

MODELES LINEAIRES ET ADDITIFS GENERALISES, MAXIMUM DE VRAISEMBLANCE LOCAL ET METHODES RELATIONNELLES EN ASSURANCE SUR LA VIE

Antoine DELWARDE
*Institut des Sciences Actuarielles UCL**

Daria KCHAKHIDZE

Scor Vie •

Laure OLIE

Scor Vie •

Michel DENUIT

Institut de Statistique & Institut des Sciences Actuarielles UCL #*

RESUME

Cet article est consacré aux modèles relationnels de prévision de la mortalité. Différentes alternatives sont envisagées: le modèle logit de Brass, le modèle à risques proportionnels de Cox, le modèle de Ewbank, Gomez De Leon et Stoto, et celui de Hannerz. Ces approches paramétriques sont unifiées dans le cadre des modèles linéaires et additifs généralisés, qui fournissent dès lors à l'actuaire un cadre de travail adéquat pour aborder les méthodes relationnelles. La mise en pratique de ces techniques est illustrée sur base de différents jeux de données. Les modèles proviennent de DELWARDE & DENUIT (2005) et sont adaptés aux méthodes relationnelles.

1. INTRODUCTION

1.1. Modèles semi-paramétriques en mortalité

Une partie importante de la littérature actuarielle est consacrée à la construction de tables de mortalité et, partant, à l'estimation des probabilités annuelles de décès q_x aux âges $x \in \{\alpha, \dots, \omega\}$, aussi appelées quotients de mortalité en démographie. Partant d'une estimation brute des q_x , souvent obtenue en rapportant un nombre de décès à un nombre d'individus soumis au risque, notée $\hat{Q} = \{\hat{q}_x, x = \alpha, \dots, \omega\}$, l'actuaire désire la plupart du temps ajuster ces quantités pour mieux faire ressortir les caractéristiques de la mortalité du groupe d'individus considéré, qu'il pressent assez régulière.

* Louvain-la-Neuve, Belgique

• Département Technique et Développement – Centre R&D sur l'Assurance de la Longévité et la Mortalité - CERDALM - La Défense, Paris

Institut de Statistique UCL, Louvain-la-Neuve, Belgique

Depuis De Moivre en 1725, de nombreux modèles paramétriques ont été proposés pour rendre compte de l'évolution de la mortalité en fonction de l'âge. Le principe est relativement simple: les quotients (ou les taux) de mortalité sont exprimés sous la forme d'une fonction de l'âge, fonction dépendant d'un ou de plusieurs paramètres (qu'il s'agira d'estimer sur base de statistiques de mortalité). Plus précisément, les méthodes d'ajustement analytique consistent à exprimer q_x sous la forme d'une fonction $q_x = f(x; \theta)$ dont la forme analytique est connue mais faisant intervenir p paramètres inconnus $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. Les valeurs de ces paramètres sont estimées à partir des \hat{q}_x (ou d'autres statistiques de mortalité appropriées). Notant $\hat{\theta}$ l'estimateur de θ , l'ajustement paramétrique de \hat{Q} consistera à lui substituer $\hat{Q}^{\text{ajustée}} = \{\hat{q}_x^{\text{ajustée}} = f(x; \hat{\theta}), x = \alpha, \dots, \omega\}$. De nombreuses fonctions ont été proposées dans la littérature. Citons pour mémoire les formules de Moivre, de Sang, de Gompertz, de Makeham, de Perks et de Heligman-Pollard. Pour plus de détails, le lecteur intéressé est invité à consulter SEAL (1979).

Aussi raisonnable qu'il puisse paraître, un modèle paramétrique n'est jamais qu'un résumé parfois trop sommaire de la réalité du phénomène observé. De plus, rien ne garantit que le modèle paramétrique soit bien spécifié. C'est pourquoi de nombreuses méthodes de lissage ont été proposées pour les quotients de mortalité. Ces méthodes consistent à substituer à la série brute \hat{Q} une série plus régulière, notée $\hat{Q}^{\text{lissée}} = \{\hat{q}_x^{\text{lissée}}, x = \alpha, \dots, \omega\}$. La mortalité n'est donc pas résumée dans un petit nombre de paramètre, mais est décrite par les $\omega - \alpha + 1$ probabilités annuelles de décès dont on a gommé les irrégularités par lissage. Etant donnée la puissance de l'outil informatique désormais à la disposition de l'actuaire, la complexité numérique inhérente aux modèles non-paramétriques n'est plus un obstacle à leur mise en pratique.

Depuis VERRALL (1996), les actuaires anglo-saxons préfèrent les méthodes semi-paramétriques d'ajustement des tables de mortalité. Cette approche allie souplesse et précision d'ajustement. Cet article a pour but de démontrer l'intérêt des méthodes semi-paramétriques dans le cadre des modèles relationnels.

1.2. Méthodes relationnelles en mortalité

De tout temps, les actuaires et les démographes ont utilisé des données de mortalité connexes pour améliorer l'estimation de la mortalité d'une population d'intérêt. L'approche relationnelle permet de relier les indicateurs démographiques relatifs à deux groupes d'individus. Ces indicateurs peuvent être les taux de mortalité, les probabilités de survie ou de décès, les espérances de vie, etc. Les deux groupes peuvent être

- deux populations distinctes
- une population et un sous-ensemble de celle-ci
- une même population à deux époques différentes.

Dans ce travail, nous nous inscrivons dans la lignée de VERRALL (1996) et nous montrons que les modèles linéaires et additifs généralisés fournissent des outils particulièrement bien adaptés à la construction de tables de mortalité. Les méthodes de construction des tables de mortalité qui en résultent sont tout à la fois intellectuellement satisfaisantes et empiriquement performantes. Ces modèles permettent également d'unifier les différentes approches suivies dans le cadre des méthodes relationnelles. Il suffit en effet d'introduire parmi les variables explicatives les caractéristiques de la mortalité de la population de référence (taux ou quotients de mortalité).

2. MODELES RELATIONNELS : REVUE DE LA LITTERATURE

2.1. Modèle à risques proportionnels de Cox

Les modèles paramétriques autorisent des comparaisons fines de la mortalité, mais sont évidemment sujets à mauvaise spécification : la forme paramétrique peut s'avérer erronée, jetant le discrédit sur les conclusions obtenues.

Le modèle à risques proportionnels proposé par COX (1972) permet de s'affranchir d'une formulation paramétrique rigide. Il postule que les taux de mortalité du groupe auquel l'actuaire s'intéresse sont proportionnels à ceux d'une population de référence, le facteur de proportionnalité ne dépendant pas de l'âge. Plus précisément, les taux μ_x sont liés à un jeu de taux de référence μ_x^{ref} par la relation

$$\mu_x = \theta \mu_x^{\text{ref}}, x = \alpha, \alpha + 1, \dots, \omega. \quad (1)$$

Il s'agit là du modèle relationnel le plus simple, liant la mortalité du groupe d'intérêt, représentée par les taux $\mu_\alpha, \mu_{\alpha+1}, \dots, \mu_\omega$ à celle de la mortalité de référence, représentée par les taux $\mu_\alpha^{\text{ref}}, \mu_{\alpha+1}^{\text{ref}}, \dots, \mu_\omega^{\text{ref}}$, à l'aide du modèle (1).

2.2. Modèle de Brass

La validité de (1) a cependant été mise en doute par de nombreux auteurs, dont BRASS (1971). En lieu et place de (1), BRASS (1971) suggère la relation

$$\text{logit}({}_x q_0) = \theta_1 + \theta_2 \text{logit}({}_x q_0^{\text{ref}}), x = \alpha, \alpha + 1, \dots, \omega, \quad (2)$$

où $\text{logit}(q) = \ln(q/(1-q))$. Il s'agit d'une relation linéaire entre les logit des quotients de mortalité d'intérêt et ceux d'une population de référence. Le recours à la transformation

logit permet d'obtenir une quantité non-contrainte (alors que $q \in (0,1)$, $\text{logit}(q) \in \mathbb{R}$). Voyez également BRASS (1969, 1974).

ZABA (1979) proposa une variante de (2). Un peu plus tard, EWBANK, GOMEZ-DE LEON & STOTO (1983) proposèrent à leur tour une forme paramétrique plus complexe liant $\text{logit}({}_x q_0)$ à $\text{logit}({}_x q_0^{\text{ref}})$.

