Université Claude Bernard Lyon 1
Institut de Science Financière et d'Assurances

Prise en compte de l'hétérogénéité
dans la construction des tables
de mortalité prospectives de place en assurance

Julien Tomas - Frédéric Planchet
Institut de Science Financière et d'Assurances
Laboratoire de recherche de Sciences Actuarielle et Financière

Ce travail a bénéficié du support de l'Institut des Actuaires et de la Chair BNP Paribas Cardif "Management de la modélisation". Les points de vue exprimés dans ce document sont ceux de l'auteur et ne reflètent pas nécessairement ceux de l'Institut des Actuaires et de BNP Paribas Cardif.

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Motivations

- Nous nous plaçons dans la situation de la construction d'une table de mortalité prospective pour laquelle les données disponibles sont
  - constituées de groupes a priori hétérogènes, et
  - observées sur des périodes différentes.

Portefeuille	Age moyen à l'entrée	Age moyen à la sortie	Expo. moyenne	Age moyen au décès	Période d'observation	
					Début	Fin
P1	40.59	48.94	8.35	60.52	01/01/1995	31/12/2010
P2	41.54	45.36	3.818	52.3	01/01/2005	31/12/2010
P3	44.43	46.69	2.26	77.04	01/07/2004	30/06/2007
P4	51.43	61.74	10.31	77.92	01/01/1996	31/12/2007
P5	41.57	46.60	5.03	55.83	01/01/2003	31/12/2009
P6	47.76	55.22	7.45	73.82	01/01/1994	31/12/2009
P7	48.81	52.79	3.97	73.51	01/01/2006	31/12/2010
P8	46.44	55.10	3.66	62.15	01/01/2005	31/12/2009

TABLE: Statistiques observées par portefeuille, population masculine.

Motivations

- Sans prise en compte explicite de l'hétérogénéité, il est nécessaire de réduire la période d'observation à l'intersection des périodes d'observation des différentes populations.
- Cela s'avère pénalisant pour la détermination des tendances d'évolution de la mortalité et l'extrapolation
- Objectif : Proposer un modèle intégrant explicitement la prise en compte de l'hétérogénéité :
  - conserver l'ensemble de l'historique disponible pour toutes les populations, et
  - augmenter la profondeur de l'historique pour la détermination des tendances de mortalité.

Approche

i. Positionner les populations les unes par rapport aux autres à partir d'*un modèle à hasard proportionnel*, de type *modèle de Cox*.

ii. Ajuster les tables de mortalité par sexe par des *méthodes non-paramétriques de vraisemblance locale* où l'exposition observée de la population est pondérée par le coefficient obtenu à l'étape **i**.

iii. Identifier les composantes de la mortalité et leur importance dans le temps par une *analyse en composantes principales*.

iv. Extrapoler les paramètres variant dans le temps par des *méthodes de séries temporelles*.

Notations & hypothèses

Nombre d'individus, de décès, exposition et forces de mortalité

- À chacune des ces observations i, on associe une indicatrice δ_i indiquant si l'individu est décédé ou non,

$$\delta_i = \begin{cases} 1 & \text{si l'individu } i \text{ est décédé,} \\ 0 & \text{sinon,} \end{cases}$$

pour $i = 1, \ldots, L_{x,t}$.

- On définit τ_i, le temps durant lequel l'individu i est observé (c'est l'exposition au risque).
- Ainsi,

$$\sum_{i=1}^{L_{x,t}} \tau_i = \mathrm{E}_{x,t} \text{ et } \sum_{i=1}^{L_{x,t}} \delta_i = D_{x,t}.$$

- La force de mortalité à l'âge $x + \tau$ dans l'année calendaire t est définie par

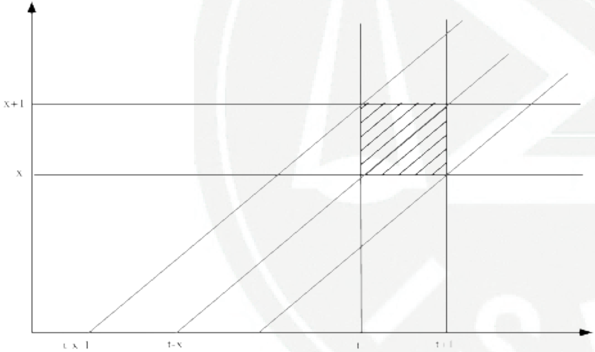
$$\varphi_{x+\tau}(t) = \lim_{\Delta\tau \to 0^+} \frac{\Pr\left[\tau < T_x(t) \leq \tau + \Delta\tau \middle| T_x(t) > \tau\right]}{\Delta\tau}$$

