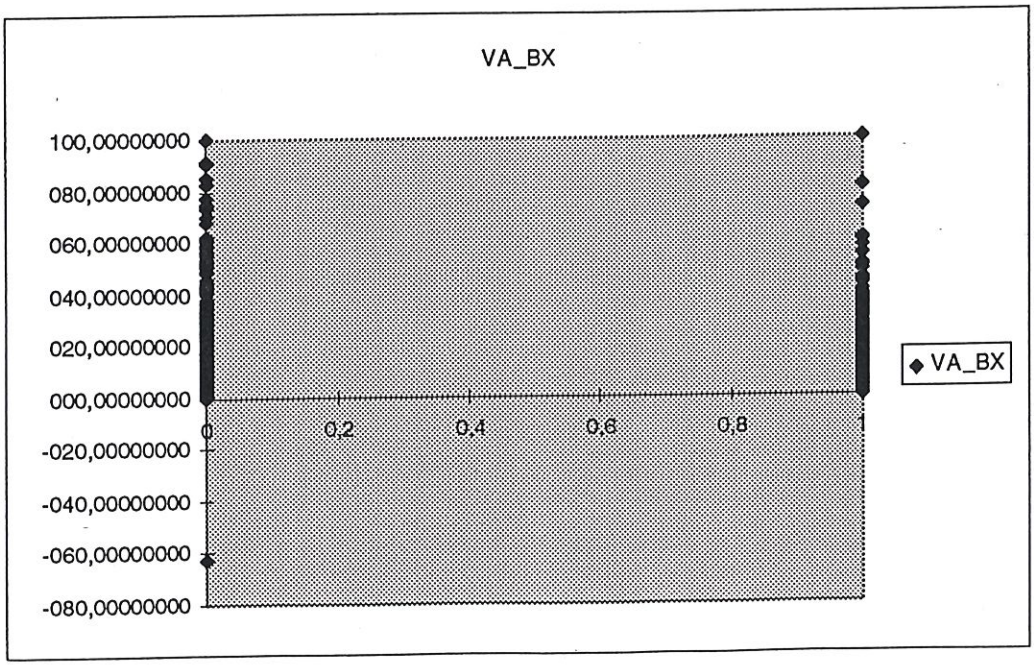

CAP_DTT
300,000000000
250,000000000
200,000000000
150,000000000
100,000000000
050,000000000
000,000000000
-50,000000000
-100,000000000
-150,000000000
0,2
0,4
0,6
0,8
CAP_DTT


CSTEMPLT
0160,00
0140,00
0120,00
0100,00
0080,00
0060,00
0040,00
0020,00
0000,00
0
0,2
0,4
0,6
0,8
1
CSTEMPLT

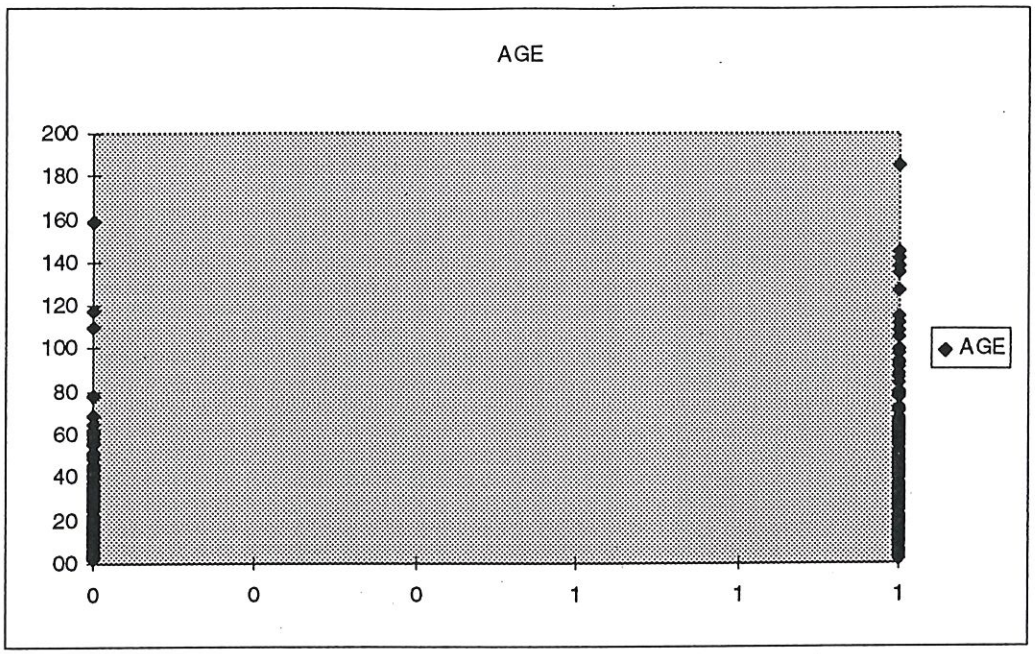

R4PRC
100,000000000
90,000000000
80,000000000
70,000000000
60,000000000
50,000000000
40,000000000
30,000000000
20,000000000
10,000000000
00,000000000
-10,000000000
0
0,2
0,4
0,6
0,8
R4PRC