Sur base de plusieurs études empiriques, le paramètre θ_2 dans le modèle (2) s'est avéré proche de 1. Si nous posons $\theta_2 = 1$ dans (2), il vient

$$\mu_x = \frac{\theta_1 \mu_x^{\text{ref}}}{1 - (1 - \theta_1)_x q_0^{\text{ref}}} \quad (3)$$

où $\theta_1 = \exp(\theta_1)$ et μ_x^{ref} est un taux instantané de mortalité de référence. Nous voyons dans ce cas que $\mu_x \rightarrow \mu_x^{\text{ref}}$ si $x \rightarrow +\infty$.

2.3. Modèle de Hannerz

HANNERZ (2001) a procédé à une analyse empirique de la validité des relations (1), (2) et (3). Parmi les jeux de données considérés par cet auteur figuraient des statistiques fournies par la fédération suédoise des sociétés d'assurance. Après avoir constaté l'inadéquation des formes paramétriques simples présentées plus haut, HANNERZ (2001) a proposé un modèle alternatif basé sur ses travaux antérieurs. Plus précisément, l'évolution de la mortalité en fonction de l'âge pour une population féminine donnée est décrite par la relation

$$\text{logit}({}_x q_0) = a_0 - a_1 x^{-1} + a_2 \frac{x^2}{2} + a_3 \frac{\exp(cx)}{c}. \quad (4)$$

Les paramètres a_0, a_1, a_2 et a_3 possèdent chacun leur signification (représentant la mortalité sur différentes tranches d'âges). Ceci a amené cet auteur au modèle relationnel

$$\text{logit}({}_x q_0) = \text{logit}({}_x q_0^{\text{ref}}) + \theta_0 - \theta_1 x^{-1} + \theta_2 \frac{x^2}{2} + \theta_3 \frac{\exp(cx)}{c} \quad (5)$$

où chaque paramètre indique la différence entre les mortalités respectives des deux groupes sur des tranches d'âges données. Ainsi,

- si $\theta_1 \neq 0$ alors les mortalités diffèrent aux âges les plus jeunes;
- si $\theta_2 \neq 0$ alors la relation entre la mortalité à l'âge adulte et durant l'enfance diffère entre les deux populations considérées;
- si $\theta_3 \neq 0$ alors les deux populations diffèrent aux âges élevés.

La relation (4) ne convient cependant pas aux hommes, en raison de la "bosse accidents", marquée chez les jeunes garçons (il s'agit des morts violentes, accidents de la route et suicides, entre 18 et 35 ans). HANNERZ (2001) propose alors un mélange de deux

lois de probabilité, l'une rendant compte des décès accidentels, et l'autre des décès naturels, chacune d'entre elles étant de la forme (4).

3. APPROCHE UNIFIEE DANS LE CADRE DES MODELES LINEAIRES ET ADDITIFS GENERALISES

Les modèles paramétriques se lisent aisément: chaque paramètre possède sa propre signification, ce qui permet une interprétation facile des résultats, et partant une explication des phénomènes sous-jacents. Dans la mesure où l'actuaire cherche moins à expliquer les causes des différences de mortalité qu'il constate qu'à en tenir compte dans le tarif, on peut imaginer avoir recours à des modèles semi-paramétriques, plus souples et autorisant des comparaisons plus fines de la mortalité.

3.1. Unification des modèles de Cox, de Brass et de Hannerz

Le point de départ de notre démarche consiste à fondre les modèles paramétriques (1)-(2)-(5) dans une classe plus large de modèles. Tout d'abord, notons que le modèle (1) est un cas particulier de

$$\ln \mu_x = f(\ln \mu_x^{\text{ref}}), x = \alpha, \alpha + 1, \dots, \omega, \quad (6)$$

en prenant $f(\xi) = \ln \theta + \xi$. Le modèle (2) s'inscrit quant à lui naturellement dans la formulation générale

$$\text{logit}({}_x q_0) = f(\text{logit}({}_x q_0^{\text{ref}})), x = \alpha, \alpha + 1, \dots, \omega, \quad (7)$$

avec $f(\xi) = \alpha + \beta \xi$. Le modèle (5) nécessite l'introduction de deux fonctions: la première de $\text{logit}({}_x q_0^{\text{ref}})$ et la seconde de l'âge x . Ainsi, on peut voir (5) comme un cas particulier de la formulation plus générale

$$\text{logit}({}_x q_0) = f_1(\text{logit}({}_x q_0^{\text{ref}})) + f_2(x), x = \alpha, \alpha + 1, \dots, \omega, \quad (8)$$

où $f_1(\xi) = \xi$ et

$$f_2(\xi) = \theta_0 - \theta_1 \xi^{-1} + \theta_2 \frac{\xi^2}{2} + \theta_3 \frac{\exp(c\xi)}{c}.$$

La structure qui se dégage de ces exemples apparaît donc clairement: les indicateurs démographiques d'intérêt sont d'abord transformés de manière à ne plus être contraints (par une transformation logarithmique pour les taux et une transformation logit pour les quotients), et ensuite ils sont expliqués à l'aide d'une fonction de leurs homologues d'une population de référence, le cas échéant en y superposant une fonction de l'âge. Nous allons exploiter cette structure pour proposer une généralisation semi-paramétrique des modèles relationnels.

L'approche préconisée dans cet article peut se résumer comme suit. Plutôt que de chercher une expression analytique pour les fonctions intervenant dans (6)-(7)-(8) puis d'en estimer ces paramètres, nous estimerons directement ces fonctions sur base des statistiques de mortalité, évitant ainsi tout risque de mauvaise spécification. Cette approche convient parfaitement à l'actuaire, qui veut tenir compte des particularités de la mortalité du groupe qu'il couvre lorsqu'il établit une tarification, sans avoir nécessairement besoin d'expliquer les différences de mortalité qu'il constate. L'estimation directe des fonctions intervenant dans (6)-(7)-(8) sans en spécifier la forme paramétrique répond à cet objectif (alors que les spécifications des démographes visent à expliquer les variations de mortalité observées entre les groupes d'individus).

3.2. Taux de mortalité constants par morceaux

3.2.1. Teneur de l'hypothèse

Cette section est consacrée à l'hypothèse que nous retiendrons par la suite. Dorénavant, nous supposons que les taux instantanés de mortalité sont constants par morceaux, i.e.

$$\mu_{x+\xi} = \mu_x \text{ pour } 0 \leq \xi < 1 \text{ et } x \text{ entier.} \quad (9)$$

3.2.2. Résultats qui en découlent

On vérifie facilement que, dans le cas où (9) est vérifiée, taux m_x et taux instantanés μ_x de mortalité coïncident. De plus,

$$\begin{aligned} \mu_x &= -\ln(1 - q_x) \\ {}_x p_x &= \exp(-\xi \mu_x) = \{p_k\} \text{ pour } 0 \leq \xi < 1. \end{aligned}$$

Enfin, l'exposition au risque à l'âge x , notée ER_x , est liée au quotient de mortalité par la relation

$$ER_x = \frac{-L_x q_x}{\ln(1 - q_x)}$$

où L_x est le nombre de survivants à l'âge x .

3.2.3. Estimation brute des taux de mortalité

Sous l'hypothèse (9), l'estimateur du maximum de vraisemblance de μ_x est donné par

$$\hat{\mu}_x = \frac{D_x}{ER_x} \quad (10)$$

où D_x est le nombre observé de décès à l'âge x , pour une exposition au risque ER_x . On en déduit alors les estimations des quotients de mortalité grâce à la formule

$$\hat{q}_x = 1 - \hat{p}_x = 1 - \exp(-\hat{\mu}_x).$$

3.3. Modèles envisagés

Idéalement, les indicateurs choisis devraient être indépendants. Ainsi, choisir de relier \hat{q}_0 et \hat{q}_0^{ref} nécessite de prendre en compte lors de l'estimation des paramètres la structure de corrélation entre les \hat{q}_0 pour différentes valeurs de x . Cette dépendance est très forte puisque

$$\hat{q}_{x+1} = \hat{q}_x + (1 - \hat{q}_x)\hat{q}_x.$$

Afin de contourner cette difficulté, notre préférence va aux probabilités annuelles de décès \hat{q}_x , aux probabilités annuelles de survie \hat{p}_x et aux taux de mortalité $\hat{\mu}_x$. Les taux ou les quotients de mortalité seront préalablement transformés de manière à lever les contraintes pesant sur eux: nous travaillerons donc avec le logarithme des taux de mortalité et avec le logit des quotients de mortalité.

Différents modèles relationnels peuvent être envisagés, en fonction des statistiques disponibles. Nous les passons en revue ci-dessous.