Notations & hypothèses

Hypothèse de mortalité constante par morceaux

- Dorénavant, on fait l'hypothèse de constance des forces de mortalité dans chaque carré unitaire du diagramme de Lexis :

$$\varphi_{x+\xi}(t+\zeta) = \varphi_x(t) \text{ pour } 0 \leq \xi < 1 \text{ et } 0 \leq \zeta < 1 \text{ et } x, t \text{ entiers.}$$



- Sous cette hypothèse, $p_x(t) = \exp\Big(-\int_0^1 \underbrace{\varphi_{x+\xi}(t+\xi)}_{=\varphi_x(t)}\Big) = \exp\big(-\varphi_x(t)\big)$.

Contents

Positionner les populations

Modèle de Cox à hazard proportionnel

- Pour la population p, nous avons le modèle suivant :

$$\varphi_x^p(t) = \alpha^p \, \varphi_x^0(t), \text{ avec } \varphi_x^0(t) \text{ inconnue et } \alpha^p = \exp\left(\boldsymbol{z}_p^T \theta_p\right).$$

- Le paramètre θ_p mesure l'influence d'appartenir à la population p sur l'intensité, i.e. la force de mortalité.

Portefeuille	coef	exp(coef)	se(coef)	z	Pr(>|z|)	Signif.
P1	0.18764	1.20640	0.02313	8.112	$4.44e-16$	$***$
P2	-0.18614	0.83015	0.02586	-7.199	$6.05e-13$	$***$
P3	0.41198	1.50980	0.02073	19.875	$< 2e-16$	$***$
P4	0.30471	1.35623	0.02359	12.917	$< 2e-16$	$***$
P5	0.30755	1.36008	0.01555	19.775	$< 2e-16$	$***$
P6	0.03091	1.03139	0.01405	2.200	0.0278	$*$
P7	-0.14341	0.86640	0.01801	-7.964	$1.67e-15$	$***$

TABLE: Coefficients estimés du modèle de Cox, population masculine.

Positionner les populations

Pondérer l'exposition au risque de chaque population I

- La contribution du $i$ème individu à la vraisemblance s'écrit,
  - si l'individu survie à son $(x+1)$ème anniversaire dans l'année calendaire t $(\delta_i = 0, \tau_i = 1)$ alors : $p_x(t) = \exp\big(-\varphi_x(t)\big)$;
  - si l'individu décède avant son $(x+1)$ème anniversaire dans l'année calendaire t $(\delta_i = 1, \tau_i < 1)$ alors : ${}_{\tau_i}p_x(t)\ \varphi_{x+\tau_i}(t) = \exp\big(-\tau_i\varphi_x(t)\big)\varphi_x(t)$.
- La contribution de l'individu i à la vraisemblance vaut donc

$$\exp\big(-\tau_i\varphi_x(t)\big)\big(\varphi_x(t)\big)^{\delta_i}.$$

- La vraisemblance devient

$$\mathcal{L}\big(\varphi_x(t)\big) = \prod_{i=1}^{L_{x,t}} \exp\big(-\tau_i\varphi_x(t)\big)\big(\varphi_x(t)\big)^{\delta_i} = \exp\big(-E_{x,t}\ \varphi_x(t)\big)\big(\varphi_x(t)\big)^{D_{x,t}},$$

- et $\log \mathcal{L}\big(\varphi_x(t)\big) = -E_{x,t}\ \varphi_x(t) + D_{x,t} \log \varphi_x(t)$.