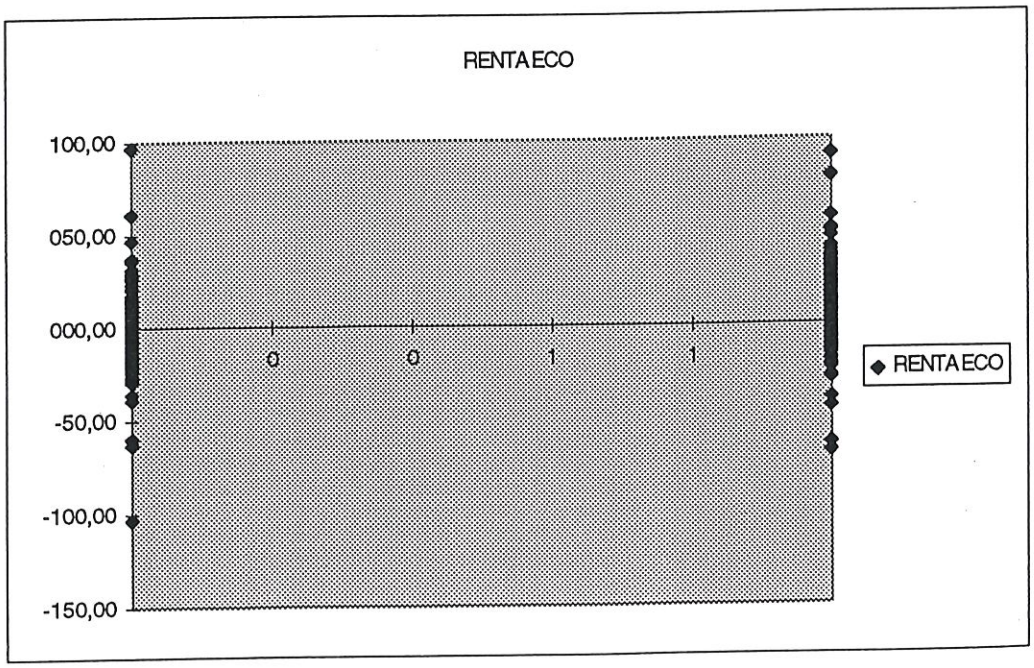

DCT_BIL
700,0000000
600,0000000
500,0000000
400,0000000
300,0000000
200,0000000
100,0000000
000,0000000
0
0,2
0,4
0,6
0,8
1
DCT_BIL