3.3.1. Seuls les taux de mortalité sont disponibles

Dans ce cas, un modèle du type

$$\ln \hat{\mu}_x = f(\ln \mu_x^{\text{ref}}) + \varepsilon_x \text{ avec } \varepsilon_x \sim \text{Nor}(0, \sigma^2), x = \alpha, \alpha + 1, \dots, \omega, \quad (11)$$

est envisageable.

Il est possible d'introduire des poids dans le modèle (11) afin de tenir compte d'une précision d'estimation différente selon l'âge. Dans ce cas, on aura recours à un ajustement par moindres carrés pondérés, les poids reflétant alors la précision de l'estimation de μ_x , généralement décroissante avec l'âge x .

Le modèle (11) est un modèle additif: la fonction $f(\cdot)$ qui lie les taux de référence aux taux de la population d'intérêt est estimée sur base des données disponibles. Notez que ce modèle est très souple: la seule hypothèse formulée à propos de la fonction f est qu'elle soit régulière. On peut donc capturer un lien non-linéaire entre les taux de mortalité.

Remarque 3.1. Le cas échéant, l'estimation \hat{f} ainsi obtenue peut ensuite servir à identifier une forme paramétrique raisonnable pour f . Ainsi, BROUHNS & DENUIT (2002) et BROUHNS, DENUIT & VERMUNT (2002) ont mis en évidence une relation linéaire sur une échelle doublement logarithmique entre la mortalité de la population belge et celles des rentiers (i.e., la fonction intervenant dans (11) est de la forme $f(\xi) = \alpha + \beta\xi$).

Si nécessaire, on peut enrichir le modèle (11) en y incluant un effet de l'âge. Ceci peut s'avérer utile dans le cas où la mortalité des deux populations diffère sur certaines

tranches d'âges. On passe alors à un modèle du type

$$\ln \hat{\mu}_x = c + f_1(x) + f_2(\ln \mu_x^{\text{ref}}) + \varepsilon_x \text{ avec } \varepsilon_x \sim \text{Nor}(0, \sigma^2), \quad (12)$$

où la constante c apparaît par souci d'identifiabilité. Il s'agit cette fois d'estimer les fonctions f_1 et f_2 apparaissant dans (12).

3.3.2. Les nombres de décès et les expositions au risque sont disponibles

Dans certains cas, les éléments constitutifs de $\hat{\mu}_x$, à savoir nombre de décès et exposition au risque correspondante, sont disponibles. On aura alors recours au modèle de régression poissonnien

$$D_x \sim \text{Poi}(\text{ER}_x \exp(f(\ln \mu_x^{\text{ref}}))), x = \alpha, \alpha + 1, \dots, \omega, \quad (13)$$

qui lève l'hypothèse d'homoscédasticité des erreurs intervenant dans (11).

Le modèle de Poisson (13) apparaît naturellement sous l'hypothèse (9). Bien entendu, le modèle poissonnien (13) n'est qu'une approximation de la réalité, puisque ce modèle permet avec une probabilité positive d'observer plus de L_x décès, ce qui est clairement impossible. Néanmoins, ce modèle fournit souvent une excellente approximation. Ceci est d'autant plus vrai qu'en vertu du théorème central-limite pour les lois discrètes, D_x est approximativement de loi de Poisson pour L_x suffisamment grand. BRILLINER (1986) a montré que sous des hypothèses raisonnables formulées à propos des processus gouvernant les naissances et les décès, la loi de Poisson apparaissait naturellement pour modéliser le nombre de décès aux différents âges.

En plus de $\ln(\mu_x^{\text{ref}})$, on peut également expliquer μ_x en fonction de l'âge x en passant à un modèle du type

$$D_x \sim \text{Poi}(\text{ER}_x \exp(c + f_1(x) + f_2(\ln \mu_x^{\text{ref}}))), x = \alpha, \alpha + 1, \dots, \omega,$$

où il s'agira d'estimer les fonctions f_1 et f_2 .

3.3.3. Seuls les quotients de mortalité sont disponibles

Lorsque seules les estimations brutes \hat{q}_x des quotients de mortalité sont disponibles, le modèle

$$\text{logit}(\hat{q}_x) = f(\text{logit}(q_x^{\text{ref}})) + \varepsilon_x \text{ avec } \varepsilon_x \sim \text{Nor}(0, \sigma^2),$$

est envisageable. Il correspond à (11), adapté pour les quotients de mortalité. Une extension tenant compte de l'âge x et de $\text{logit}(q_x^{\text{ref}})$ est aussi envisageable, de même qu'une pondération reflétant la précision de l'estimation des quotients en fonction de l'âge.

3.3.4. Les nombre de décès et les effectifs en début de période sont disponibles

Dans ce cas, deux options sont envisageables. Soit l'exposition au risque est reconstituée sous l'hypothèse (9), de sorte qu'on peut alors recourir au modèle de régression poissonnien (13) pour D_x . Soit on recourt au modèle de régression binomial

$$D_x \sim \text{Bin}\left(L_x, \frac{\exp(f(\text{logit}(q_x^{\text{ref}})))}{1 + \exp(f(\text{logit}(q_x^{\text{ref}})))}\right), x = \alpha, \alpha + 1, \dots, \omega.$$

A nouveau, une fonction de l'âge peut être introduite dans le modèle.

4. INFERENCE PAR MAXIMUM DE VRAISEMBLANCE LOCAL

4.1. Famille exponentielle naturelle

Supposons que les observations y_i (dans notre cas, les logarithmes des taux de mortalité, les logit des probabilités de décès ou de survie, les nombres de décès, etc.) sont indépendantes mais pas identiquement distribuées. L'indice générique i représente l'âge dans les problèmes relatifs à la mortalité (âge en années révolues, ou classes d'âges de 3 ou 5 ans par exemple). Spécifiquement, la densité $f(\cdot | \theta_i, \phi)$ de y_i provient de la famille exponentielle linéaire à un paramètre θ_i . La forme générique de cette densité (discrète ou continue) est

$$f(y_i | \theta_i, \phi) = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right), y_i \in S, \quad (14)$$

où S est un sous-ensemble des entiers non-négatifs N ou des réels R . Le paramètre θ_i , encore appelé paramètre naturel, est le paramètre d'intérêt tandis que ϕ est considéré comme un paramètre de nuisance. Les fonctions $b(\cdot)$ et $c(\cdot)$ déterminent la loi en question.

Les lois de probabilité utilisées plus haut pour modéliser la mortalité possèdent une densité pouvant se mettre sous la forme (14), comme en témoigne le Tableau 1.

Loi de proba.	Support S	Densité de proba.	Param. naturel $\theta = \theta(\mu)$	$b(\theta)$	ϕ
Bernoulli $\text{Bin}(1, \mu)$	$\{0, 1\}$	$\mu^y (1 - \mu)^{1-y}$	$\ln \frac{\mu}{1 - \mu}$	$\ln(1 + \exp(\theta))$	1
Poisson $\text{Poi}(\mu)$	N	$\exp(-\mu) \frac{\mu^y}{y!}$	$\ln \mu$	$\exp(\theta)$	1
Normale $\text{Nor}(\mu, \sigma^2)$	R	$\frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{\sigma^2} (y - \mu)^2\right)$	μ	$\theta^2/2$	σ^2

Table 1. Quelques lois de probabilité dont la densité est de la forme (14).

Remarque 4.1. Nous verrons qu'il est parfois utile d'introduire une pondération dans (14). Ceci nous amène à considérer que la densité de y_i est de la forme

$$f(y_i | \theta_i, \phi, \omega_i) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{\phi} \omega_i + c(y_i, \phi / \omega_i) \right\}, y_i \in S, \quad (15)$$

où ω_i est le poids accordé à la i ème observation.

On peut montrer que $E[Y] = b'(\theta)$ et $\text{Var}[Y] = \phi b''(\theta) / \omega$. Dès lors, la variance de Y apparaît comme le produit de deux fonctions: la première, $b''(\theta)$, qui dépend uniquement du paramètre θ (et donc de la moyenne μ) est appelée fonction variance et notée $V(\mu)$; la seconde est indépendante de θ et dépend uniquement de ϕ . Les fonctions variance $V(\cdot)$ intervenant dans les différents modèles présentés au Tableau 1 sont reprises au Tableau 2. Il est important de noter que, le cas de la loi normale mis à part, la variance est toujours fonction de la moyenne.

