

www.Mcours.com

Site N°1 des Cours et Exercices Email: contact@mcours.com

MÉTHODES STATISTIQUES EN FINANCE

Pierre Clauss

Institut de Gestion de Rennes
Master 1 Finance

OBJECTIF DE L'ENSEIGNEMENT

Ce cours de 15h est composé de 5 sessions de 3h dont 2 sessions de cours magistral et 3 sessions d'atelier sur un projet en groupe.

Ce cours de Méthodes Statistiques en Finance a pour double objectif de dépasser la méthode des Moindres Carrés Ordinaires pour l'estimation des modèles de régression linéaire simple en familiarisant l'étudiant avec d'autres approches et notamment celle du maximum de vraisemblance ; ainsi que de découvrir la régression logistique, largement utilisée en Finance dans le cadre du *credit scoring* par exemple.

Ce cours permettra d'approfondir les techniques de modélisation statistique étudiées en Licence et d'acquérir des réflexes importants en statistiques dans la perspective des cours de Master 2 Finance. Des applications seront en outre proposées sur des données réelles à l'aide du logiciel Excel.

TABLE DES MATIÈRES

INTRODUCTION	4
1 ÉLÉMENTS DE THÉORIE FINANCIÈRE	5
1.1 CAPM	5
1.1.1 Modèle	5
1.1.2 Portefeuille de marché	6
1.1.3 Présence d'alpha	6
1.2 Credit scoring	7
1.2.1 Définition du risque de crédit	7
1.2.2 Notation des agences	7
1.2.3 Scoring des émetteurs	9
2 RÉGRESSION LINÉAIRE SIMPLE	10
2.1 Exemple introductif	10
2.2 Estimation	11
2.2.1 Avant toute estimation : éviter les régressions fallacieuses	11
2.2.2 Hypothèses	13
2.2.3 Moindres Carrés Ordinaires	13
2.2.4 Méthode des Moments	15
2.2.5 Méthode du Maximum de Vraisemblance	16
2.2.6 Propriétés des estimateurs	17
2.2.7 Estimation récursive versus estimation glissante	19
2.3 Validation du modèle	20
2.3.1 Coefficient de détermination	20
2.3.2 Tests statistiques	20
2.3.3 Backtest	24
3 RÉGRESSION LOGISTIQUE	25
3.1 Inadéquation des MCO	25
3.2 Introduction d'une variable latente	26
3.3 Estimation par Maximum de Vraisemblance	26
CONCLUSION	28

INTRODUCTION

Estimer un paramètre de régression est une chose assez répandue dans de nombreux métiers aujourd'hui pour aider à la décision et mieux comprendre son environnement. Néanmoins, entre utiliser la méthode et la comprendre ainsi que bien l'appliquer, il existe parfois un décalage important. En effet, il semble essentiel de comprendre d'où vient l'estimateur et d'en comprendre les hypothèses de mise en oeuvre pour ne pas faire des erreurs basiques dommageables. C'est l'objectif essentiel de ce cours : aller un peu plus dans la boîte noire de l'estimation de modèles statistiques simples et très usités et les appliquer sur des données financières. Car ce domaine d'application est très exigeant et a montré certaines faiblesses durant la crise, entre autres dans des problématiques d'estimation (nous pensons à la titrisation de dettes subprimes parfois mal calibrées).

Estimer c'est tenter, à partir d'une modélisation que l'on considère adéquate, de déterminer le plus précisément possible la valeur du paramètre de ce modèle. Une action française va réagir à l'indice phare du marché boursier français CAC 40. Nous pouvons avoir l'intuition que cette relation est linéaire. Le modèle, qui peut être justifié par la théorie financière, nous l'étudierons plus bas, énonce alors que la rentabilité de l'action est sensible linéairement à celle du CAC 40. Ainsi, il va incomber au statisticien de déterminer cette sensibilité. Ce dernier ne remet pas en cause le modèle précédent mais cherche à estimer rigoureusement la sensibilité. Bien entendu, le modèle peut s'avérer peu adéquat et des méthodes statistiques peuvent alors permettre d'en proposer de plus judicieux.

Plusieurs étapes sont indispensables à l'estimation d'un modèle :

1. spécifier le modèle financier, économique, sociologique, physique, etc.,
2. en tirer un modèle statistique (ou économétrique),
3. vérifier les hypothèses essentielles sur les données réelles pour appliquer l'estimation,
4. estimer,
5. déterminer les propriétés de l'estimateur : écart relatif à la vraie valeur inconnue (biais), dispersion et loi de l'estimateur,
6. valider le pertinence du modèle sur les données.

Ce cours sera composé de trois parties.

La première s'attachera à développer certains éléments de théorie financière (CAPM et credit scoring) qui nous permettront de déterminer des modèles financiers que l'on pourra alors estimer. Après avoir rappelé la Méthode des Moindres Carrés Ordinaires, la deuxième partie approfondira les techniques de régression linéaire simple avec la Méthode des Moments et le Maximum de Vraisemblance. Enfin, la dernière partie développera la régression logistique, traitant de l'explication d'une variable qualitative par des variables continues.

CHAPITRE 1

ÉLÉMENTS DE THÉORIE FINANCIÈRE

Nous allons nous attacher à étudier dans ce cours deux modélisations financières classiques :

1. Le Capital Asset Pricing Model (CAPM) permettant de théoriser la formation de toute rentabilité d'un titre financier ou d'un portefeuille de titres par leur sensibilité au risque de marché.
2. La notation des émetteurs de dette (pays, entreprises, particuliers) dont on va modéliser plus particulièrement la probabilité de défaillance. Cet aspect est essentiel dans le cadre de la réglementation prudentielle issue du comité de Bâle.

1.1 CAPM

1.1.1 Modèle

Le Capital Asset Pricing Model (CAPM) ou Modèle d'Équilibre des Actifs Financiers (MEDAF) a été développé par Sharpe en 1964 [12], Lintner en 1965 [9] et Mossin en 1966 [11]. Les hypothèses de ce modèle sont les suivantes :

1. les investisseurs exigent une rentabilité d'autant plus forte que le risque est élevé : il existe donc une relation croissante entre rendement et risque,
2. un actif sans risque est disponible,
3. les anticipations sont identiques pour tous les investisseurs.

D'après ces hypothèses, nous pouvons exprimer l'espérance de la rentabilité r_i d'un portefeuille ou d'un actif risqué i en fonction de celle de l'actif sans risque r_f et de celle du portefeuille de marché r_M , qui est celui que tous les investisseurs possèdent :

$$\mathbb{E} r_i = r_f + \beta_i (\mathbb{E} r_M - r_f)$$

Nous pouvons aussi réécrire ce modèle de la manière suivante :

$$\rho_i = \mathbb{E} r_i - r_f = \beta_i (\mathbb{E} r_M - r_f) \quad (1.1)$$

pour faire apparaître la prime de risque ρ_i . Ainsi, plus les investisseurs seront exposés au portefeuille de marché, plus ils prendront de risque et plus leur rémunération consécutive potentielle sera théoriquement élevée.

Ce modèle explique donc la prime de risque d'un actif ou d'un portefeuille d'actifs i par sa sensibilité β_i par rapport au portefeuille de marché. Cette dernière dénote le caractère plus ou moins agressif d'un actif relativement au portefeuille de marché. Ainsi, si $\beta_i > 1$, l'actif est dit "offensif", et si $\beta_i < 1$, il est dit "défensif" relativement au marché considéré ($\beta_i = 1$ correspond à une prise de risque similaire à celle prise

par le portefeuille de marché). L'indicateur β est très utilisé par les praticiens car simple d'utilisation.

Cependant, outre les hypothèses restrictives du modèle d'équilibre, il possède aussi quelques limites de mise en oeuvre empiriques.

1.1.2 Portefeuille de marché

Il semble difficile de déterminer précisément le portefeuille de marché. Souvent, il est réduit à l'indice phare de la place financière dans laquelle est évalué l'actif ou le portefeuille. Seulement, ces indices n'intègrent pas tous les actifs risqués de l'univers d'investissement, comme le voudrait la théorie, puisqu'il se restreint aux actions.

Les poids alloués aux actions dans ces indices varient suivant les places financières. Ils peuvent être relatifs :

- à la valeur des titres : c'est le cas du plus vieil indice fondé en 1884, le Dow Jones Industrial Average composé de 30 entreprises américaines importantes,
- aux capitalisations boursières,
- ou au flottant, défini comme la part de la capitalisation que l'on peut échanger sur les marchés (inférieur le plus souvent à la capitalisation boursière totale), traduisant donc la liquidité du titre. La pondération par le flottant est le cas de la plupart des indices aujourd'hui. Ainsi, le CAC 40 l'est depuis le 1er décembre 2003, après avoir été pondéré par les capitalisations.

Les créances bancaires, l'immobilier, le capital humain manquent donc dans la composition de ces indices et sont difficiles à mesurer précisément.

1.1.3 Présence d'alpha

Le problème traité par ce paragraphe est la significativité possible d'une constante α_i dans le modèle statistique de marché, qui s'écrit :

$$r_i - r_f = \alpha_i + \beta_i (r_M - r_f) + \varepsilon_i$$

Théoriquement, le CAPM implique l'Absence d'Opportunité d'Arbitrage (AOA) entre les actifs, c'est-à-dire qu'en théorie la relation (1.1) est toujours vérifiée. Si cela n'est pas le cas, l'actif est mal évalué et selon l'Hypothèse d'Efficiency des Marchés (HEM), l'anomalie sera résorbée rapidement par les arbitrageurs et ne devrait pas perdurer.