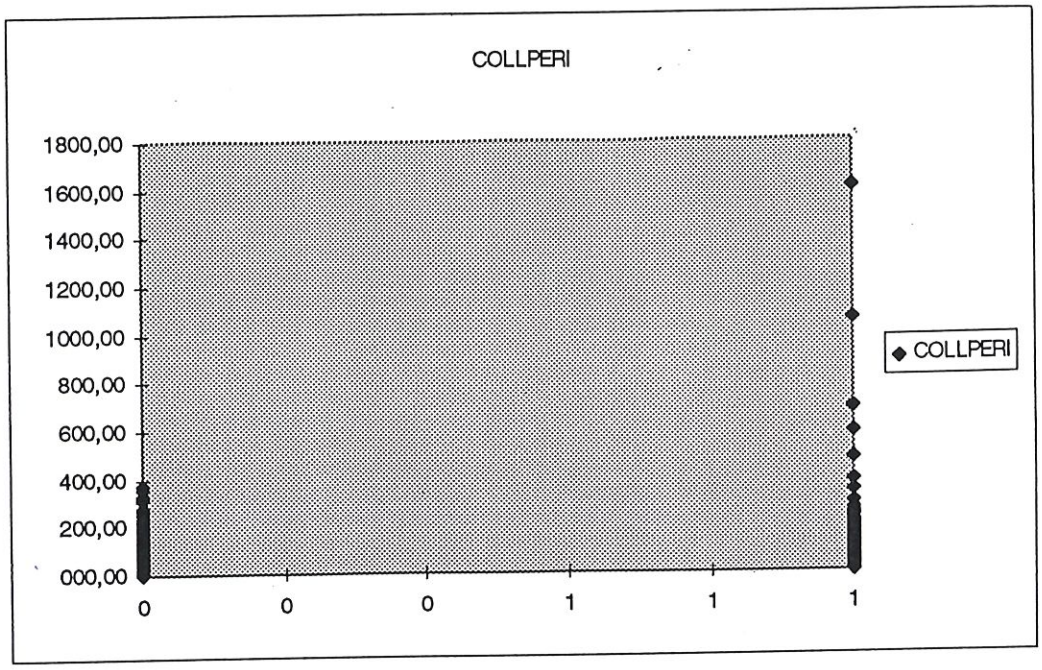

CVCAFRN
6000,0000000
5000,0000000
4000,0000000
3000,0000000
2000,0000000
1000,0000000
0000,0000000
-1000,0000000
0,2
0,4
0,6
0,8
CVCAFRN

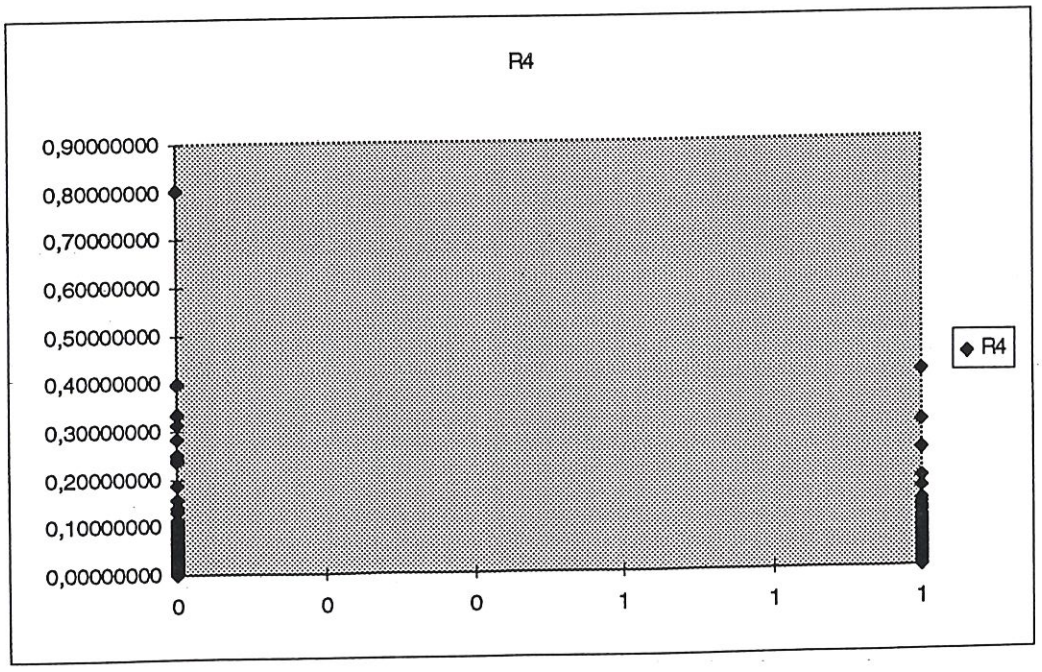

VA_BX
100,00000000
080,00000000
060,00000000
040,00000000
020,00000000
000,00000000
-020,00000000
-040,00000000
-060,00000000
-080,00000000
0
0,2
0,4
0,6
0,8
1
VA_BX