Loi de probabilité	Moyenne $\mu = b'(\theta)$	Fonction variance $V(\mu) = b''(\theta)$	Variance $\text{Var}[Y] = \phi b''(\theta) / \omega$
Bernoulli	$\mu = \frac{\exp \theta}{1 + \exp \theta}$	$\mu(1 - \mu)$	$\frac{\mu(1 - \mu)}{\omega}$
Poisson	$\mu = \exp \theta$	μ	$\frac{\mu}{\omega}$
Normale	$\mu = \theta$	1	$\frac{\sigma^2}{\omega}$

Table 2. Structure moyenne-variance associée aux lois de probabilité usuelles dont la densité est de la forme (15).

4.2. Modèle de régression

Lions à présent la variable réponse y_i aux variables explicatives \mathbf{x}_i . Ces variables explicatives sont généralement l'âge, le logarithme d'un taux de mortalité d'une population de référence ou le logit d'un quotient de mortalité d'une population de référence. Notant μ_i la moyenne de Y_i , on suppose que les variables explicatives sont liées à la moyenne par l'intermédiaire d'un score η_i , i.e.

$$g(\mu_i) = \eta_i = c + \sum_{j=1}^p f_j(x_{ij}) \quad (16)$$

où la fonction monotone et dérivable g est appelée fonction de lien, et où les fonctions $f_j(\cdot)$, $j=1, \dots, p$, supposées régulières, traduisent l'influence des variables explicatives x_1, \dots, x_p sur la réponse y . L'appellation de modèle additif provient de la forme additive du

score dans (16). Plus généralement, la forme additive (16) peut être abandonnée au profit de

$$g(\mu_i) = \eta_i = c + f(x_{i1}, x_{i2}, \dots, x_{ip}).$$

4.3. Lien canonique

Chacune des lois de probabilité de la famille exponentielle linéaire possède une fonction de lien spécifique, dite fonction de lien canonique, définie par $\theta = \eta$, où θ est le paramètre canonique. Le lien canonique est tel que $g(\mu_i) = \theta_i$. Or, $\mu_i = b'(\theta_i)$ d'où $g^{-1} = b'$. Les fonctions canoniques sont reprises au Tableau 3. On voit ainsi que la loi de Poisson privilégie le lien logarithmique, alors que la loi de Bernoulli privilégie le lien logit.

Loi de probabilité	Fonction de lien canonique
Normale	$\eta = \mu$
Poisson	$\eta = \ln \mu$
Bernoulli	$\eta = \ln \mu - \ln(1 - \mu)$

Table 3. Liens canoniques associés aux lois de probabilité usuelles dont la densité est de la forme (15).

4.4. Méthode Loess

4.4.1. Importance du cas gaussien

Les ajustements des modèles linéaires et additifs généralisés passent par une succession d'ajustements de pseudo-variables à l'aide de techniques de moindres carrés pondérés. Il s'agit donc en fait de maîtriser les techniques relatives aux modèles à erreurs gaussiennes.

Commençons par considérer le modèle élémentaire suivant. Nous disposons de n observations (x_i, y_i) , $i = 1, \dots, n$, où la variable à expliquer y_i et le régresseur x_i sont continues. L'influence de x_i sur y_i est modélisée à l'aide d'une fonction $f(\cdot)$, supposée régulière, à un bruit additif près, i.e.,

$$y_i = f(x_i) + \varepsilon_i \text{ où } \varepsilon_i \sim \text{No}(0, \sigma^2). \quad (17)$$

Les erreurs ε_i sont de surcroît supposées indépendantes. La spécification (17) permet donc de s'affranchir de la contrainte de linéarité imposée dans la régression classique. L'estimation de $f(\cdot)$ se fera à l'aide des techniques de lissage et nous nous intéresserons plus particulièrement à des lisseurs linéaires, en ce sens que $\hat{f}(x_i)$ peut s'exprimer comme une combinaison linéaire des valeurs y_1, \dots, y_n , i.e.

$$\hat{f}(x_i) = \sum_{j=1}^n h_{ij} y_j \quad (18)$$

où les poids $h_{ij} = h(x_i, x_j)$ dépendent de l'endroit x_i où la réponse $f(x_i)$ doit être estimée. Si on définit le vecteur $\mathbf{f} = (f(x_1), \dots, f(x_n))^t$, (18) permet d'écrire la relation matricielle $\hat{\mathbf{f}} = \mathbf{H}\mathbf{y}$ (par convention, tous les vecteurs sont des colonnes et l'exposant t indique la transposition).

4.4.2. Principe: moindres carrés pondérés

Cette méthode proposée par CLEVELAND (1979) fait partie de la famille des régressions polynomiales locales. Elle consiste à approximer localement $f(\cdot)$ par une droite⁴. C'est ce qu'ont proposé CLEVELAND ET AL. (1988, 1991). L'idée est d'utiliser les λ plus proches voisins de x afin d'estimer $f(x)$. Le voisinage s'apprécie par rapport aux variables explicatives: on utilise les λ observations dont les variables explicatives sont les plus proches de x pour estimer la réponse $f(x)$.

La méthode Loess peut se décomposer comme suit : disposant des n observations (x_i, y_i) , $i = 1, 2, \dots, n$,

1. les λ plus proches voisins de x sont identifiés (soit $V(x)$ l'ensemble de ceux-ci);
2. la distance $\Delta(x)$ séparant x du plus éloigné de ses λ plus proches voisins est calculée comme

$$\Delta(x) = \max_{i \in V'(x)} |x - x_i| ;$$

3. les poids

$$w_i(x) = K\left(\frac{|x - x_i|}{\Delta(x)}\right)$$

sont assignés à chaque élément de $V(x)$. La fonction $K(\cdot)$ assignant les poids $w_i(x)$ aux observations de $V(x)$ doit avoir les propriétés suivantes:

- (a) $K(u) \geq 0$ pour tout u
- (b) $K(u) = 0$ pour $u > 1$
- (c) K est non-croissante sur $(0, 1)$.

Comme à l'accoutumée, nous utiliserons la fonction $K(\cdot)$ donnée par

$$K(u) = \begin{cases} (1 - u^3)^3 & \text{pour } 0 \leq u < 1 \\ 0 & \text{sinon.} \end{cases}$$

On donne donc plus de poids aux points du voisinage les plus proches de x .

⁴Notez qu'on peut également approximer localement f par une constante, auquel cas, on retrouve des moyennes mobiles pondérées, ou par un polynôme de degré 2. Nous ne considérerons ici que l'ajustement linéaire local.

4. $\hat{f}(x)$ est obtenue en régressant les y_i , $i \in V(x)$, sur les x_i correspondants à l'aide d'un ajustement des moindres carrés pondérés et en se servant de la droite de régression pour prédire la réponse correspondant à x .

Cette approche fournit une réponse de la forme (18).

Dans le cas d'un seul régresseur, la valeur ajustée $\hat{y}_i = \hat{\beta}_0(x_i) + \hat{\beta}_1(x_i)x_i$ s'obtient en déterminant $\hat{\beta}_0(x_i)$ et $\hat{\beta}_1(x_i)$ de façon à minimiser

$$\sum_{k \in V'(x_i)} w_k(x_i)(y_k - \beta_0(x_i) - \beta_1(x_i)x_k)^2$$

ce qui fournit

$$\hat{y}_i = \sum_{k=1}^n h_k(x_i)y_k,$$

où $h_k(x_i)$ ne dépend pas des y_j , $j=1, \dots, n$ ($h_k(x_i)$ ne dépend que des régresseurs).

Si on s'intéresse à la réponse pour une valeur x non observée, le modèle utilisé pour estimer $f(x)$ est donc

$$y_i = \beta_0(x) + \beta_1(x)x_i + \varepsilon_i \text{ pour } i \in V(x)$$

où les estimations $\hat{\beta}_0(x)$ et $\hat{\beta}_1(x)$ des paramètres $\beta_0(x)$ et $\beta_1(x)$ s'obtiennent en minimisant

$$\sum_{k \in V'(x)} w_k(x)(y_k - \beta_0(x) - \beta_1(x)x_k)^2.$$

Ceci donne finalement $\hat{f}(x) = \hat{\beta}_0(x) + \hat{\beta}_1(x)x$.