Néanmoins, dans la réalité, des α_i significativement non nuls perdurent plus longtemps que prévu par l'HEM. Cela peut alors signifier deux choses :

- Soit le modèle n'est pas à remettre en cause et alors la prime de risque observée ρ_i^{obs} est arbitrageable mais les arbitrageurs de la théorie ne sont pas aussi efficaces que prévu. Alors une rentabilité *gratuite* apparaît dans le cas où α_i est par exemple positif : $\rho_i^{\text{obs}} = \alpha_i + \rho_i^{\text{th}} \Leftrightarrow \rho_i^{\text{th}} = \rho_i^{\text{obs}} - \alpha_i$. Le modèle est supposé bon, donc la prime de risque ρ_i^{obs} devrait diminuer pour atteindre la prime de risque théorique ρ_i^{th} et donc le prix de l'action augmenter (prime de risque au dénominateur du prix de l'action). L'action est bien sous-évaluée. Le marché est alors inefficace pour l'actif i considéré. La rentabilité sera élevée puisque une partie *gratuite* α_i est présente.
- Soit la prime de risque n'est pas entièrement décrite par sa seule sensibilité au portefeuille de marché. D'autres facteurs de risque sont certainement évalués par les intervenants des marchés. Fama et French [6] ont ainsi proposé un modèle d'équilibre plus réaliste complétant le CAPM des autres facteurs de risque pouvant être rémunérés par la prime de risque (capitalisation boursière et valorisation).

Pour étudier la significativité ou non des alphas, il faudra estimer le modèle et c'est ce que nous étudierons dans le second chapitre de ce cours.

1.2 Credit scoring

1.2.1 Définition du risque de crédit

Aujourd'hui, la gestion du risque pris par les banques est devenue fondamentale. Alors que le risque de marché est défini comme le risque de perte lié aux variations des marchés (nous en avons étudié une explication dans la section précédente : la chute du portefeuille de marché), le risque de crédit est lié aux variations de la qualité de la signature d'un émetteur. Ainsi, un prêteur X est soumis via l'achat d'une obligation émise par Y au risque que ce dernier ne rembourse pas le capital investi, à la maturité de l'obligation. Si Y ne peut pas faire face à ses engagements, cela signifie que Y a fait *défait*. Les exemples historiques les plus marquants de défaut sont ceux d'Enron, Parmalat, et plus récemment les ménages *subprimes* américains. Tout l'enjeu est alors de modéliser la défaillance d'un émetteur de dette plus ou moins risquée dans le but de la mesurer.

Le risque de défaillance se mesure à travers différents événements que peut subir un émetteur : l'évolution possible de sa notation ou *rating* (*upgrade* ou *downgrade*) émis par les agences de notation telles Moody's et Standard & Poor's par exemple, la modification de la cotation de l'entreprise, la variation du spread de crédit, le défaut de paiement, la faillite. Ces événements peuvent influencer plus ou moins sur la qualité de la signature de l'émetteur, augmentant alors, dans le cas d'une dégradation de la signature, le risque de défaut de l'entreprise.

Outre la probabilité de défaillance, deux autres paramètres interviennent dans la mesure du risque de crédit : le taux de recouvrement, qui quantifie ce que récupère le créancier lorsque l'entreprise fait défaut, et la dépendance entre les défauts des émetteurs.

Pour modéliser la probabilité de défaut, trois possibilités s'offrent à nous :

- l'approche par les ratings et par les évolutions en terme de notations des émetteurs : à chaque *rating* est associée une probabilité de défaut d'autant plus élevée que le *rating* est mauvais,
- l'approche *structurelle* modélisant le défaut en prenant en compte le bilan comptable de l'entreprise et le processus financier conduisant au défaut (utilisation des résultats de Merton [10]),
- et l'approche sous *forme réduite* modélisant le défaut sans référence au processus financier sous-jacent. Le défaut est imprévisible et sa loi est gouvernée par un processus stochastique (de Poisson à intensité constante le plus souvent).

Nous allons nous intéresser dans cette section à l'approche par les notations. Cette méthodologie s'appuie sur le constat qu'un événement de défaut est consécutif à une dégradation progressive des ratings de l'émetteur. En outre, elle fait partie des fondements des réglementations du Comité de Bâle. Deux approches sont possibles : la première élabore des notations à l'aide d'opinions d'experts et la seconde utilise les techniques de scoring.

1.2.2 Notation des agences

Les ratings des agences de notation sont déterminés sur la base d'opinions indépendantes, objectives, crédibles et transparentes. Cette opinion qualitative est en général présentée sous la forme d'une ou plusieurs lettres symbolisant la qualité de crédit de l'émetteur. L'ensemble de ces notations forme l'échelle de notation. Les ratings, qui s'appliquent aussi bien à l'émetteur qu'à l'instrument de dette émis, varient selon l'horizon auquel ils s'appliquent. Nous distinguons des ratings court terme (d'horizon égal ou inférieur à 1 an) et des ratings de long terme. Les agences ont des échelles différentes mais un point commun essentiel est la distinction entre les deux catégories *investment grade* et *sub-investment grade* ou *speculative grade* (cf. tableau 1.1). La première catégorie rassemble des firmes ayant une relative stabilité dans leurs modèles de développement et un niveau de risque modéré. La seconde catégorie inclut des émetteurs aux caractéristiques financières incertaines dont la probabilité de faire défaut est plus élevée.

Les ratings sont issus de la compétitivité de la firme, la qualité du management et de la politique suivie, de la situation du marché de la firme, ainsi que de ratios financiers, dépendant aussi du secteur de la société. Ils font l'objet d'un suivi régulier afin de refléter au mieux l'évolution du client.

	Moody's	Stanbard & Poor's
<i>Investment grade</i>	Aaa	AAA
	Aa	AA
	A	A
	Baa	BBB
<i>Speculative grade</i>	Ba	BB
	B	B
	Caa	CCC

TABLE 1.1 – Ratings de Moody's et Stanbard & Poor's

Un point important est le fait que le rating ne reflète pas la cherté d'une dette mais le jugement sur la qualité de la signature d'une dette et sa capacité de remboursement.

Les probabilités de défaut, en fonction de la zone géographique, du secteur, de la période, de l'horizon, correspondent à la moyenne des fréquences de défaut par rapport à l'ensemble des firmes de même rating.

Les agences de notation fournissent aussi des matrices de transition indiquant pour un horizon donné la probabilité de migration qu'un émetteur ou instrument appartenant à une classe de rating passe dans une autre classe en cours de période. Mais ces matrices sont-elles stables dans le temps et permettent-elles ainsi d'anticiper les migrations futures ?

Si c'est le cas, nous pouvons faire l'hypothèse que les matrices de transition sont markoviennes, c'est-à-dire que la seule connaissance de l'information en t permet de déterminer les valeurs de la matrice en $t + 1$. Cette hypothèse est néanmoins très discutée (cf. de Servigny, Métayer et Zelenko [5] pour une revue des débats en cours).

Nous supposons que l'on dispose de 7 niveaux de rating distincts hors défaut soit 8 états pour les créances. On note l'espace état $I = \{AAA, AA, A, BBB, BB, B, CCC, D\}$ dénombrable et fini avec D la classe des titres en défaut. On note $P = P(0, 1) = (p_{ij})_{i,j \in I \times I}$ la matrice 8×8 de transition à 1 an et nous remarquons que P a la forme suivante, avec en lignes les ratings en t et en colonnes les ratings en $t + 1$:

$$P = P(0, 1) = \begin{pmatrix} R & L \\ 0 & 1 \end{pmatrix}$$

avec R un matrice de taille 7×7 de transition d'une classe de crédit à l'autre, L un vecteur 7×1 de probabilités de défaut, 0 le vecteur 1×7 de probabilités nulles de sortir de l'état de défaut, et la valeur 1 traduit l'état absorbant de défaut. P étant telle que $0 \leq p_{ij} < \infty \forall i, j$ et $\sum_j p_{ij} = 1 \forall i$, la matrice de transition est qualifiée de *stochastique*.

Les probabilités de transition sont modélisées à l'aide d'un chaîne de Markov à états finis.

Définition 1.2.1. *Un processus $X = (X_n)_n$ est une chaîne de Markov homogène à états finis discrets I de matrice de transition P si on a :*

$$\mathbb{P}[X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i] = \mathbb{P}[X_{n+1} = j | X_n = i] = p_{ij}$$

Propriété 1.2.1. *L'équation de Chapman-Kolmogorov implique que :*

$$P(0, m + n) = P(0, n)P(0, m)$$

ou encore :

$$P(0, n) = P(0, 1)^n = P^n = \left(p_{ij}^{(n)} \right)_{i,j}$$

avec $P(0, n)$ la matrice de transition déterminée aujourd'hui pour un horizon de n années, et $p_{ij}^{(n)}$ l'élément (i, j) de la matrice P^n (probabilité de passer de i à j en n années). Nous avons aussi :

$$\mathbb{P}[X_{n+m} = j | X_m = i] = p_{ij}^{(n)}$$

Les probabilités de défaut à n'importe quel horizon sont donc déterminées grâce à la modélisation de la matrice de transition à un an par chaîne de Markov homogène.