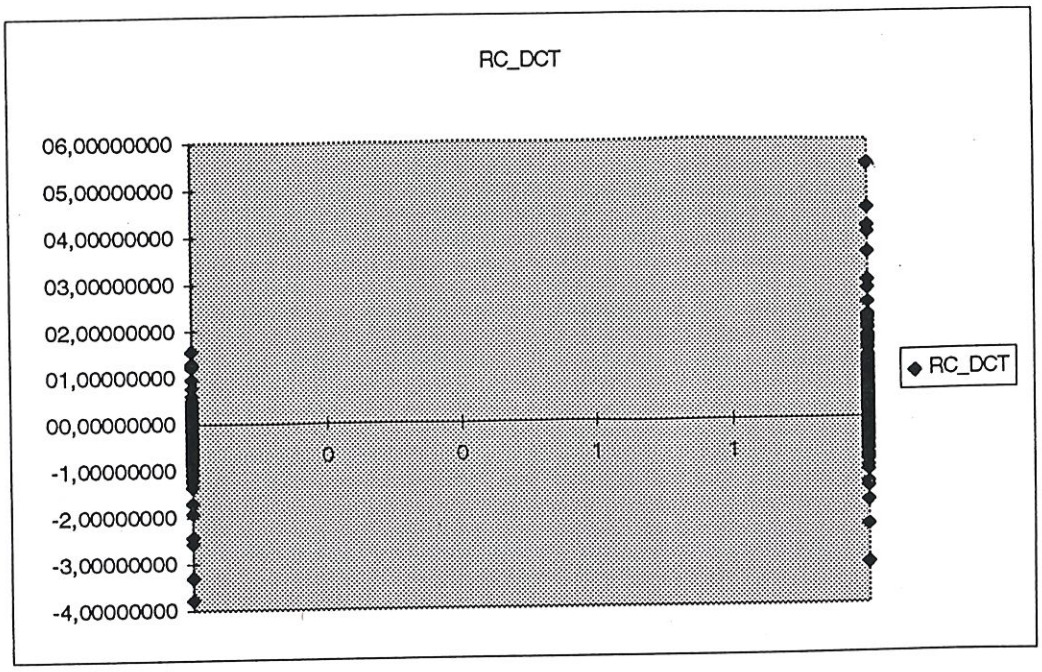
## C.2 *comtrans.xls*

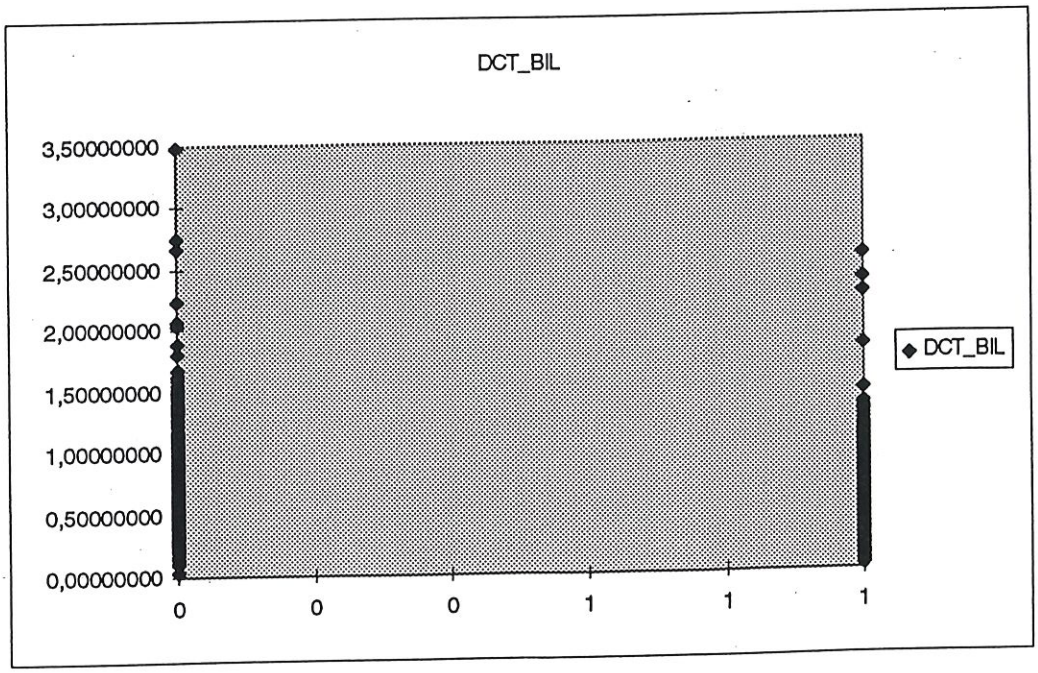



COLLPERI
1800,00
1600,00
1400,00
1200,00
1000,00
800,00
600,00
400,00
200,00
000,00
0
0
0
1
1
1
COLLPERI


R4
0,90000000
0,80000000
0,70000000
0,60000000
0,50000000
0,40000000
0,30000000
0,20000000
0,10000000
0,00000000
0
0
0
1
1
1
R4


RC_DCT
06,00000000
05,00000000
04,00000000
03,00000000
02,00000000
01,00000000
00,00000000
-1,00000000
-2,00000000
-3,00000000
-4,00000000
0
0
1
1
RC_DCT


DCT_BIL
3,50000000
3,00000000
2,50000000
2,00000000
1,50000000
1,00000000
0,50000000
0,00000000
0
0
0
1
1
1
DCT_BIL