Positionner les populations

Pondérer l'exposition au risque de chaque population II

- La vraisemblance $\mathcal{L}\big(\varphi_x(t)\big)$ est proportionnelle à la vraisemblance de Poisson basée sur $D_{x,t} \sim \mathcal{P}\big(E_{x,t}\,\varphi_x(t)\big)$.
- Ainsi pour population p : $D^p_{x,t} \sim \mathcal{P}\left(E^p_{x,t}\,\varphi^p_x(t)\right)$.
- Du fait de l'additivité de la loi de Poisson :

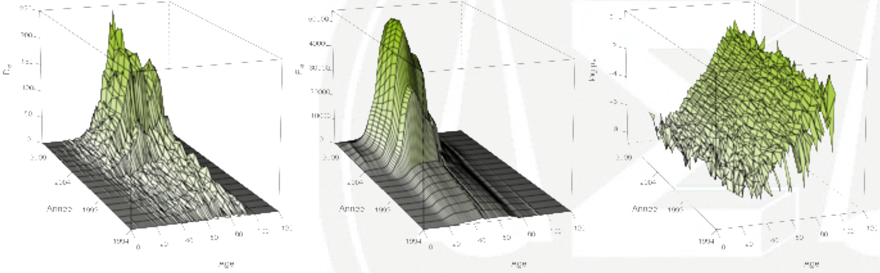
$$\sum_p D^p_{x,t} \sim \mathcal{P}\left(\sum_p E^p_{x,t}\,\varphi^p_x(t)\right), \text{d'où } \sum_p D^p_{x,t} \sim \mathcal{P}\left(\sum_p E^p_{x,t}\,\alpha_p\,\varphi^0_x(t)\right).$$

- On considérele modèle suivant :

$$D^\circ_{x,t} \sim \mathcal{P}\left(E^\circ_{x,t}\,\varphi^0_x(t)\right),$$

où $D^\circ_{x,t} = \sum_p D^p_{x,t}$ et $E^\circ_{x,t} = \sum_p E^p_{x,t}\,\alpha_p$.

Données agrégées et pondérées



(a) $D^{\circ}_{x,t}$ (b) $E^{\circ}_{x,t}$ (c) $\log \varphi_x(t)$

FIGURE: Statistiques agrégées et pondérées, population masculine.

Contents

Ajustement de la mortalité

Méthode de vraisemblance locale

- Estimer $\varphi_x^0(t)$ en considérant le modèle suivant :

$$D^\circ_{x,t} \sim \mathcal{P}\left(E^\circ_{x,t} \exp(\psi(x,t))\right),$$

  où $D^\circ_{x,t} = \sum_p D^p_{x,t}$, $E^\circ_{x,t} = \sum_p E^p_{x,t}\,\alpha_p$ et $\psi(x,t)$ est une fonction lisse non-spécifiée.

- L'idée principale : chaque force de mortalité est étroitement liée à ses voisines.

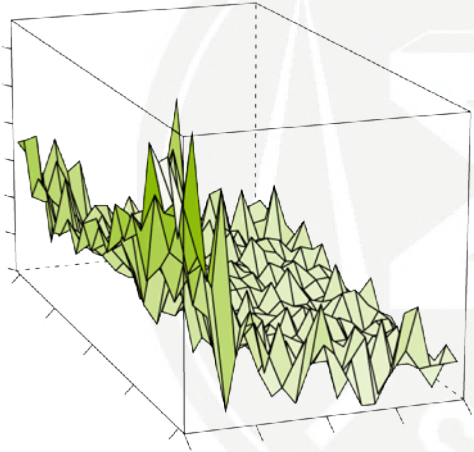
- Les observations φ_j^0 dans le voisinage de φ_i^0 contiennent de l'information à propos de la valeur de ψ au point $i = (x_i, t_i)$.

- Une estimation améliorée de φ_i^0 peut être obtenue en lissant les estimations individuelles φ_j^0.

- Ainsi la mortalité n'est pas résumée en a petit nombre de paramètres mais décrite par les n forces de mortalité.