1.2.3 Scoring des émetteurs

En général, les ratings des agences sont faits sur des États et entreprises de grande taille. Concernant les petites et moyennes entreprises ou encore les particuliers, les probabilités de défaut peuvent être estimées à l'aide de méthodes de scoring.

Les variables explicatives, issues d'informations comptables, sont des ratios économiques et financiers statistiquement significatifs pour différencier les entreprises susceptibles d'être en difficulté de celles qui sont plus saines (performances, solvabilité, autonomie financière, etc.). Un *mapping* spécifique permet de relier le scoring continu et l'échelle de notation discrète.

Deux techniques principales de scoring sont utilisées pour calibrer la probabilité de défaillance : l'analyse discriminante linéaire de Fisher et la régression logistique. Nous étudierons la seconde dans le dernier chapitre de ce cours.

CHAPITRE 2

RÉGRESSION LINÉAIRE SIMPLE

2.1 Exemple introductif

Prenons l'exemple du CAPM, Capital Asset Pricing Model, (étudié en section 1.1). Supposons pour simplifier que les rentabilités du CAC 40, que nous notons x , peuvent être de l'ordre de 3 niveaux : -5% , 0% et 10% . Pour chacun de ces niveaux, nous allons nous intéresser à collecter pour une action cotée sur le marché boursier français la rentabilité mesurée au même moment. Cette rentabilité de l'action que nous notons y est inconnue au départ, c'est donc une variable aléatoire.

En outre, lorsque le CAC 40 va faire 0% , l'action peut prendre des valeurs différentes suivant le temps (cf. tableau 2.1).

	1	2	3	4	5	6	7	8	9	...
CAC 40 x	0%	0%	-5%	10%	-5%	0%	-5%	10%	10%	...
Action y	1%	-0.5%	-9%	5%	-2%	-2%	-7%	8%	12%	...

TABLE 2.1 – Possibles valeurs temporelles de la rentabilité de l'action y suivant la rentabilité du CAC 40

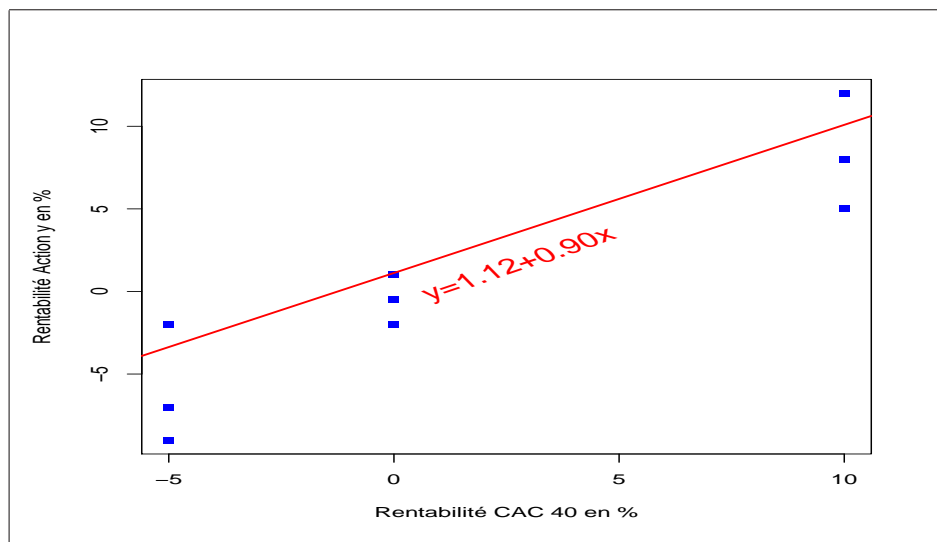


FIGURE 2.1 – Nuage de points entre CAC 40 et action y

Ainsi suivant les valeurs de x , nous obtenons différentes valeurs aléatoires pour y . Si nous notons $\mathbb{E}[y|x]$

l'espérance ou encore la moyenne des rentabilités de l'action y sachant la rentabilité du CAC 40 x , nous pouvons écrire le modèle financier CAPM sous la forme suivante dénommée fonction de régression simple (à une variable) :

$$\mathbb{E}[y|x] = \alpha + \beta x$$

car c'est en moyenne que nous avons un résultat de dépendance entre les 2 variables.

y est la variable dépendante et x est la variable explicative. Entre la moyenne définie par la fonction de régression et la donnée réelle, nous avons un écart appelé l'erreur ou bruit ε . Bien entendu, ce terme d'erreur est aléatoire :

$$\varepsilon = y - \mathbb{E}[y|x] = y - \alpha - \beta x$$

On peut ainsi écrire :

$$y = \mathbb{E}[y|x] + \varepsilon = \alpha + \beta x + \varepsilon$$

ce qui fait distinguer l'explication de la variable y entre la partie systématique $\mathbb{E}[y|x]$ et la partie spécifique ε

A partir de cette fonction de régression simple, nous posons certaines hypothèses qui vont déterminer un modèle statistique ou économétrique. Un critère important à observer aussi est la variance conditionnelle : $\text{Var}[y|x]$ qui détermine la dispersion pour chaque niveau de rentabilité du CAC 40 de celles associées pour l'action y . Cette dispersion est supposée dans le cadre de ce cours comme égale pour chaque niveau. Enfin, le dernier critère d'intérêt est la covariance entre les rentabilités de l'action y à savoir que dans le temps nous les supposons non corrélées : $\text{Cov}(y_i, y_j) = 0$ pour $i \neq j$.

Pour plus d'informations sur la régression linéaire simple, nous renvoyons à Carter Hill, Griffiths & Lim [3] et Tassi [13].

2.2 Estimation

2.2.1 Avant toute estimation : éviter les régressions fallacieuses

Supposons tout d'abord que nous avons un échantillon de données (x_i, y_i) de taille n . Nous notons x la variable aléatoire dont sont issues les données x_i et y pour les données y_i . Nous supposons que les réalisations x_i sont indépendantes et équi-distribuées, ainsi que les réalisations y_i .

Il est donc nécessaire avant de réaliser l'estimation d'un modèle de régression linéaire simple sur cet échantillon de vérifier que les données statistiques x_i et y_i sont stationnaires pour éviter toute régression fallacieuse (*spurious regression*). Une régression fallacieuse apparaît lorsque nous essayons de régresser une variable intégrée d'ordre 1 (non-stationnaire) sur une autre variable non-stationnaire. Ce problème a été mis en évidence par Granger et Newbold [7]. Une régression fallacieuse va conclure à des relations erronées entre variables, car statistiquement invalides. Il est donc primordial de travailler avec des données stationnaires. Rappelons qu'un processus temporel est stationnaire (au sens faible) si sa moyenne, sa variance (qui est finie) et sa covariance entre 2 instants sont indépendantes du temps et le processus est stable *en moyenne* dans le temps.

En Finance, les données sont rarement stationnaires : en effet, les prix P_t des indices boursiers ou des portefeuilles d'investissement ne sont pas stationnaires ; une simple différenciation permet de rendre les données de prix stationnaires : les rentabilités financières R_t le deviennent alors.

$$R_t = \ln \frac{P_t}{P_{t-1}} \approx \frac{P_t}{P_{t-1}} - 1$$

Étude de la stationnarité du S&P 500

Étudions la série temporelle de l'indice S&P 500 (cf. [8] pour plus de précisions). Nous y appliquons au préalable la transformation de Box-Cox [2] et plus particulièrement celle du logarithme, dans le but de *stabiliser* la variance des données. Nous pouvons conclure tout de même à la non-stationnarité de l'indice américain. En effet :

1. la Figure 2.2 révèle une tendance globale à la hausse,
2. elle révèle aussi la décroissance lente des autocorrélations significativement non nulles de la série temporelle des valeurs du S&P 500.

Il est possible de réaliser aussi le test statistique de Dickey-Fuller augmenté concluant à la présence d'une racine unitaire dans le processus de l'indice et donc à sa non-stationnarité. mais cela dépasse le cadre de ce cours.

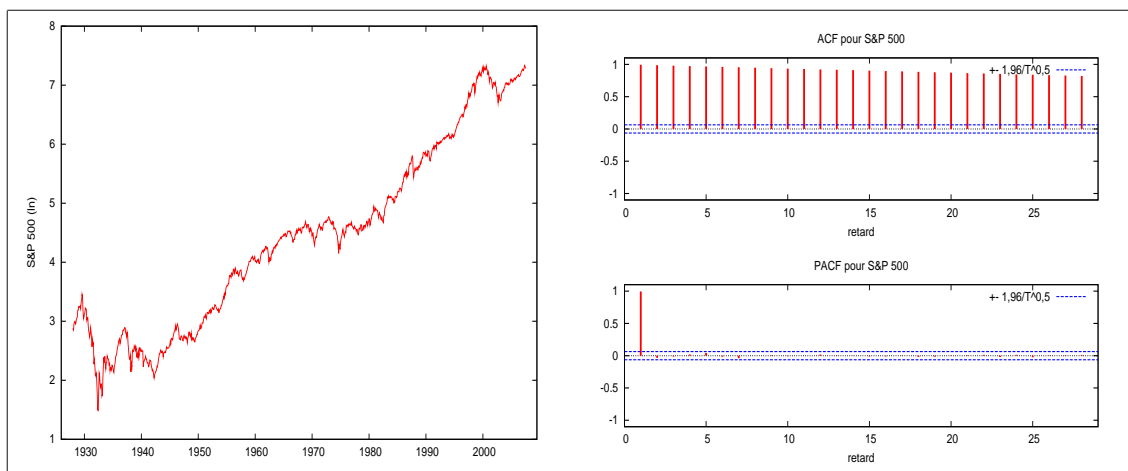


FIGURE 2.2 – Évolution et autocorrélogrammes de l'indice S&P 500 en log

Il nous faut donc stationnariser les données en les différenciant une fois. Ces nouvelles données se révèlent alors stationnaires :

1. pas de tendance dans la série temporelle des différences premières des logarithmes du S&P 500 (cf. Figure 2.3),
2. l'autocorrélogramme de la Figure 2.3 ne révèle pas de décroissance lente des autocorrélations.