4.4.3. Intervalles de confiance

Si on note $\hat{\mathbf{y}}$ le vecteur des valeurs ajustées $(\hat{f}(x_1), \dots, \hat{f}(x_n))^t$ et $\hat{\varepsilon}$ le vecteur des résidus, on a

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \text{ et } \hat{\varepsilon} = (\mathbf{I} - \mathbf{H})\mathbf{y}.$$

Par conséquent, $\hat{\mathbf{y}}$ et $\hat{\varepsilon}$ obéissent tous deux aux lois normales multivariées de matrice variance covariance $\sigma^2 \mathbf{H}\mathbf{H}^t$ et $\sigma^2 (\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})^t$, respectivement. Ceci nous permet d'obtenir des intervalles de confiance pour $f(x)$. Si nous posons

$$\delta_k = \text{trace}((\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})^t)^k \text{ pour } k = 1, 2,$$

on voit facilement que

$$\mathbb{E} \left[\sum_{i=1}^n \hat{\varepsilon}_i^2 \right] = \sigma^2 \delta_1$$

de sorte que

$$\hat{\sigma}^2 = \frac{1}{\delta_1} \sum_{i=1}^n \varepsilon_i^2$$

est un estimateur sans biais de σ^2 . De plus,

$$\frac{\delta_1^2}{\delta_2} \times \frac{\hat{\sigma}^2}{\sigma^2}$$

suit approximativement la loi chi-carrée à δ_1^2/δ_2 degrés de liberté (où δ_1^2/δ_2 sera arrondi à l'entier le plus proche). Par conséquent,

$$\frac{\hat{f}(x) - f(x)}{\hat{\sigma} \sqrt{\sum_{i=1}^n h_i^2(x)}}$$

obéit approximativement à la loi de Student à δ_1^2/δ_2 degrés de liberté. Ceci permet d'obtenir des intervalles de confiance pour la réponse $f(\cdot)$ en différents points x .

4.4.4. Degré de lissage

Comme le lecteur aura pu s'en rendre compte, l'approche Loess dépend donc du nombre λ de points contenu dans le voisinage $V(x_0)$. Le nombre de plus proches voisins, le plus souvent exprimé en pourcentage de la taille du jeu de données, joue le rôle de paramètre de lissage. La sélection d'une valeur optimale pour λ est considérée plus bas (en minimisant le critère AICC1, présenté dans la section suivante).

Lors de la comparaison des modèles, il est bon de disposer d'une mesure de la complexité de ceux-ci. Ceci peut s'effectuer à l'aide d'un nombre de degrés de liberté associé aux lisseurs satisfaisant (18). Le nombre de degrés de liberté DF_λ est fourni par la trace de la matrice \mathbf{H}_λ dont les éléments sont les h_{ij} intervenant dans (18). Ce choix provient du fait que dans le modèle de régression linéaire classique, la trace de la matrice qui fait passer des observations y_i aux valeurs ajustées $\hat{y}_i = \hat{\beta}' \mathbf{x}_i$ vaut le nombre de paramètres (i.e. la dimension de β).

4.4.5. Mesure de la qualité de l'ajustement

La mesure de la qualité de l'ajustement est plus complexe pour les modèles où il s'agit d'estimer une fonction que pour les modèles paramétriques classiques. La plupart des critères font intervenir à la fois une mesure de la qualité de l'ajustement et une mesure de la complexité du modèle. La minimisation de tels critères permet également de sélectionner le paramètre de lissage.

Notons $\hat{\sigma}^2$ la somme des carrés des résidus. HURVICH & SIMONOFF (1998) ont proposé le critère suivant:

$$AICC1 = n \ln \hat{\sigma}^2 + n \frac{\delta_1/\delta_2(n + DF_\lambda)}{\delta_1^2/\delta_2 - 2}$$

appelé AIC corrigé, où

$$\delta_1 = \text{trace}((\mathbf{I} - \mathbf{H}_\lambda)'(\mathbf{I} - \mathbf{H}_\lambda))$$

$$\delta_2 = \text{trace}(((\mathbf{I} - \mathbf{H}_\lambda)'(\mathbf{I} - \mathbf{H}_\lambda))^2).$$

Le premier terme intervenant dans AICC1 mesure la qualité de l'ajustement tandis que le deuxième évalue la complexité du modèle. Ce critère permet de sélectionner la valeur optimale de λ (i.e. celle minimisant AICC1).

4.4.6. Extension à des régresseurs multiples

L'extension de la méthode Loess à plus d'un régresseur est immédiate. On considère alors un modèle de la forme

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{ip}) + \varepsilon_i.$$

Il suffit d'approcher localement la fonction des variables explicatives par un hyperplan et ensuite de procéder comme expliqué plus haut dans le cas d'un seul régresseur. Il importe dans ce cas que les différents régresseurs aient des valeurs comparables (afin que les voisinages multidimensionnels ne soient pas déterminés sur la seule base de la variable prenant les plus grandes valeurs). Pour cela, les variables explicatives sont souvent normalisées.

Remarque 2. Dans notre cas, la taille des jeux de données à traiter étant relativement modeste, il n'est pas nécessaire d'utiliser des " $k-d$ trees" pour définir des voisinages multidimensionnels.

4.4.7. Maximum de vraisemblance local

La méthode Loess s'étend naturellement aux modèles généralisés additifs. On aura alors recours à des ajustements locaux dans des modèles linéaires généralisés, par maximum de vraisemblance (exactement comme procédait la méthode Loess à partir d'une log-vraisemblance gaussienne). En pratique, ceci permet d'utiliser la procédure GENMOD de SASSTAT avec les poids proposés dans Loess. Plus précisément, étant donné un point x où on veut estimer la réponse, on détermine un voisinage $V(x)$ et des poids $w_i(x)$ exactement comme dans Loess.

Par exemple, si on veut de la sorte estimer les q_x , on déterminera le voisinage $V(x)$ et on estimera en x la fonction f intervenant dans

$$q_x = \frac{\exp(f(x))}{1 + \exp(f(x))}$$

en minimisant

$$L(\beta_0(x), \beta_1(x)) = \sum_{\xi \in V(x)} w_\xi(x) \ln \ell_\xi(\beta_0(x), \beta_1(x))$$

où

$$\ell_{\xi}(\beta_0(x), \beta_1(x)) = \binom{L_{\xi}}{d_{\xi}} (q_{\xi}(x))^{d_{\xi}} (1 - q_{\xi}(x))^{L_{\xi} - d_{\xi}}$$

avec

$$q_{\xi}(x) = \frac{\exp(\beta_0(x) + \beta_1(x)\xi)}{1 + \exp(\beta_0(x) + \beta_1(x)\xi)}.$$

Finalement,

$$\hat{q}_x = \frac{\exp(\hat{f}(x))}{1 + \exp(\hat{f}(x))}$$

avec

$$\hat{f}(x) = \hat{\beta}_0(x) + \hat{\beta}_1(x)x.$$

5. ILLUSTRATIONS NUMERIQUES

5.1. Portefeuille 1

5.1.1. Présentation des données

Le premier jeu de données concerne un portefeuille standard d'assurance décès en France. Nous disposons de l'exposition au risque initiale et du nombre de décès de 30 à 70 ans, séparément pour les hommes et les femmes.

Les tables de référence potentielles sont:

- la table TD88-90 qui est souvent utilisée en tarification des produits décès en France
- la table AFAM utilisée dans le même but au Royaume-Uni.

5.1.2. Approche par maximum de vraisemblance local

Dans un premier temps, nous proposons de recourir au modèle binomial

$$D_x \sim \text{Bin}(L_x, q_x)$$

où q_x est fonction de l'âge et du logit du quotient de mortalité d'éventuelles tables de référence. Ce modèle est estimé par la méthode du maximum de vraisemblance local. Cette méthode possède l'avantage de pouvoir comparer plusieurs combinaisons de fonctions de variables explicatives, et différents voisinages, via le critère *AICC1* défini précédemment. Nous testons dans un premier temps les combinaisons suivantes de variables explicatives:

Modèle	Hommes	Femmes
1	Age	Age
2	$logit(q^{AM92})$	$logit(q^{AF92})$
3	$logit(q^{TD88-90})$	$logit(q^{TD88-90})$
4	$Age, logit(q^{AM92})$	$Age, logit(q^{AF92})$
5	$Age, logit(q^{TD88-90})$	$Age, logit(q^{TD88-90})$
6	$logit(q^{AM92}), logit(q^{TD88-90})$	$logit(q^{AF92}), logit(q^{TD88-90})$
7	$Age, logit(q^{AM92}), logit(q^{TD88-90})$	$Age, logit(q^{AF92}), logit(q^{TD88-90})$

Définissons δ comme le nombre d'âges pris en considération dans le voisinage entourant chaque âge x fixé. Plus précisément, chaque voisinage sera donc constitué par $2\delta + 1$ points (excepté près des bornes de l'intervalle d'observation, i.e. 30-70 dans notre cas).