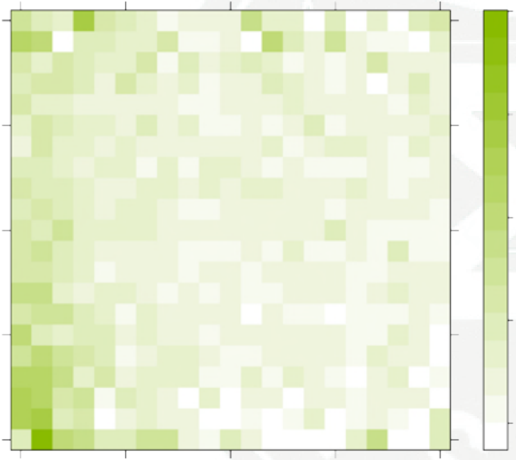
Méthode de vraisemblance locale

Comment estimer ψ ? Illustration de l'idée générale



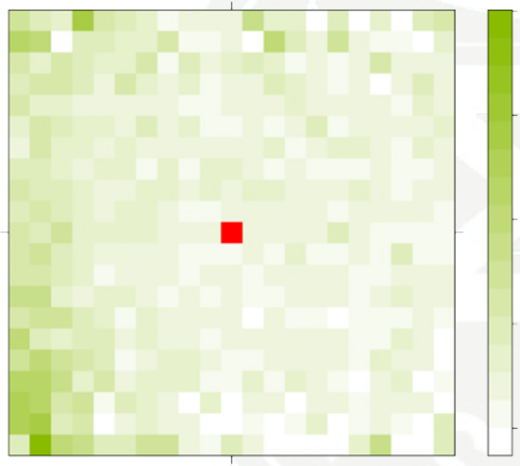
Méthode de vraisemblance locale

Comment estimer ψ ? Illustration de l'idée générale



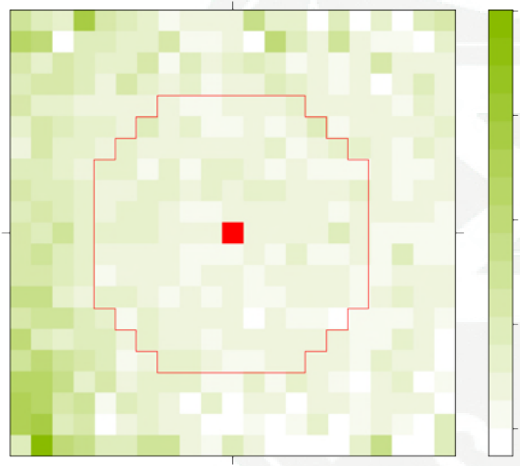
Méthode de vraisemblance locale

Comment estimer ψ ? Illustration de l'idée générale



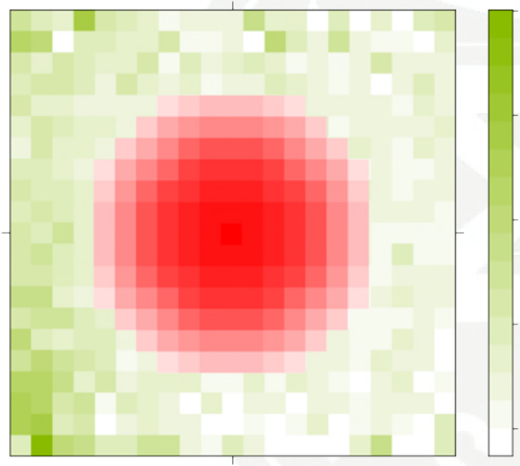
Méthode de vraisemblance locale

Comment estimer ψ ? Illustration de l'idée générale



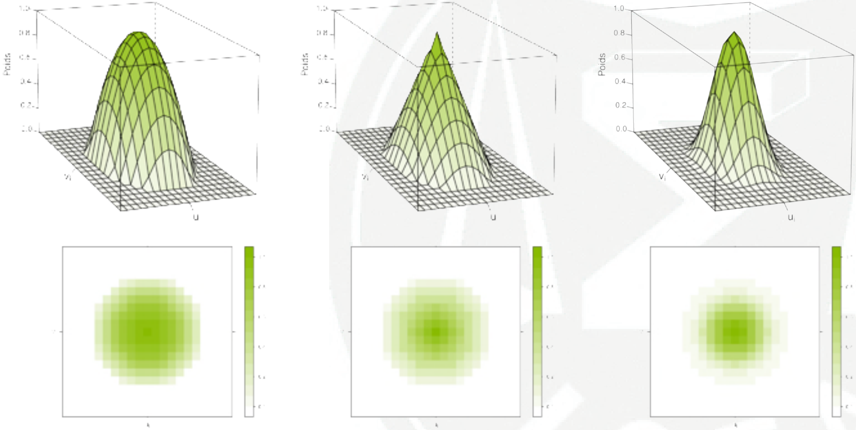
Méthode de vraisemblance locale

Comment estimer ψ ? Illustration de l'idée générale



Méthode de vraisemblance locale

La fonction de poids

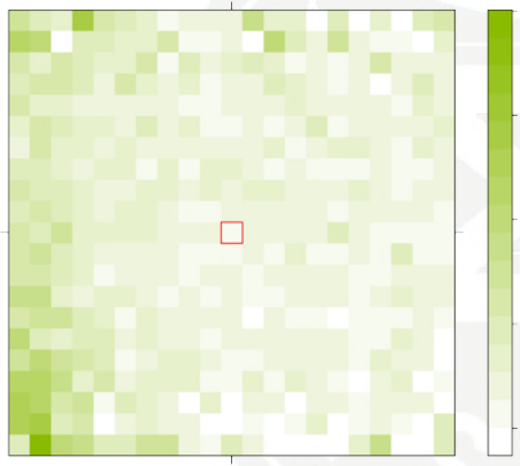


(a) Epanechnikov (b) Triangulaire (c) Triweight

FIGURE: Système de pondération pour des fonctions de poids avec $h = 7$

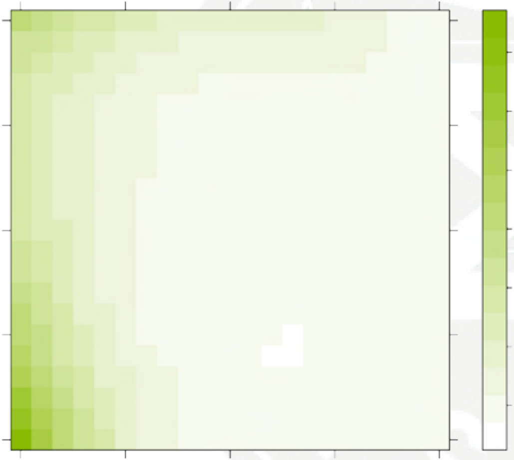
Méthode de vraisemblance locale

Comment estimer ψ ? Illustration de l'idée générale