Les différences premières de la série des logarithmes du S&P 500 sont équivalentes aux log-rentabilités. Nous comprenons alors pourquoi les rentabilités seront des objets adéquats à la modélisation statistique, puisqu'elles se révèlent être stationnaires *en moyenne*.

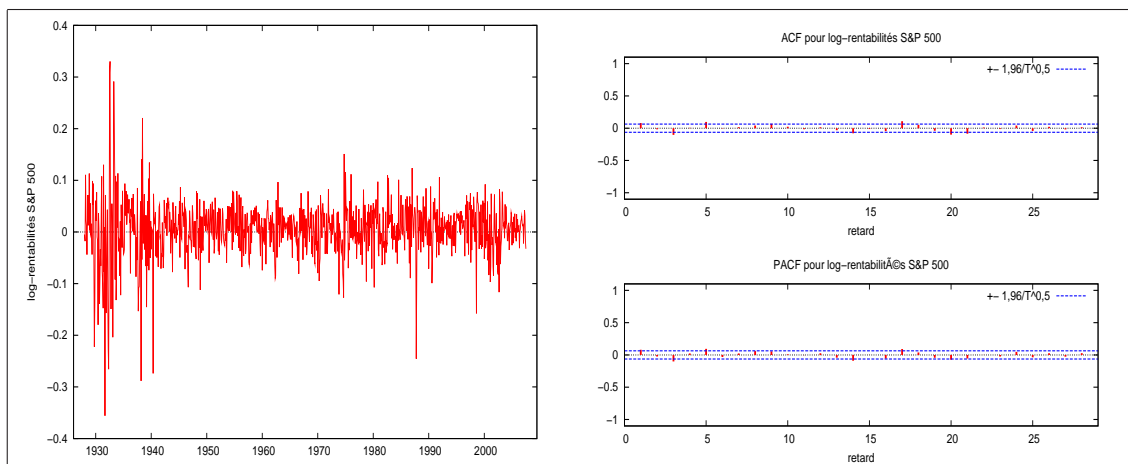


FIGURE 2.3 – Évolution et autocorrélogrammes des différences premières l'indice S&P 500 en log

2.2.2 Hypothèses

Soit le modèle de régression linéaire simple suivant sur notre échantillon (x_i, y_i) :

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad \forall i = 1, \dots, n$$

Une remarque concernant la notation dans ce cours est importante à ce stade : nous utiliserons, dans les hypothèses qui suivront, soit les réalisations x_i, y_i et ε_i , soit les variables dont sont issues ces réalisations x, y et ε , ceci revenant au même. En outre, nous utiliserons les majuscules X et Y lorsque nous traiterons de moments empiriques déterminées sur les réalisations x_i et y_i .

Les hypothèses de ce modèle sont les suivantes :

Hypothèse 2.2.1. $\mathbb{E}[\varepsilon|x] = 0$ (ou $\mathbb{E}\varepsilon = 0$) ce qui est équivalent à retrouver la fonction de régression $\mathbb{E}[y|x] = \alpha + \beta x$. Dans le jargon économétrique, on stipule aussi que la variable x est exogène (y est au contraire endogène).

Hypothèse 2.2.2. $\text{Var}[\varepsilon|x] = \text{Var}[y|x] = \sigma^2$ (ou $\text{Var}\varepsilon = \text{Var}y = \sigma^2$) : en d'autres termes, les données sont homoscédastiques et l'incertitude est la même concernant la relation de régression pour toutes les observations.

Hypothèse 2.2.3. $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$: les erreurs ne sont donc pas auto-corrélées.

Les hypothèses que nous venons de stipuler sont soit sur des moments conditionnels soit des moments inconditionnels. Pourquoi ? Nous abordons ici la différence qu'il existe entre les approches statistique et économétrique de la régression linéaire. En effet, les résultats et la méthode semblent similaires entre les deux approches à la nuance près qu'en statistique, il est supposé que les variables explicatives ne sont pas aléatoires (héritage de la logique des expériences où l'on teste le comportement de y à partir de celui qu'on impose sur x) alors qu'en économétrie, l'expérience est plus difficile à mettre en place, ainsi les variables explicatives sont des variables aléatoires : on travaille alors sur des moments conditionnels à la réalisation aléatoire et non voulue par l'expérience de x .

Il va maintenant falloir passer à l'estimation du modèle et donc de la vraie valeur inconnue β_0 par l'estimateur $\hat{\beta}$. Plusieurs possibilités s'offrent alors à nous pour donner une valeur à $\hat{\beta}$:

1. la méthode des moindres carrés,
2. la méthode des moments,
3. la méthode du maximum de vraisemblance.

2.2.3 Moindres Carrés Ordinaires

Cette méthode est relativement intuitive car elle a pour principe de réduire les erreurs ε_i au minimum. On peut le voir aussi graphiquement : on souhaite alors avec les MCO tracer la droite (le modèle est linéaire) qui soit la plus centrale possible au sein du nuage de points (cf. Figure 2.1). Dans les cas, il est nécessaire de se munir d'une norme ou distance qui nous permettra de déterminer le critère à minimiser.

Les erreurs peuvent être soit positives soit négatives : il est donc nécessaire tout d'abord de rendre la somme des écarts ε_i positive. On peut penser alors à deux distances potentielles : la valeur absolue et le carré. Ce dernier est privilégié car il permet des calculs relativement simples. Il faut savoir néanmoins que le carré va pondérer plus fortement les erreurs élevées relativement à la valeur absolue ; cette distance est donc moins robuste statistiquement.

Revenons donc au critère des MCO : l'objectif plus rigoureux est donc de minimiser la somme de toutes les erreurs au carré :

$$\min \sum_{i=1}^n \varepsilon_i^2 = \min \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 = \min S(\alpha, \beta)$$

Propriété 2.2.1. L'estimateur du paramètre β issu des MCO est :

$$\hat{\beta}^{MCO} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} = \frac{\text{Cov}(X, Y)}{\text{Var } X}$$

$$\hat{\alpha}^{MCO} = \bar{Y} - \hat{\beta}^{MCO} \bar{X}$$

avec $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ et $\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Preuve. Minimisons $S(\alpha, \beta)$. Pour cela, il faut résoudre le système :

$$\begin{cases} \frac{\partial S(\hat{\alpha}, \hat{\beta})}{\partial \alpha} = 0 \\ \frac{\partial S(\hat{\alpha}, \hat{\beta})}{\partial \beta} = 0 \end{cases} \Leftrightarrow \begin{cases} -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \\ -2 \sum_{i=1}^n x_i (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \end{cases} \Leftrightarrow \begin{cases} \hat{\alpha} + \hat{\beta}\bar{X} = \bar{Y} \\ \hat{\beta} \sum_{i=1}^n x_i^2 + \hat{\alpha} \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \end{cases}$$

Ce qui donne :

$$\begin{cases} \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \\ \hat{\beta} \sum_{i=1}^n x_i^2 + \bar{Y} \sum_{i=1}^n x_i - \hat{\beta}\bar{X} \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \end{cases}$$

Ainsi l'estimateur de β par MCO est :

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^n x_i y_i - \bar{Y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{X} \sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n x_i (y_i - \bar{Y})}{\sum_{i=1}^n x_i (x_i - \bar{X})} = \frac{\sum_{i=1}^n x_i (y_i - \bar{Y}) - \bar{X} \sum_{i=1}^n (y_i - \bar{Y})}{\sum_{i=1}^n x_i (x_i - \bar{X}) - \bar{X} \sum_{i=1}^n (x_i - \bar{X})} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} = \frac{\hat{\text{Cov}}(X, Y)}{\hat{\text{Var}} X} \end{aligned}$$

car $\sum_{i=1}^n (x_i - \bar{X}) = \sum_{i=1}^n (y_i - \bar{Y}) = 0$.

$\hat{\text{Cov}}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$ et $\hat{\text{Var}} X = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$ sont les estimateurs empiriques sans biais des covariance $\text{Cov}(x, y) = \mathbb{E}[(x - \mathbb{E} x)(y - \mathbb{E} y)]$ et variance $\text{Var } x = \mathbb{E}(x - \mathbb{E} x)^2$ calculées à partir des données réelles, comme l'est $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ pour l'espérance $\mathbb{E} x$.