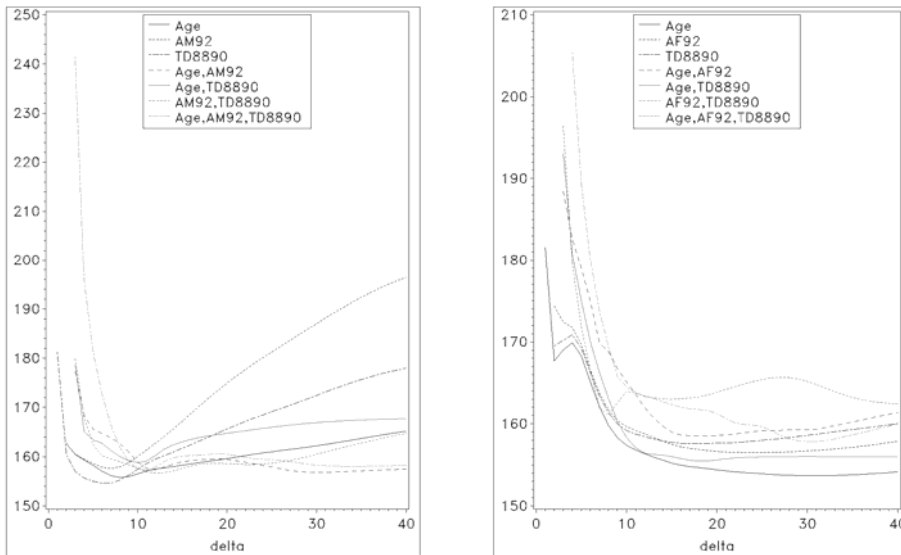


Figure 1. Maximum de vraisemblance local dans modèle binomial: critère AICC1 pour quelques modèles et quelques voisinages. δ représente le nombre d'âges pris dans le voisinage de chaque côté de l'âge considéré. Les hommes sont à gauche et les femmes à droite.

Les valeurs du critère $AICC1$ pour l'ensemble de ces modèles sont reprises à la Figure 1. Chez les hommes, le minimum est constaté pour le modèle avec la table TD 88-90 et $\delta = 6$. Il s'agit donc du modèle optimal au sens du maximum de vraisemblance local. Au contraire, le modèle féminin ne fait intervenir que l'âge, avec la valeur optimale $\delta = 31$. Ce constat peut s'expliquer par le fait que la table TD 88-90, utilisée dans le cadre d'opérations en cas de décès, est basée (conformément à un principe de prudence) sur la mortalité masculine. Il semble donc naturel que cette table soit retenue chez les hommes mais pas chez les femmes.

L'un des avantages de la méthode du maximum de vraisemblance et de pouvoir fournir des intervalles de confiance pour le lissage. En effet, s'agissant de modèle linéaires généralisés locaux, de tels intervalles peuvent être facilement obtenus. La Figure 2 expose ainsi le lissage opéré par les 2 modèles optimaux sur le score (i.e. le logit des quotients de mortalité observés). Des intervalles de confiance à 95% sont également visibles.

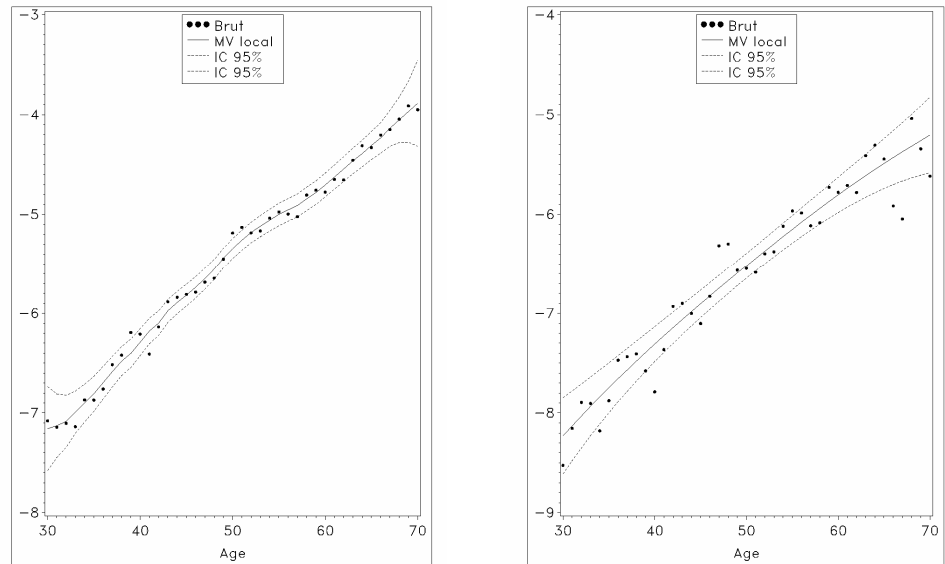


Figure 2. Maximum de vraisemblance local dans modèle binomial: prédiction du score pour le modèle optimal (TD 88-90 pour les hommes, l'âge pour les femmes). Les hommes sont à gauche et les femmes à droite.

5.2. Portefeuille 2 - France - assurance décès - analyse en nombre d'assurés

5.2.1. Présentation des données

Le deuxième portefeuille traite d'assurance décès en France avec des petits effectifs. Nous disposons de l'exposition au risque initiale, de l'exposition au risque centrale et du nombre de décès de 23 à 79 ans. C'est un cas idéal pour comparer les résultats obtenus à partir du modèle binomial

$$D_x \sim \text{Bin}(L_x, q_x)$$

où q_x est fonction de l'âge et du logit du quotient de mortalité d'éventuelles tables de référence, et du modèle de Poisson

$$D_x \sim \text{Poi}(ER_x \mu_x)$$

où μ_x est fonction de l'âge et du logarithme du taux instantané de mortalité d'éventuelles tables de référence.

5.2.2. Approche par maximum de vraisemblance local

Comme pour le premier portefeuille, commençons par choisir le modèle optimal. Nous choisissons de tester les modèles suivants:

Modèle	Variables
1	Age
2	x^{AF92}
3	x^{AM92}
4	$x^{TD88-90}$
5	Age, x^{AF92}
6	Age, x^{AM92}
7	$Age, x^{TD88-90}$
8	$Age, x^{AF92}, x^{AM92}, x^{TD88-90}$

La variable x représente respectivement le logit du quotient de mortalité dans le cas du modèle binomial et le logarithme du taux instantané de mortalité dans le cas du modèle de Poisson.

La Figure 3 illustre les valeurs du critère $AICC1$ pour quelques voisinages et les modèles choisis. Nous choisissons donc la table AM 92 et $\delta = 47$ pour le modèle binomial comme pour celui de Poisson. Les valeurs lissées du logit du quotient de mortalité (modèle binomial) et du logarithme du taux instantané de mortalité (modèle de Poisson) sont visibles à la Figure 4, avec intervalles de confiance à 95%. Remarquons que les deux graphiques sont très semblables.

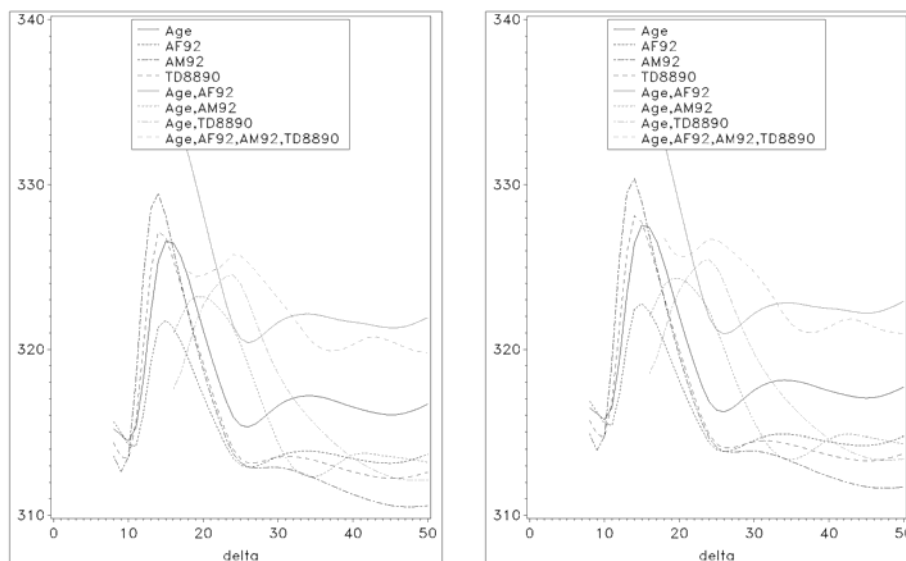


Figure 3. Maximum de vraisemblance local: critère AICC1 pour quelques modèles et quelques voisinages. δ représente le nombre d'âges pris dans le voisinage de chaque côté de l'âge considéré. Le modèle binomial est à gauche et le modèle de Poisson à droite.