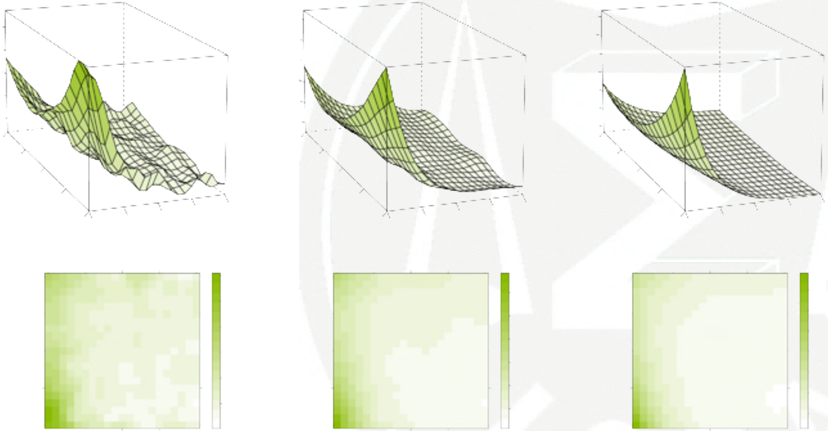
Méthode de vraisemblance locale

Comment estimer ψ ? Illustration de l'idée générale



Méthode de vraisemblance locale

Comparaison de l'ajustement avec différents voisinages



(a) $h = 3$ (29 obs.) (b) $h = 5$ (81 obs.) (c) $h = 10$ (317 obs.)

FIGURE: Comparaison de l'ajustement avec différents voisinages

Sélection des paramètres

Equilibre entre le biais et la variance

- On doit garder à l'esprit que l'ajustement et donc la sélection des paramètres de lissage est un compromis entre deux objectifs :
  - L'élimination des irrégularités, et
  - l'ajustement désiré dans la progression des forces de mortalité.
- La stratégie est d'évaluer un nombre d'ajustements *candidats* et d'utiliser un critère pour sélectionner parmi les ajustements celui qui aura le score le plus faible,

$$AIC = \sum_{i=1}^{n} D(y_i, (\theta(\widehat{\mu}_i))) + 2\,\upsilon,$$

où υ sont les degrés de liberté.