En effet, rappelons que :

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{X})^2 &= \sum_{i=1}^n (x_i - \mathbb{E} x + \mathbb{E} x - \bar{X})^2 \\
 &= \sum_{i=1}^n (x_i - \mathbb{E} x)^2 + \sum_{i=1}^n (\bar{X} - \mathbb{E} x)^2 - 2 \sum_{i=1}^n (x_i - \mathbb{E} x) (\bar{X} - \mathbb{E} x) \\
 &= \sum_{i=1}^n (x_i - \mathbb{E} x)^2 + n (\bar{X} - \mathbb{E} x)^2 - 2 (\bar{X} - \mathbb{E} x) \left(\sum_{i=1}^n x_i - n \mathbb{E} x \right) \\
 &= \sum_{i=1}^n (x_i - \mathbb{E} x)^2 - n (\bar{X} - \mathbb{E} x)^2
 \end{aligned}$$

Donc :

$$\begin{aligned}
 \mathbb{E} \left[\sum_{i=1}^n (x_i - \bar{X})^2 \right] &= \sum_{i=1}^n \mathbb{E} (x_i - \mathbb{E} x)^2 - n \mathbb{E} (\bar{X} - \mathbb{E} \bar{X})^2 \\
 &= n \text{Var } x - n \text{Var } \bar{X} = n \text{Var } x - n \text{Var} \left[\frac{1}{n} \sum_{i=1}^n x_i \right] \\
 &= (n-1) \text{Var } x
 \end{aligned}$$

car $\text{Cov}(x_i, x_j) = 0 \forall i \neq j$.

Ainsi :

$$\mathbb{E} \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \right] = \mathbb{E} [\hat{\text{Var}} X] = \text{Var } x$$

□

2.2.4 Méthode des Moments

Le principe de cette méthode est d'utiliser l'information que l'on possède sur des moments pour construire des conditions pouvant identifier les estimateurs $\hat{\beta}$ et $\hat{\alpha}$. Pour aboutir à une solution, il faut qu'il y ait donc au moins deux conditions des moments.

Propriété 2.2.2. *L'estimateur du paramètre β issu de la MM est :*

$$\begin{aligned}
 \hat{\beta}^{MM} &= \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} = \hat{\beta}^{MCO} \\
 \hat{\alpha}^{MM} &= \bar{Y} - \hat{\beta}^{MM} \bar{X} = \hat{\alpha}^{MCO}
 \end{aligned}$$

$$\text{avec } \bar{X} = \sum_{i=1}^n x_i \text{ et } \bar{Y} = \sum_{i=1}^n y_i.$$

Preuve. A partir de la première hypothèse $\mathbb{E}[\varepsilon|x] = 0$, nous pouvons spécifier les conditions des moments souhaitées :

- (i) $\mathbb{E}[\varepsilon|x] = 0 \Rightarrow \mathbb{E}[\mathbb{E}[\varepsilon|x]] = 0 \Rightarrow \mathbb{E}\varepsilon = 0 \Rightarrow \mathbb{E}[y - \alpha - \beta x] = 0$
- (ii) $\mathbb{E}[\varepsilon|x] = 0 \Rightarrow \mathbb{E}[x\mathbb{E}[\varepsilon|x]] = 0 \Rightarrow \mathbb{E}[\mathbb{E}[x\varepsilon|x]] = 0 \Rightarrow \mathbb{E}[x\varepsilon] = 0 \Rightarrow \mathbb{E}[x(y - \alpha - \beta x)] = 0$

Si nous remplaçons ces deux conditions des moments par leur contrepartie empirique, nous obtenons un système d'équations identifiantes :

$$(i) \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$$

$$(ii) \frac{1}{n} \sum_{i=1}^n x_i (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0$$

Nous retrouvons alors le système d'équations déterminant l'estimation des paramètres par MCO. □

L'espérance conditionnelle

Voici trois propriétés essentielles aux espérances conditionnelles :

$$(i) \mathbb{E} [\mathbb{E} [y|x]] = \mathbb{E} y$$

$$(ii) \mathbb{E} [g(x)y|x] = g(x)\mathbb{E} [y|x]$$

$$(iii) \mathbb{E} [y|x] = \mathbb{E} y \text{ si } x \perp y$$

2.2.5 Méthode du Maximum de Vraisemblance

La méthode du Maximum de Vraisemblance se fonde sur le principe de maximiser la probabilité que la modélisation particulière que nous imposons à l'échantillon des données soit *vraisemblable*. Il faut donc ajouter contrairement aux deux méthodes précédentes une nouvelle hypothèse de modélisation des données. La plus classique est l'hypothèse que les données sont issues d'une loi de probabilité gaussienne.

Hypothèse 2.2.4. $\varepsilon|x \sim \mathcal{N}(0, \sigma^2)$ c'est-à-dire que les erreurs sont gaussiennes conditionnellement à x .

Propriété 2.2.3. L'estimateur du paramètre β issu de la Méthode du MV est :

$$\hat{\beta}^{MV} = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^N (x_i - \bar{X})^2} = \hat{\beta}^{MCO}$$

$$\hat{\alpha}^{MV} = \bar{Y} - \hat{\beta}^{MV} \bar{X} = \hat{\alpha}^{MCO}$$

$$\text{avec } \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \text{ et } \bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Preuve. A partir de l'hypothèse de loi des résidus, nous pouvons définir la loi de la réalisation de la variable endogène $y_i|x_i$: $y_i|x_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma^2)$. La densité de $y_i|x_i$ s'écrit alors :

$$f(y_i|x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2 \right]$$

Comme les données $y_i|x_i$ sont indépendantes et identiquement distribuées (i.i.d.), la vraisemblance de

l'échantillon (y_1, \dots, y_n) est la suivante :

$$\begin{aligned}\mathcal{L}(y|x; \alpha, \beta, \sigma) &= \prod_{i=1}^n f(y_i|x_i) = \prod_{i=1}^n \left(\frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2 \right] \right) \\ &= \sigma^{-n} (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \right] \\ \ln \mathcal{L}(y|x; \alpha, \beta, \sigma) &= -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\end{aligned}$$

Maximiser la vraisemblance revient à maximiser la log-vraisemblance et d'après l'expression ci-dessus à minimiser l'expression suivante en α et β :

$$\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 = \sum_{i=1}^n \varepsilon_i^2$$

Nous retrouvons donc la fonction $S(\alpha, \beta)$ des MCO à minimiser. □

2.2.6 Propriétés des estimateurs

Nous allons étudier les espérances (et ainsi les biais), les variances et les lois des estimateurs $\hat{\alpha}$ et $\hat{\beta}$.

Propriété 2.2.4. *Les estimateurs $\hat{\alpha}$ et $\hat{\beta}$ sont sans biais.*

Preuve.

$$\begin{cases} y_i = \alpha + \beta x_i + \varepsilon_i \\ \bar{Y} = \alpha + \beta \bar{X} + \bar{\varepsilon} \end{cases}$$

Ainsi, nous avons $y_i - \bar{Y} = \beta (x_i - \bar{X}) + \varepsilon_i - \bar{\varepsilon}$. Si nous l'intégrons dans l'expression de $\hat{\beta}$, nous obtenons :

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{X}) (y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} = \frac{\sum_{i=1}^n (x_i - \bar{X}) (\beta (x_i - \bar{X}) + \varepsilon_i - \bar{\varepsilon})}{\sum_{i=1}^n (x_i - \bar{X})^2} \\ &= \beta + \frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{X}) - \bar{\varepsilon} \sum_{i=1}^n (x_i - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2} = \beta + \frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2}\end{aligned}\tag{2.1}$$

Ainsi, comme $\mathbb{E}[\varepsilon|x] = 0$ d'après l'hypothèse 2.2.1,

$$\mathbb{E}[\hat{\beta}|x] = \beta + \mathbb{E} \left[\frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2} \middle| x \right] = \beta + \frac{1}{\sum_{i=1}^n (x_i - \bar{X})^2} \sum_{i=1}^n (x_i - \bar{X}) \mathbb{E}[\varepsilon_i|x] = \beta$$

Pour $\hat{\alpha}$, nous avons :

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = \alpha + \beta \bar{X} + \bar{\varepsilon} - \hat{\beta} \bar{X} = \alpha + (\beta - \hat{\beta}) \bar{X} + \bar{\varepsilon}.\tag{2.2}$$

Ainsi $\mathbb{E}[\hat{\alpha}|x] = \alpha$ □

Propriété 2.2.5. Les variances de $\hat{\alpha}$ et $\hat{\beta}$ ont les expressions suivantes :

$$\text{Var} [\hat{\alpha}|x] = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{X})^2}$$

$$\text{Var} [\hat{\beta}|x] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{X})^2}$$

Preuve. D'après l'équation (2.1) et la définition de la variance d'une somme de variables aléatoires, nous avons :

$$\begin{aligned} \text{Var} [\hat{\beta}|x] &= \text{Var} \left[\beta + \frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2} \middle| x \right] \\ &= \frac{1}{\left[\sum_{i=1}^n (x_i - \bar{X})^2 \right]^2} \left\{ \sum_{i=1}^n (x_i - \bar{X})^2 \text{Var} [\varepsilon_i|x] + 2 \sum_{1 \leq i < j \leq n} (x_i - \bar{X})(x_j - \bar{X}) \text{Cov} (\varepsilon_i, \varepsilon_j) \right\} \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \end{aligned}$$

d'après les hypothèses 2.2.2 et 2.2.3.