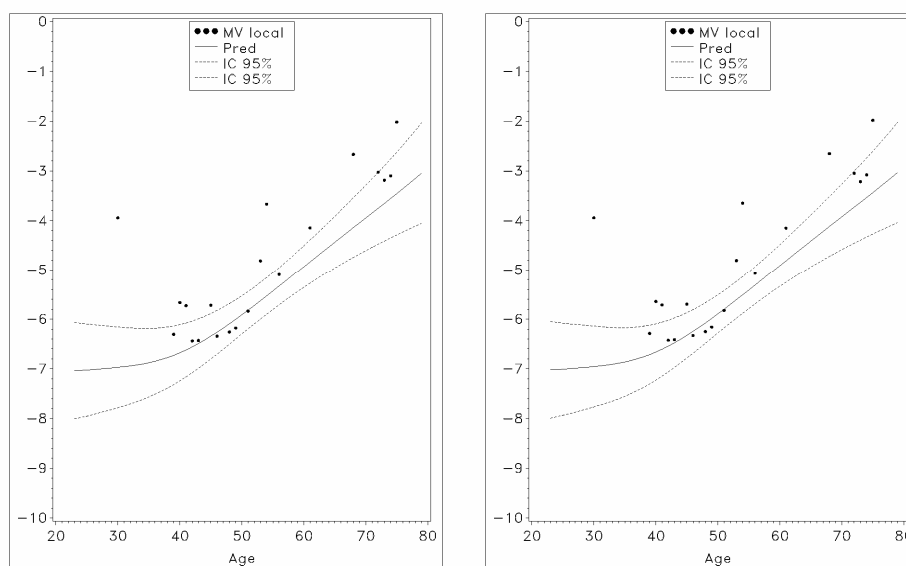


Figure 4. Maximum de vraisemblance local: prédiction du score pour le modèle optimal (AM 92). Le modèle binomial est à gauche et le modèle de Poisson à droite.

Afin de comparer les modèles, estimons les valeurs lissées par le modèle de Poisson du logit du quotient de mortalité, via la formule

$$\begin{aligned} \text{logit}(q_x^{\text{Poi}}) &= \ln \left(\frac{q_x^{\text{Poi}}}{1 - q_x^{\text{Poi}}} \right) \\ &= \ln \left(\frac{1 - \exp(-\mu_x^{\text{Poi}})}{\exp(-\mu_x^{\text{Poi}})} \right). \end{aligned} \quad (19)$$

De la même manière, les valeurs lissées par le modèle binomial du logarithme du taux instantané de mortalité sont données par

$$\ln(\mu_x^{\text{Bin}}) = \ln(-\ln(1 - q_x^{\text{Bin}})). \quad (20)$$

La Figure 5 permet ainsi de constater que les modèles engendrent des valeurs lissées très semblables.

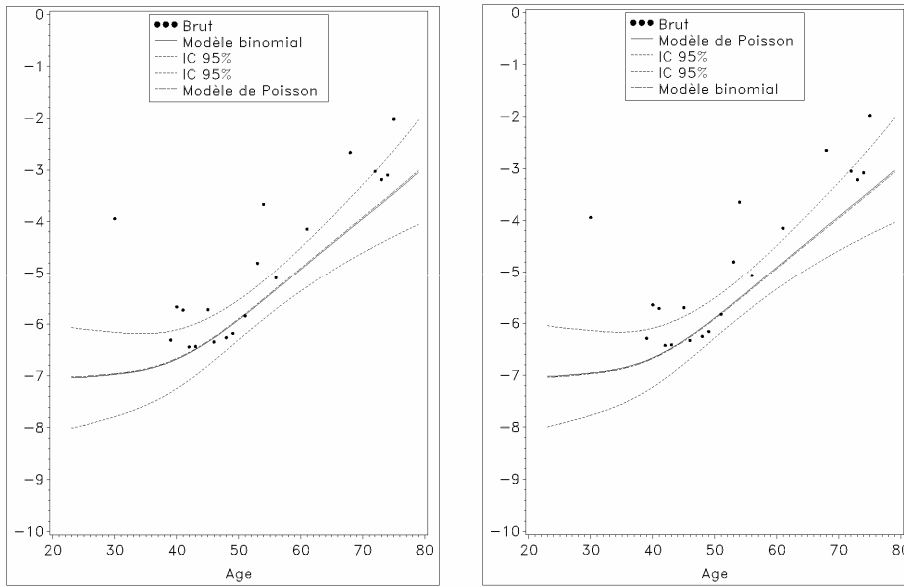


Figure 5. Maximum de vraisemblance local: prédiction du score pour le modèle optimal (AM 92). A gauche, le lissage de $\text{logit}(q_x)$ par le modèle binomial comparé à celui du modèle de Poisson après la transformation (19). A droite, le lissage de $\log(\mu_x)$ par le modèle de Poisson comparé à celui du modèle binomial après la transformation (20).

5.3. Portefeuille 3 - France - assurance décès - analyse en montant

5.3.1. Présentation des données

Le dernier portefeuille traite à nouveau d'assurance décès en France. Nous disposons de l'exposition au risque initiale, du nombre de décès de 14 à 74 ans pour les hommes et de 12 à 74 pour les femmes. Cette fois, les montants déboursés suite aux décès sont également disponibles.

5.3.2. Approche par maximum de vraisemblance local

Comme précédemment, nous commençons par sélectionner le modèle binomial puisque l'exposition au risque initiale est disponible. Divers voisinages sont envisagés et quelques tables de référence sont comparées. Au vu des résultats obtenus à partir des portefeuilles précédents, nous restreignons nos analyses aux modèles simples composés d'une unique table de référence:

Modèle	Hommes	Femmes
1	Age	Age
2	$logit(q^{AM92})$	$logit(q^{AF92})$
3	$logit(q^{TD88-90})$	$logit(q^{TD88-90})$

La Figure 6 compare les valeurs du critère $AICCl$ obtenues à partir des différents modèles et voisinages. Dans le cas masculin comme dans le cas féminin, c'est le plus grand voisinage ($\delta = 60$) qui est préféré. En d'autres termes, chaque voisinage comprend l'entièreté des données (seuls les poids diffèrent). Comme pour le premier portefeuille, c'est la table TD 88-90 qui est sélectionnée chez les hommes, tandis que l'âge est préféré chez les femmes, et ce tant pour l'analyse en nombre que pour celle en montant. Les valeurs lissées obtenues à partir de ces modèles optimaux sont illustrées à la Figure 7.