Méthode de vraisemblance locale

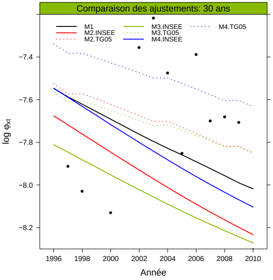
Les modèles proposés

Modèle	Formule	Réf.	Estimation	
			Vrais. loc.	Min. dist.
M1	$D^{\circ}_{x,t} \sim \mathcal{P}\left(E^{\circ}_{x,t}\ \exp\big(f_1(x,t)\big)\right)$		M1	
M2	$D^{\circ}_{x,t} \sim \mathcal{P}\left(E^{\circ}_{x,t}\ \exp\Big(f_2\big(\log\big(\varphi^{\text{ref}}_x(t)\big)\big)\Big)\right)$	INSEE	M2.INSEE	
		TG05	M2.TG05	
M3	$D^{\circ}_{x,t} \sim \mathcal{P}\left(E^{\circ}_{x,t}\ \varphi^{\text{ref}}_x(t)\ \exp\big(f_1(x,t)\big)\right)$	INSEE	M3.INSEE	
		TG05	M3.TG05	
M4	$\text{logit}\ \varphi_x(t) = \alpha + \beta\ \text{logit}\ \varphi^{\text{ref}}_x(t) + \epsilon_{x,t}$	INSEE		M4.INSEE
		TG05		M4.TG05

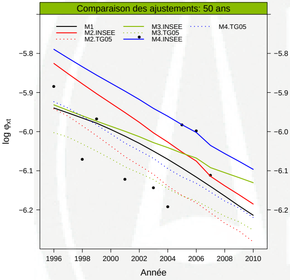
TABLE: Description des modèles de lissage.

Méthode de vraisemblance locale

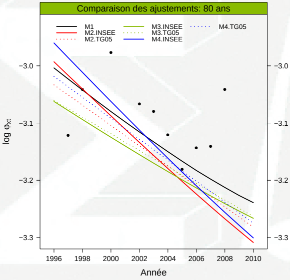
Comparaison graphique des ajustements



(a) 30 ans



(b) 50 ans



(c) 80 ans

FIGURE: Comparaison graphique des ajustements (log), population masculine

Tests et quantités pour comparer les ajustements

		M1	M2.INSEE	M2.TG05	M3.INSEE	M3.TG05	M4.INSEE	M4.TG05
χ^2		1099.24	1202.18	1252.04	1241.99	1310.53	1393.64	1416.34
R^2		0.9465	0.9423	0.9476	0.9449	0.9484	0.9101	0.9466
MAPE (%)		22.28	22.78	22.24	22.03	21.86	25.79	24.73
Déviance		1107.23	1193.44	1244.03	1204.31	1282.41	1447.64	1441.37
Résidus standardisés	> 2 > 3	56 6	61 10	67 11	64 12	75 15	73 16	84 17
Test du SMR	SMR ξ^{SMR} p-valeur	0.9821 3.8878 $1e-04$	0.9963 0.7921 0.2141	0.9954 0.9942 0.1601	0.9985 .0.3098 0.3784	1.0029 0.6175 0.2684	0.9774 4.9155 0	0.9939 1.3128 0.0946
Test de Wilcoxon	W ξ^{W} p-valeur	229508 2.4977 0.0125	222115 1.5731 0.1157	216420 0.8609 0.3893	212709 .0.3969 0.6915	215311 0.7223 0.4701	239969 3.8059 $1e-04$	220406 1.3594 0.1740
Test des signes	$+(-)$ ξ^{SIG} p-valeur	430(485) 1.7852 0.0742	436(479) 1.3885 0.1650	436(479) 1.3885 0.1650	453(462) 0.2645 0.7914	456(459) 0.0661 0.9473	395(520) 4.0993 0	409(506) 3.1737 0.0015
Test des runs	Nbre de runs ξ^{RUN} p-valeur	442 0.9858 0.3243	440 1.1596 0.2462	424 2.2204 0.0264	434 1.618 0.0014	418 2.6789 0.1057	396 3.6378 $3e-04$	376 5.1758 0

Contents

Identifier les composantes de la mortalité

Décomposition de la surface

- Décomposer la surface via une expansion de fonctions de base

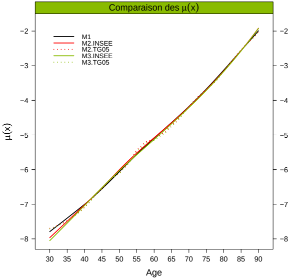
$$\log \widehat{\varphi}_x(t) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\, \phi_k(x) + \varepsilon_t(x), \quad \varepsilon_t(x) \sim \mathcal{N}\big(0, \upsilon(x)\big),$$

où $\mu(x)$ est la moyenne temporelle de $\log \widehat{\varphi}_x(t)$ et $\{\phi_k(x)\}$ est l'ensemble des fonctions de bases orthogonales.