Rappelons que d'après (2.2) $\hat{\alpha} = \alpha + (\beta - \hat{\beta}) \bar{X} + \bar{\varepsilon}$. On a donc :

$$\begin{aligned} \text{Var} [\hat{\alpha}|x] &= \bar{X}^2 \text{Var} [\hat{\beta}|x] + \text{Var} [\bar{\varepsilon}|x] + 2\bar{X} \text{Cov} (\beta - \hat{\beta}, \bar{\varepsilon}) \\ &= \frac{\sigma^2 \bar{X}^2}{\sum_{i=1}^n (x_i - \bar{X})^2} + \frac{\sigma^2}{n} - 2\bar{X} \text{Cov} (\hat{\beta}, \bar{\varepsilon}) \end{aligned}$$

Or :

$$\begin{aligned} \text{Cov} (\hat{\beta}, \bar{\varepsilon}) &= \text{Cov} \left(\frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2}, \frac{1}{n} \sum_{i=1}^n \varepsilon_i \right) \\ &= \frac{1}{n \sum_{i=1}^n (x_i - \bar{X})^2} \sum_{i=1}^n (x_i - \bar{X}) \sum_{j=1}^n \text{Cov} (\varepsilon_i, \varepsilon_j) \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \sum_{i=1}^n (x_i - \bar{X}) = 0 \end{aligned}$$

Ainsi,

$$\begin{aligned} \text{Var} [\hat{\alpha}|x] &= \frac{\sigma^2 \bar{X}^2}{\sum_{i=1}^n (x_i - \bar{X})^2} + \frac{\sigma^2}{n} = \frac{\sigma^2}{n} \left(1 + \frac{n\bar{X}^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \right) \\ &= \frac{\sigma^2 \sum_{i=1}^n x_i^2 + \sum_{i=1}^n \bar{X}^2 - 2\bar{X} \sum_{i=1}^n x_i + n\bar{X}^2}{n \sum_{i=1}^n (x_i - \bar{X})^2} = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{X})^2} \end{aligned}$$

□

Propriété 2.2.6. Les estimateurs $\hat{\alpha}$ et $\hat{\beta}$ ont pour loi, sous l'hypothèse 2.2.4 :

$$\begin{aligned} \hat{\alpha}|x &\sim \mathcal{N} \left(\alpha, \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{X})^2} \right) \\ \hat{\beta}|x &\sim \mathcal{N} \left(\beta, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \right) \end{aligned}$$

Preuve. L'expression (2.1) permet de vérifier que $\hat{\beta}$ est une combinaison linéaire des résidus ε_i qui sont gaussiens conditionnellement à x . Comme une combinaison de gaussiennes est gaussienne, $\hat{\beta}$ est gaussien avec les moyenne et variance définies précédemment. De même pour $\hat{\alpha}$. □

2.2.7 Estimation récursive versus estimation glissante

Un problème peut apparaître lorsque l'estimateur $\hat{\beta}$ se révèle instable suivant des sous-échantillons de l'échantillon total des données. En effet, il n'est pas rare que les sensibilités $\hat{\beta}$ évoluent dans le temps par exemple, si on travaille avec des données temporelles. Quelles méthodes d'estimation utiliser alors pour capter la variabilité des $\hat{\beta}$?

La question essentielle qui se pose est en fait celle de la partie de l'échantillon total des données à considérer pour déterminer les paramètres de la régression. Deux possibilités s'offrent à nous :

- Soit on fixe le point de départ de la construction du sous-échantillon (y_1, \dots, y_{n_r}) et on *empile* les données au fur et à mesure qu'elles apparaissent : le deuxième sous-échantillon devient alors (y_1, \dots, y_{n_r+1}) et ainsi de suite. On réalise des régressions récursives.
- Soit on fixe une taille d'échantillon n_g inférieure à n et on construit plusieurs sous-échantillons de taille n_g au sein de l'échantillon total. On fait ainsi *glisser* nos échantillons, de taille fixe n_g , au fur et à mesure du temps : (y_1, \dots, y_{n_g}) pour le premier sous-échantillon, (y_2, \dots, y_{n_g+1}) pour le deuxième, et ainsi de suite. On réalise alors des régressions glissantes.

Pour capter la réactivité des $\hat{\beta}$, nous privilégierons la régression glissante. Le choix de la taille n_g (comme pour celui de n_r) des sous-échantillons glissants n'a pas de règle figée. En fait, il dépend d'un arbitrage entre réactivité des estimateurs et bonnes propriétés statistiques. L'objectif étant de capter le plus précisément possible le moment où le régime des $\hat{\beta}$ a changé, plus l'échantillon sera de petite taille, moins il sera pollué par les valeurs du passé (si on travaille avec des données temporelles) mais plus il aura des propriétés statistiques mauvaises.

2.3 Validation du modèle

2.3.1 Coefficient de détermination

La qualité du modèle statistique peut être déterminée par le coefficient de détermination, qui nous fournit une indication sur la proximité entre les données observées (y_1, \dots, y_n) et les valeurs ajustées par le modèle $(\hat{y}_1, \dots, \hat{y}_n)$ avec $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \forall i = 1, \dots, n$.

Définition 2.3.1. *Le coefficient de détermination mesure le rapport entre la variance expliquée par le modèle c'est-à-dire la variance des valeurs ajustées et celle des observations totale :*

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2}$$

$0 \leq R^2 \leq 1$ et plus R^2 est proche de 1, plus la variance résiduelle est faible ; le modèle linéaire explique donc une grande partie de la dispersion des données.

Propriété 2.3.1. *Le coefficient de détermination est aussi égal au carré du coefficient de corrélation linéaire empirique de Bravais-Pearson entre x et y noté $\rho(x, y)$.*

Preuve.

$$\begin{aligned} R^2 &= \frac{\sum_{i=1}^n (\hat{\alpha} + \hat{\beta}x_i - \bar{Y})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} = \frac{\sum_{i=1}^n (\bar{Y} - \hat{\beta}\bar{X} + \hat{\beta}x_i - \bar{Y})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} \\ &= \hat{\beta}^2 \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} = \left(\frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^N (x_i - \bar{X})^2} \right)^2 \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} \\ &= \frac{\left[\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) \right]^2}{\sum_{i=1}^N (x_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{Y})^2} = \rho^2(x, y) \end{aligned}$$

□

2.3.2 Tests statistiques

Test de Student

La question importante que l'on peut se poser aussi pour valider le modèle statistique est si le paramètre à estimer est significativement nul ou pas. Sous l'hypothèse 2.2.4, nous pouvons procéder à ces tests sur les paramètres. Nous déterminons alors les tests statistiques suivants :

$$\begin{cases} H_0 : \alpha = 0 \\ H_a : \alpha \neq 0 \end{cases} \quad \begin{cases} H_0 : \beta = 0 \\ H_a : \beta \neq 0 \end{cases}$$

Pour construire les statistiques de test, nous avons besoin tout d'abord de déterminer un estimateur sans biais de la variance des résidus.

Propriété 2.3.2. *L'estimateur sans biais de la variance des résidus s'écrit :*

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$$

avec $\hat{\varepsilon}_i = y_i - \hat{y}_i$.

Preuve. Nous avons :

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i^2 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left\{ (\alpha - \hat{\alpha}) + (\beta - \hat{\beta}) x_i + \varepsilon_i \right\}^2 \\ &= \sum_{i=1}^n \left\{ (\beta - \hat{\beta}) (x_i - \bar{X}) + \varepsilon_i - \bar{\varepsilon} \right\}^2 \quad \text{d'après (2.2)} \\ &= (\beta - \hat{\beta})^2 \sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 + 2(\beta - \hat{\beta}) \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon}) (x_i - \bar{X}) \\ &= (\beta - \hat{\beta})^2 \sum_{i=1}^n (x_i - \bar{X})^2 + \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 - 2(\beta - \hat{\beta}) \sum_{i=1}^n (x_i - \bar{X})^2 \\ \text{d'après (2.1) : } &\sum_{i=1}^n \varepsilon_i (x_i - \bar{X}) = (\hat{\beta} - \beta) \sum_{i=1}^n (x_i - \bar{X})^2 \\ &= \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 - (\beta - \hat{\beta})^2 \sum_{i=1}^n (x_i - \bar{X})^2 \end{aligned}$$

Ainsi,

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^n \hat{\varepsilon}_i^2 | x \right] &= \mathbb{E} \left[\left(\sum_{i=1}^n \varepsilon_i^2 - n\bar{\varepsilon}^2 \right) | x \right] - \sum_{i=1}^n (x_i - \bar{X})^2 \text{Var} [\hat{\beta} | x] \\ &= n\sigma^2 - n(\text{Var} [\bar{\varepsilon} | x] + \mathbb{E}^2 [\bar{\varepsilon} | x]) - \sigma^2 = (n-2)\sigma^2 \end{aligned}$$

Donc il s'en suit que $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$ est bien un estimateur sans biais de σ^2 .

□

Théorème 2.3.1 (Cochran). *La somme des carrés de n variables aléatoires indépendantes de loi Normale centrée réduite suit un loi du Khi-deux à n degrés de libertés. Donc :*

$$\sum_{i=1}^n \left(\frac{\hat{\varepsilon}_i}{\sigma} \right)^2 | x = (n-2) \frac{\hat{\sigma}^2}{\sigma^2} | x \sim \chi_{(n-2)}^2$$

car $\varepsilon_i | x \sim \mathcal{N}(0, \sigma^2)$ d'après l'hypothèse 2.2.4.