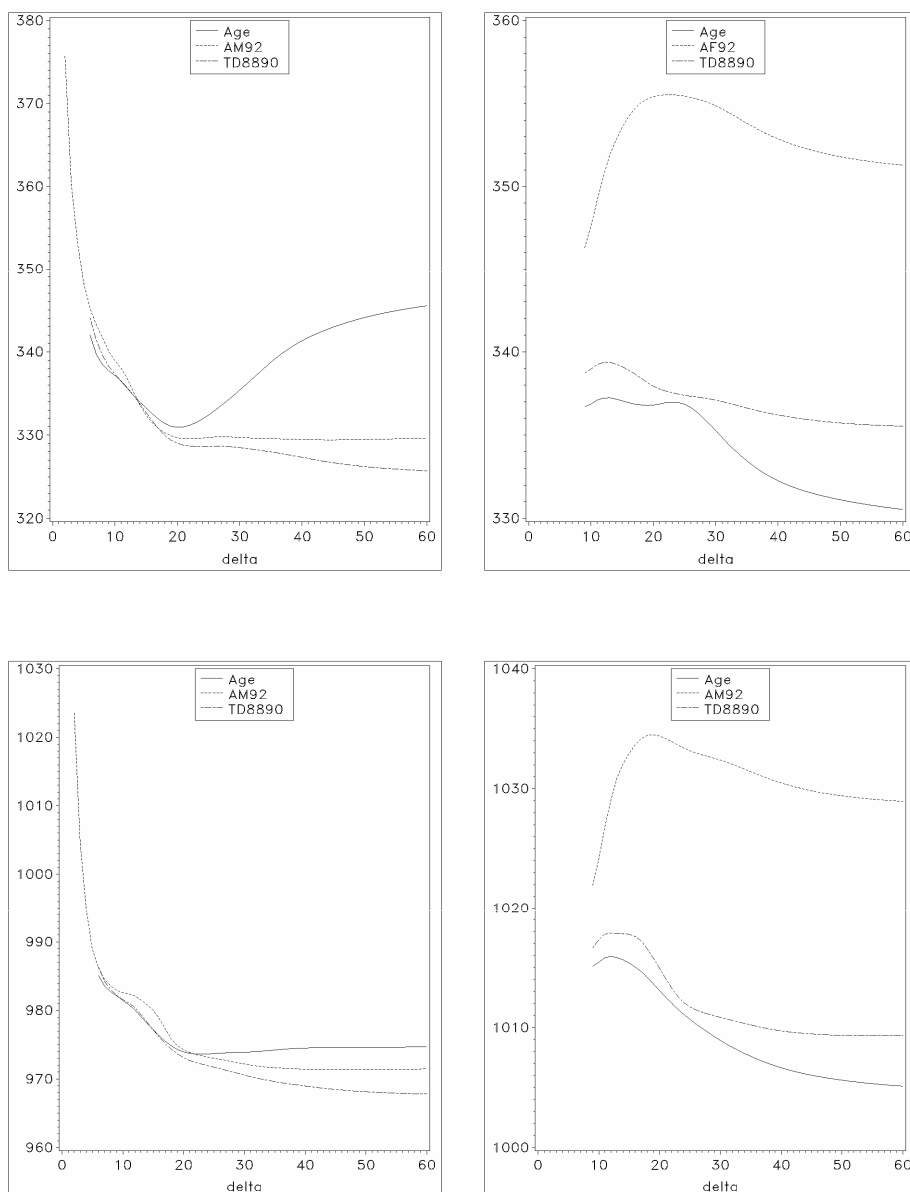


Figure 6. Maximum de vraisemblance local dans modèle binomial: critère AICC1 pour quelques modèles et quelques voisinages. δ représente le nombre d'âges pris dans le voisinage de chaque côté de l'âge considéré. Les hommes sont à gauche et les femmes à droite; analyse en nombre au-dessus, en montant en dessous.

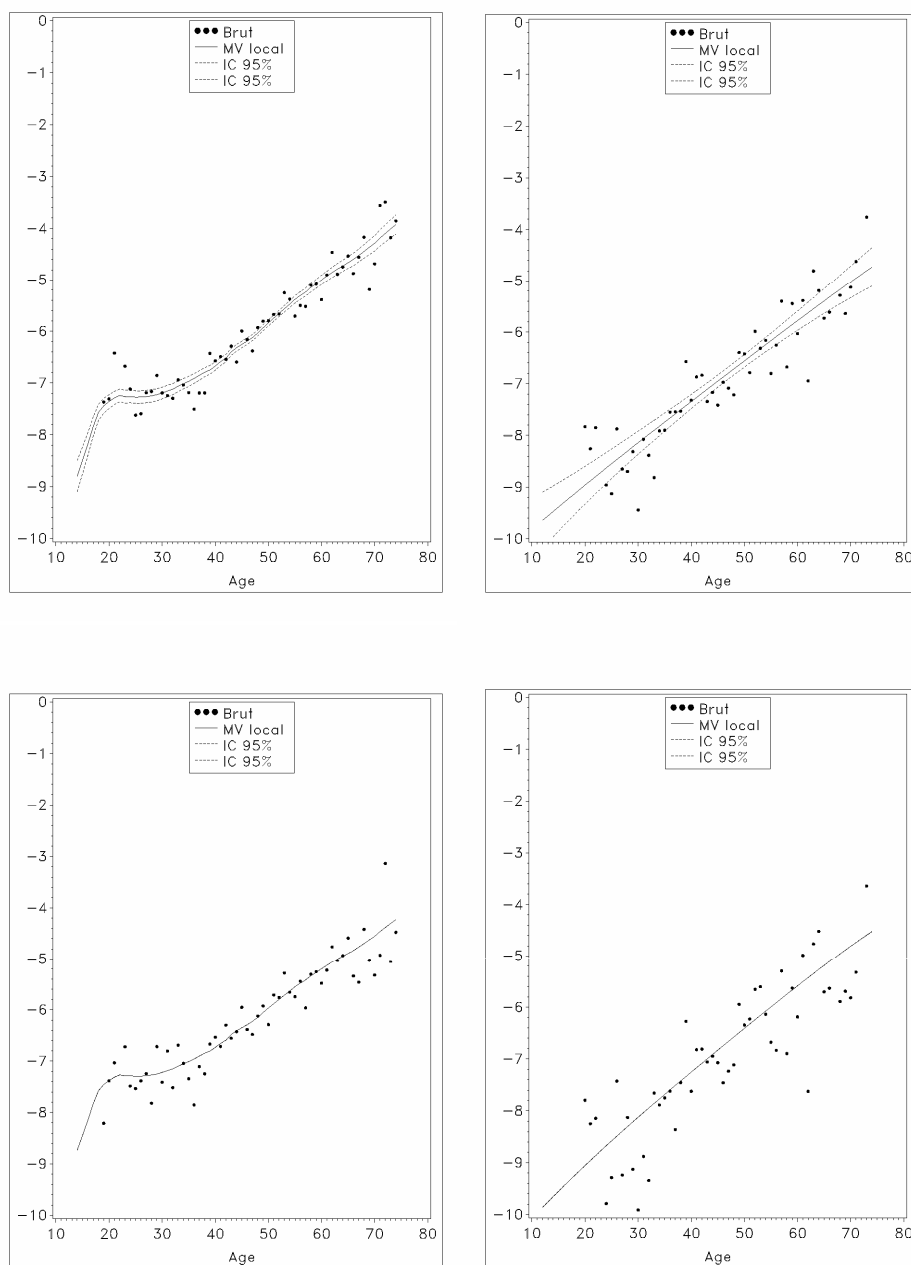


Figure 7. Maximum de vraisemblance local dans modèle binomial: prédiction du score pour le modèle optimal (TD 88-90 pour les hommes, l'âge pour les femmes). Les hommes sont à gauche et les femmes à droite;analyse en nombre au-dessus, en montant en-dessous.

REFERENCES

- BRASS, W. (1969). A generation method for projecting death rates. In *Population growth and the brain drain; Technique s and Methods of Study*, Bechhofer F. editor, Edinburgh University Press, pp. 75-91.
- BRASS, W. (1971). On the scale of mortality. In *Biological Aspects of Mortality*, Brass W. editor, Taylor & Francis Ltd., London, pp. 69-110.
- BRASS, W. (1974). Mortality models and their uses in demography. *Transactions of the Faculty of Actuaries* **33**, 122-133.
- BRILLINGER, D.R. (1986). The natural variability of vital rates and associated statistics. *Biometrics* **42**, 693-734.
- BROUHNS, N., & DENUIT, M. (2002). Risque de longévité et rentes viagères. III. Elaboration de tables de mortalité prospectives pour la population assurée belge, et évaluation du coût de l'antiselection. *Belgian Actuarial Bulletin* **2**, 64-72.
- BROUHNS, N., DENUIT, M., & VERMUNT, J.K. (2002). A Poisson log-bilinear approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* **31**, 373-393.
- CLEVELAND, W.S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829-836.
- CLEVELAND, W.S., & DEVLIN, S.J. (1988). Locally weighted regression: An approach to regression by local fitting. *Journal of the American Statistical Association* **83**, 596-610.
- CLEVELAND, W.S., DEVLIN, S.J., & GROSSE, E. (1988). Regression by local fitting. *Journal of Econometrics* **37**, 87-114.
- CLEVELAND, W.S., & GROSSE, E. (1991). Computational methods for local regression. *Statistics and Computing* **1**, 47-62.
- COX, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society – Series B Statistical Methodology* **34**, 187-220.
- DELWARDE, A., & DENUIT, M. (2005). *Construction de Tables de Mortalité Périodiques et Prospectives*. Economica, Paris (à paraître).
- EWBANK, D.C., GOMEZ-LEON, J.C., & STOTO, M.A. (1983). A reducible four-parameter system of model life tables. *Population Studies* **37**, 105-127.
- HANNERZ, H. (2001). An extension of relational methods in mortality estimation. *Demographic Research* **4**, Article 10.
- HURWICH, C.M., & SIMONOFF, J.S. (1998). Smoothing parameter selection in

- nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society - Series B* **60**, 271-293.
- SEAL, H. (1979). The fitting of a mathematical graduation formula: A historical review with illustrations. *German Actuarial Bulletin* **14**, 237-253.
- VERRALL, R.J. (1996). A unified framework for graduation. Actuarial Research Paper **91**, Department of Actuarial Science and Statistics, City University, London.
- ZABA, B. (1979). The four-parameter logit life table system. *Population Studies* **33**, 89-100.