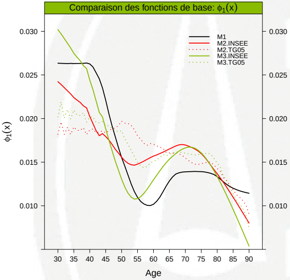
- L'idée principale est d'identifier les composantes de la mortalité et leur importance dans le temps.
- La décomposition utilisant une base orthogonale est obtenue via l'analyse en composante principale.
- Le nombre K de paramètres est déterminé par leur pouvoir explicatif.

Identifier les composantes de la mortalité

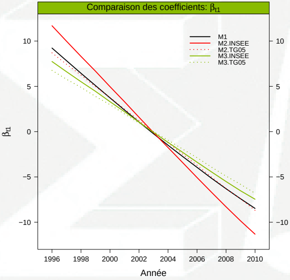
Comparaison graphique $\widehat{\mu}(x)$, $\widehat{\phi}_1(x)$ et $\widehat{\beta}_{t,1}$



(a) $\widehat{\mu}(x)$



(b) $\widehat{\phi}_1(x)$



(c) $\widehat{\beta}_{t,1}$

FIGURE: Comparaison graphique des composantes de la mortalité, population masculine

Contents

Extrapoler les paramètres variant dans le temps

- Ajuster des modèles de série temporelle à chaque $\widehat{\beta}_{t,1}$.
- On considère la gamme complète ARIMA(p,d,q) avec $d = 0, 1, 2$ et $p, q = 0, 1, 2, 3, 4$.
- Les paramètres p, d et q, sont sélectionnés selon le critère BIC.

Modèle & composant		Modèle pour les $\beta_{t,k}$	
M1	$k = 1$	ARIMA(1,2,0)	$\beta_t = 2\,\beta_{t-1} - \beta_{t-2} + \mu + \phi(\beta_{t-1} - 2\,\beta_{t-2} + \beta_{t-3} - \mu) + Z_t$
M2.INSEE	$k = 1$	ARIMA(0,2,0)	$\beta_t = 2\,\beta_{t-1} - \beta_{t-2} + Z_t$
M2.TG05	$k = 1$	ARIMA(0,1,0) with drift	$\beta_t = \beta_{t-1} + d + Z_t$
M3.INSEE	$k = 1$	ARIMA(0,2,0)	$\beta_t = 2\,\beta_{t-1} - \beta_{t-2} + Z_t$
M3.TG05	$k = 1$	ARIMA(0,1,0) with drift	$\beta_t = \beta_{t-1} + d + Z_t$

TABLE: Description des modèles pour les paramètres variant dans le temps.

- Extrapoler les coefficients $\left\{\beta_{t,1}\right\}$ en utilisant les modèles de séries temporelles.

Extrapoler les paramètres variant dans le temps

Comparaison graphique

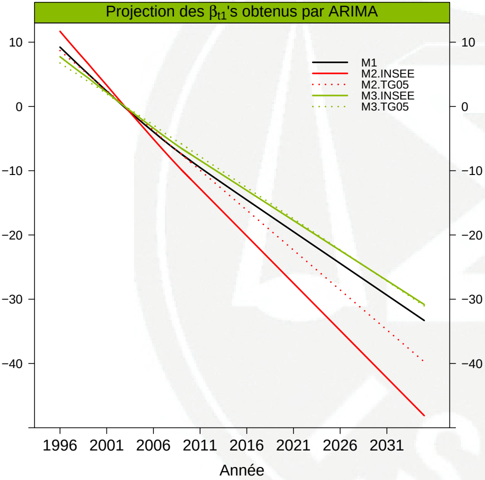
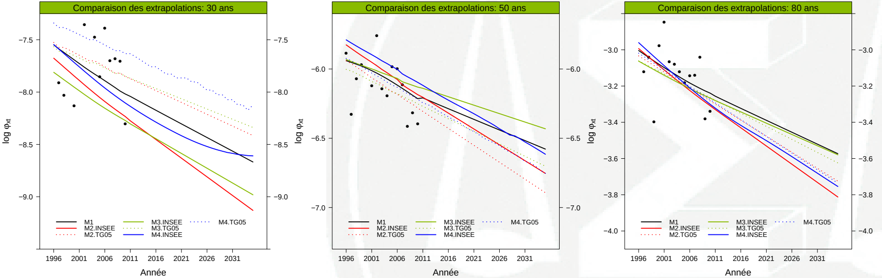