Nous pouvons ainsi maintenant déterminer pour β la statistique de test suivante :

$$t = \frac{|\hat{\beta} - \beta|}{\hat{\sigma}_{\hat{\beta}}} = \frac{|\hat{\beta} - \beta| \sigma_{\hat{\beta}}}{\sigma_{\hat{\beta}} \hat{\sigma}_{\hat{\beta}}} = \frac{|\hat{\beta} - \beta| \sigma}{\sigma_{\hat{\beta}} \hat{\sigma}}$$

On sait d'après la propriété 2.2.6 et le théorème 2.3.1 que :

$$\begin{aligned} \frac{\hat{\beta} - \beta}{\sigma_{\hat{\beta}}} | x &\sim \mathcal{N}(0, 1) \\ (n-2) \frac{\hat{\sigma}^2}{\sigma^2} | x &\sim \chi_{(n-2)}^2 \end{aligned}$$

Or d'après la définition d'une loi de Student, nous pouvons écrire que :

$$t|x = \frac{\frac{|\hat{\beta} - \beta|}{\sigma_{\hat{\beta}}}}{\sqrt{\frac{(n-2)\hat{\sigma}^2}{\sigma^2}}} |x \sim St_{n-2}$$

Sous H_0 la statistique de test devient :

$$t = \frac{|\hat{\beta}|}{\hat{\sigma}_{\hat{\beta}}}$$

Concernant le test sur α , cette statistique devient :

$$t = \frac{|\hat{\alpha}|}{\hat{\sigma}_{\hat{\alpha}}}$$

Rappelons, d'après la propriété 2.2.5, que :

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{X})^2}} \quad \hat{\sigma}_{\hat{\alpha}} = \sqrt{\frac{\hat{\sigma}^2 \frac{\sum_{i=1}^n x_i^2}{n}}{\sum_{i=1}^n (x_i - \bar{X})^2}}$$

En se fixant α comme niveau de test, nous accepterons l'hypothèse nulle si la statistique de test t est inférieure au quantile de la loi de Student de probabilité $1 - \alpha/2$.

Test de Durbin-Watson

Ce test permet de vérifier si la covariance entre 2 résidus est bien nulle, avec la contrainte que ces résidus se suivent. Il teste donc la présence d'auto-corrélation à l'ordre 1 entre les résidus ε_i . On considère alors le processus suivant :

$$\hat{\varepsilon}_i = \rho \hat{\varepsilon}_{i-1} + u_i$$

où u_i est un bruit blanc.

Le test est alors :

$$\begin{cases} H_0 : \rho = 0 \\ H_a : \rho \neq 0 \end{cases}$$

La statistique de test est :

$$DW = \frac{\sum_{i=2}^n (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^n \hat{\varepsilon}_i^2}$$

Cette statistique varie entre 0 et 4 et vaut 2 en l'absence d'auto-corrélation à l'ordre 1 entre les résidus.

Test de Jarque et Bera

Ce test permet de valider l'hypothèse de normalité des résidus estimés. Il faut pour cela déterminer les skewness β_1 et kurtosis β_2 , correspondant respectivement aux coefficients d'asymétrie et d'aplatissement développés par Pearson¹ :

$$\beta_1 = \frac{\left[\mathbb{E} (\varepsilon - \mathbb{E} \varepsilon)^3 \right]^2}{\left[\mathbb{E} (\varepsilon - \mathbb{E} \varepsilon)^2 \right]^3} = \frac{\mu_3^2}{\mu_2^3} = \frac{\mu_3^2}{(\sigma^2)^3} \quad \beta_2 = \frac{\mathbb{E} (\varepsilon - \mathbb{E} \varepsilon)^4}{\left[\mathbb{E} (\varepsilon - \mathbb{E} \varepsilon)^2 \right]^2} = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$$

Fisher a développé deux autres coefficients d'asymétrie et d'aplatissement :

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \quad \gamma_2 = \beta_2 - 3$$

Souvent on utilise γ_1 pour définir le skewness car il peut être positif ou négatif contrairement à β_1 qui est toujours positif. Enfin, γ_2 est aussi appelé le kurtosis en excès ou encore le kurtosis normalisé.

Pour une loi Normale, nous avons :

$$\beta_1 = \gamma_1 = 0 \quad \beta_2 = 3 \Rightarrow \gamma_2 = 0$$

Estimons ces coefficients d'asymétrie et d'aplatissement. Précisons que nous devons utiliser les estimateurs des résidus de la régression linéaire simple $\hat{\varepsilon} = y - \hat{y}$. Or nous avons $\tilde{\varepsilon} = \bar{Y} - \tilde{Y} = \bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X} = 0$ d'après l'expression de $\hat{\alpha}^{\text{MCO}}$ (cf. propriété 2.2.1). Ainsi, nous pouvons estimer les skewness et kurtosis à l'aide des contreparties empiriques des coefficients d'asymétrie et d'aplatissement de Pearson suivantes :

$$\hat{\beta}_1 = \frac{\left(\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^3 \right)^2}{\left(\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \right)^3} \quad \hat{\beta}_2 = \frac{\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^4}{\left(\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \right)^2} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}'} \right)^4$$

$$\text{avec } \hat{\sigma}' = \sqrt{\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2}$$

Nous avons aussi les estimateurs des coefficients d'asymétrie et d'aplatissement de Fisher suivants :

$$\hat{\gamma}_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}'} \right)^3 \quad \hat{\gamma}_2 = \hat{\beta}_2 - 3$$

La statistique de Jarque et Bera (cf. [1]) s'écrit alors :

$$JB = \frac{n}{6} \left(\hat{\gamma}_1^2 + \frac{1}{4} \hat{\gamma}_2^2 \right)$$

Sous l'hypothèse nulle de normalité des résidus, la statistique JB suit une loi du Khi-deux à 2 degrés de liberté. Nous rejetons la normalité si JB est supérieure au quantile de probabilité $1 - \alpha$ de la loi $\chi_{(2)}^2$.

D'autres tests statistiques plus complexes et précis existent mais vont au-delà du cadre de ce cours.

Ajoutons aussi qu'il est fortement conseillé en premier lieu de grapher les résidus pour observer si les hypothèses 2.2.1, 2.2.2, 2.2.3 et 2.2.4 (via un QQ-plot par exemple pour cette dernière hypothèse) tiennent. C'est l'étape préalable indispensable avant d'effectuer le moindre test statistique.

1. Précisons que les espérances sont bien conditionnelles à x mais pour ne pas surcharger les expressions des skewness et kurtosis, nous ne l'avons pas spécifié.

2.3.3 Backtest

Le principe du backtest est de distinguer sur l'échantillon total des données un sous-échantillon d'apprentissage de taille n_a sur lequel on va estimer les paramètres du modèle et un sous-échantillon de backtest de taille n_b sur lequel on va tester la robustesse du modèle. Deux indicateurs peuvent alors être déterminés :

- Le Root Mean Squared Error (RMSE) égal à $\sqrt{\frac{1}{n_b} \sum_{i=1}^{n_b} (\hat{y}_i - y_i)^2}$ avec \hat{y}_i la valeur prévue par le modèle défini sur l'échantillon d'apprentissage à partir des observations x_i de l'échantillon de backtest ; ces n_b prévisions sont alors comparées aux valeurs y_i de l'échantillon de backtest réellement réalisées.
- Le taux de bon classement égal à $\frac{\text{card}[\hat{y}_i * y_i > 0]}{n_b}$ permettant de juger le bon sens de la prévision \hat{y}_i .
Il est possible d'ajouter des seuils pour ne prendre en compte que les prévisions importantes en valeur absolue.

CHAPITRE 3

RÉGRESSION LOGISTIQUE

Pour plus d'informations sur la méthodologie de la régression logistique nous renvoyons à Thomas [14].

3.1 Inadéquation des MCO

Nous nous plaçons dans le contexte du credit scoring où l'un des objectifs est d'estimer la probabilité de défaillance d'une entreprise, d'un État ou d'un particulier. Nous supposons que le défaut est modélisé par une variable y_i binaire, avec $i = 1, \dots, n$, qui peut prendre comme valeur 0 (pas de défaut) ou 1 (défaut). Tentons d'appliquer en premier lieu le modèle de la régression linéaire simple étudié dans le chapitre précédent. Nous supposons donc que le défaut y_i de l'entreprise i est expliqué par une seule variable x_i (un ratio d'endettement par exemple).

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Cette modélisation présente de nombreux inconvénients et la technique des Moindres Carrés Ordinaires, classiquement utilisée dans les modèles de régression à variable dépendante y continue, ne peut être appliquée. En effet, on peut énumérer trois limites à l'application des MCO.

Tout d'abord, rien ne contraint le modèle à ce que la probabilité estimée soit comprise entre 0 et 1.

Ensuite, les résidus de la régression ne peuvent prendre que 2 valeurs : $1 - \alpha - \beta x_i$ lorsque $\{y_i = 1\}$ avec la probabilité $\alpha + \beta x_i$ car $\mathbb{P}(y_i = 1|x_i) = \mathbb{P}(y_i = 1|x_i) * 1 + \mathbb{P}(y_i = 0|x_i) * 0 = \mathbb{E}[y_i|x_i] = \alpha + \beta x_i$ d'après l'hypothèse 2.2.1. La seconde valeur que peut prendre le résidu est $-\alpha - \beta x_i$ avec la probabilité complémentaire $1 - \alpha - \beta x_i$. Le cadre d'analyse des MCO considérant les résidus continus se prête donc mal à la modélisation de y_i .

Enfin, le calcul de la variance des résidus donne :

$$\begin{aligned} \text{Var} [\varepsilon_i|x_i] &= \mathbb{E} [\varepsilon_i^2|x_i] \\ &= \{\varepsilon_i^2|y_i = 1\} * \mathbb{P}(y_i = 1|x_i) + \{\varepsilon_i^2|y_i = 0\} * \mathbb{P}(y_i = 0|x_i) \\ &= (1 - \alpha - \beta x_i)^2(\alpha + \beta x_i) + (-\alpha - \beta x_i)^2(1 - \alpha - \beta x_i) \\ &= (\alpha + \beta x_i)(1 - \alpha - \beta x_i) \end{aligned}$$

Le modèle est donc hétéroscédastique par construction, la variance des résidus dépendant des x_i et n'étant ainsi pas constante. L'hypothèse 2.2.2 n'est pas vérifiée.

3.2 Introduction d'une variable latente

Il est donc nécessaire d'utiliser une technique alternative et pour cela nous allons introduire une variable latente y_i^* (inobservable) telle que :

$$\begin{cases} y_i = 1 & \text{si } y_i^* > c \\ y_i = 0 & \text{si } y_i^* \leq c \end{cases}$$

c correspond à la valeur seuil que nous pouvons normaliser à 0. On va pouvoir à l'aide de cette variable latente continue appliquer le modèle de régression linéaire simple : $y_i^* = \alpha + \beta x_i + \varepsilon_i$.

Nous considérons alors la réalisation de la variable dépendante binaire y_i comme provenant d'une règle de décision sur la variable latente y_i^* s'écrivant :

$$\begin{cases} \mathbb{P}(y_i = 1|x_i) = \mathbb{P}(y_i^* > 0|x_i) = \mathbb{P}(\varepsilon_i > -\alpha - \beta x_i) = 1 - F(-\alpha - \beta x_i) = F(\alpha + \beta x_i) \\ \mathbb{P}(y_i = 0|x_i) = \mathbb{P}(y_i^* \leq 0|x_i) = \mathbb{P}(\varepsilon_i \leq -\alpha - \beta x_i) = F(-\alpha - \beta x_i) = 1 - F(\alpha + \beta x_i) \end{cases}$$

avec F une fonction de répartition symétrique autour de 0 définie conditionnellement à x .

Il faut bien remarquer que dans la modélisation de la variable binaire y_i , nous ne modélisons plus la variable en elle-même, mais les probabilités de réalisation des événements $\{y_i = 1\}$ et $\{y_i = 0\}$.

3.3 Estimation par Maximum de Vraisemblance

Pour estimer les paramètres $\hat{\alpha}$ et $\hat{\beta}$ par maximum de vraisemblance, il nous faut spécifier la fonction de répartition F des résidus conditionnellement à x . Dans la littérature, nous avons le choix entre la loi logistique (modèle Logit), la loi normale (modèle Probit) et la loi de Gompertz, non symétrique (modèle Gompit). Nous ne verrons ici que le cas de la loi logistique, plus facile à manipuler, qui a pour fonction de répartition Λ :

$$\Lambda(z_i) = \frac{\exp(z_i)}{1 + \exp(z_i)}$$

Nous estimons le modèle par maximum de vraisemblance. En supposant que les données y_i sont i.i.d., la vraisemblance s'écrit :

$$\begin{aligned} \mathcal{L}(y|x; \alpha, \beta) &= \prod_{i=1}^n [\Lambda(\alpha + \beta x_i)]^{y_i} [1 - \Lambda(\alpha + \beta x_i)]^{1-y_i} \\ &= \prod_{i=1}^n \left[\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right]^{y_i} \left[\frac{1}{1 + \exp(\alpha + \beta x_i)} \right]^{1-y_i} \end{aligned}$$

Et la log-vraisemblance a pour expression :

$$\begin{aligned} \ln \mathcal{L}(y|x; \alpha, \beta) &= \sum_{i=1}^n \{y_i (\alpha + \beta x_i) - y_i \ln [1 + \exp(\alpha + \beta x_i)] - (1 - y_i) \ln [1 + \exp(\alpha + \beta x_i)]\} \\ &= \sum_{i=1}^n \{y_i (\alpha + \beta x_i) - \ln [1 + \exp(\alpha + \beta x_i)]\} \end{aligned}$$

Les conditions de premier ordre sont :

$$\begin{cases} \frac{\partial \ln \mathcal{L}(y|x; \hat{\alpha}, \hat{\beta})}{\partial \alpha} = 0 \\ \frac{\partial \ln \mathcal{L}(y|x; \hat{\alpha}, \hat{\beta})}{\partial \beta} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n \left[y_i - \frac{\exp(\hat{\alpha} + \hat{\beta} x_i)}{1 + \exp(\hat{\alpha} + \hat{\beta} x_i)} \right] = 0 \\ \sum_{i=1}^n \left[x_i y_i - \frac{x_i \exp(\hat{\alpha} + \hat{\beta} x_i)}{1 + \exp(\hat{\alpha} + \hat{\beta} x_i)} \right] = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n [y_i - \Lambda(\hat{\alpha} + \hat{\beta} x_i)] = 0 \\ \sum_{i=1}^n x_i [y_i - \Lambda(\hat{\alpha} + \hat{\beta} x_i)] = 0 \end{cases}$$

Ces conditions de premier ordre peuvent être assimilées à celles des MCO d'un modèle de régression non-linéaire qui s'écrirait $y_i = \Lambda(\alpha + \beta x_i) + v_i$ avec v_i les résidus. Dans ce cadre, nous ne pouvons pas exprimer les estimateurs comme des fonctions simples des observations (x_i, y_i) . Il faut donc utiliser des méthodes numériques d'optimisation pour maximiser la log-vraisemblance.

Propriété 3.3.1. *Nous obtenons alors l'estimation de la probabilité de défaut sachant x_i notée \hat{p}_i pour l'entreprise i , correspondant à l'événement $\{y_i = 1\}$, suivante :*

$$\hat{p}_i = \Lambda(\hat{\alpha} + \hat{\beta}x_i) = \frac{\exp(\hat{\alpha} + \hat{\beta}x_i)}{1 + \exp(\hat{\alpha} + \hat{\beta}x_i)}$$

Précisons que les valeurs des paramètres estimés $\hat{\alpha}$ et $\hat{\beta}$ n'ont pas d'intérêt en soi ; seul leur signe est une information utilisable : en effet, il indique si la variable associée influence la probabilité à la hausse ou à la baisse.

CONCLUSION

Ce cours a permis je l'espère d'appréhender des méthodes statistiques classiques (régression linéaire, régression logistique) dans un cadre financier. Ces méthodes, quoique très connues et simples d'utilisation, doivent néanmoins être comprises et surtout mises en oeuvre rigoureusement suivant leurs hypothèses.

Bien entendu, de nombreux développements existent que nous n'avons pas pu étudier dans ce cours : régression multiple (plus d'une variable explicative), validation du modèle avec sélection des variables les plus explicatives, colinéarité, critères d'information (Akaike, Schwarz), distance de Cook, tests d'homoscédasticité, cointégration, entre autres. Mais aussi analyse discriminante pour le credit scoring ainsi que les réseaux de neurones, les modèles de durées de vie, les arbres de décision et les machines à support vectoriel par exemple.

BIBLIOGRAPHIE

- [1] BERA, A.K. & JARQUE, C.M., 1987, *A Test for Normality of Observations and Regression Residuals*, International Statistical Review, 55, 163-172.
- [2] BOX, G.E.P. & COX, D.R., 1964, *An Analysis of Transformations*, Journal of the Royal Statistical Society, 26, 211-243.
- [3] CARTER HILL, R., GRIFFITHS, W.E & LIM, G.C., 2008, *Principles of Econometrics*, Wiley.
- [4] CLAUSS, P., 2011, *Gestion de Portefeuille*, Dunod.
- [5] DE SERVIGNY, A., METAYER, B. & ZELENKO, I., 2006, *Le risque de Crédit*, Dunod.
- [6] FAMA, E. & FRENCH, K., 1993, *Common Risk Factors in the Returns on Stocks and Bonds*, Journal of Financial Economics, 33, 3-56.
- [7] GRANGER, C.W.G. & NEWBOLD, P., 1974, *Spurious Regressions in Econometrics*, Journal of Econometrics, 26, 1045-1066.
- [8] LARDIC, S. & MIGNON, V., 2002, *Économétrie des Séries Temporelles Macroéconomiques et Financières*, Economica.
- [9] LINTNER, J., 1965, *The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets*, Review of Economics and Statistics, 47, 13-37.
- [10] MERTON, R.C., 1974, *On the Pricing of Corporate Debt : the Risk Structure of Interest Rates*, Journal of Finance, 29, 449-470.
- [11] MOSSIN, J., 1966, *Equilibrium in a Capital Asset Market*, Econometrica, 34, 768-783.
- [12] SHARPE, W., 1964, *Capital Asset Prices : a Theory of Market Equilibrium under Conditions of Risk*, Journal of Finance, 19, 425-442.
- [13] TASSI, Ph., 1989, *Méthodes Statistiques*, Economica.
- [14] THOMAS, A., 2000, *Économétrie des variables qualitatives*, Dunod.