FIGURE: Projection des coefficients estimés $\beta_{t,1}$.

Contents

Projection des ajustements

Prise en compte de l'hétérogénéité

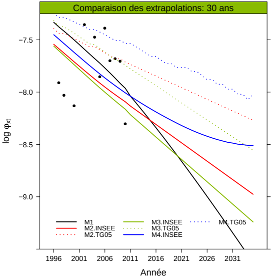


(a) 30 ans (b) 50 ans (c) 80 ans

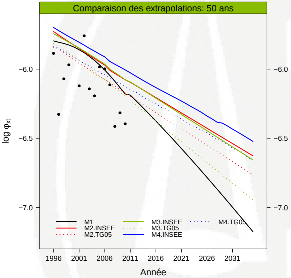
FIGURE: Comparaisons des projections pour plusieurs âges, log, population masculine

Projection des ajustements

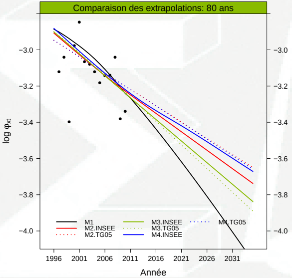
Sans prise en compte de l'hétérogénéité



(a) 30 ans



(b) 50 ans



(c) 80 ans

FIGURE: Comparaisons des projections pour plusieurs âges, log, population masculine

Remarques

- Les modèles possèdent les caractéristiques suivantes en commun :
  - Le niveau global de la mortalité diminue avec le temps
  - Ces améliorations sont plus importantes pour les jeunes adultes que pour les âges élevés.
- Néanmoins, les modèles divergent sur le niveau et la rapidité des ces améliorations.
- Nous devons donc peser les forces et faiblesses de chaque modèle pour valider la table de mortalité.
- La détermination des tendances de mortalité est fortement biaisée sans prise en compte explicite de l'hétérogénéité, notamment pour les modèles endogènes.

Ce travail a bénéficié du support de l'Institut des Actuaires et de la Chair BNP Paribas Cardif "Management de la modélisation". Les points de vue exprimés dans ce document sont ceux de l'auteur et ne reflètent pas nécessairement ceux de l'Institut des Actuaires et de BNP Paribas Cardif.

Références

Brass, W. (1971). On the scale of mortality. In W. Brass (ed.), editor, *Biological aspects of demography*. London : Taylor & Francis.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer-Verlag New-York, Inc., second edition.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, **34**(2), 187–220. Disponible ici.

Delwarde, A. and Denuit, M. (2003). Importance de la période d'observation et des âges considérés dans la projection de la mortalité selon la méthode de Lee-Carter. *Belgian Actuarial Bulletin*, **3**(1), 1–21. Disponible ici.

Delwarde, A. and Denuit, M. (2005). *Construction de tables de mortalité périodiques et prospectives*. Assurance Audit Actuariat. Economica.

Delwarde, A., Kachkhdze, D., Olie, L., and Denuit, M. (2004). Modèles linéaires et additifs géneralisés, maximum de vraisemblance local et méthodes relationelles en assurance sur la vie. *Bulletin Français d'Actuariat*, **6**(12), 77–102. Disponible ici.

Tomas, J. (2011). A local likelihood approach to univariate graduation of mortality. *Bulletin Français d'Actuariat*, **11**(22), 105–153. Disponible ici.

Tomas, J. and Planchet, F. (2013). Multidimensional smoothing by adaptive local kernel-weighted log-likelihood with application to long-term care insurance. *Insurance : Mathematics & Economics*, **52**(3), 573–589. Disponible ici.