

Mémoire présenté devant l'Université de Paris Dauphine  
pour l'obtention du diplôme du Master Actuariat  
et l'admission à l'Institut des Actuaire  
le 20 novembre 2018

Par : Corentin Guilbaud

Titre : Nouveaux modèles d'analyse et de projection de la mortalité, application à la population française

Confidentialité :  Non     Oui    (Durée :  1 an     2 ans)

---

*Les signataires s'engagent à respecter la confidentialité ci-dessus*

*Membre présent du jury de l'Institut  
des Actuaire :*

*Entreprise :*

Nom : Sia Partners

**SIAPARTNERS**

Signature :

*Membre présent du Jury du Master  
Actuariat de Paris Dauphine :*

*Directeur de Mémoire en entreprise :*

Nom : M. Santiago Fiallos

Signature :

---

*Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)*

*Secrétariat :*

*Signature du responsable entreprise*

*Bibliothèque :*

*Signature du candidat*



## Résumé

---

Ce travail a pour objet la modélisation de la mortalité dans les dimensions des âges et des années. À des fins de tarification, de provisionnement et de gestion de contrats d'assurance vie, les actuaires analysent la mortalité historique afin d'inférer la mortalité future. La majorité des modèles usuels décomposent les taux de mortalité en plusieurs facteurs. Ces modèles ajustent bien la mortalité historique mais sont peu fiables à des fins de prédiction. La principale difficulté est de déterminer les structures de dépendance spatio-temporelles liant les taux de mortalité sur des échantillons relativement petit.

Des études ont montré qu'à un instant donné les taux de mortalité à tous les âges présentent une forme caractéristique. Des modèles ont donc été développés pour la capturer de façon fine. Nous nous appuyons sur ces derniers afin de développer notre propre modèle de mortalité paramétrique répondant à certaines exigences techniques à propos du calibrage. Nous montrons que le modèle utilisé ajuste de façon fine les taux de mortalité de la population française de 1950 à 2016. L'analyse de ce modèle montre que les événements démographiques sont clairement visibles dans ses coefficients. Cela ouvre de perspectives quant à la compréhension fine de la mortalité française et à sa projection.

Nous proposons enfin des modèles permettant de projeter dans le temps les coefficients de la forme paramétrique ajustée. Appliquée à la mortalité cette approche est récente. Elle a été utilisée afin de projeter la mortalité aux grands âges. Nous en proposons une généralisation appliquée à la mortalité adulte. Cependant nous devons projeter des séries temporelles multidimensionnelles, dont les composantes sont dépendantes entre elles. La littérature nous amène à considérer des modèles de type réseau de neurones pour cette tâche. Ces derniers permettent de prendre en compte les dépendances spatio-temporelles potentiellement non linéaires entre les taux de mortalité. Nous utilisons des méthodes de régularisation afin d'obtenir des projections cohérentes, ancrées sur les tendances complexes apprises des données passées.

En conclusion nous proposons une approche d'analyse et de projection de la mortalité centrée sur la donnée qui est une alternative aux modèles existants.

---

*Mots-clés : Modèle Paramétrique de Mortalité ; Série Temporelle Multidimensionnelle ; Réseau de Neurones ; Prédiction de Mortalité.*

## Abstract

---

The purpose of this work is to model mortality. For the purpose of pricing, funding and managing life insurance contracts, actuaries analyze historical mortality in order to infer future mortality. Most common models decompose mortality rates into several factors. These models fit historical mortality well but are unreliable for prediction purposes. The main difficulty is to determine the spatio-temporal dependence structures linking mortality rates on relatively small samples.

Studies have shown that in a given year, mortality rates at all ages have a characteristic shape. Models have therefore been developed to capture it in a fine way. We rely on these to develop our own parametric mortality model that meets certain technical requirements for calibration. We show that the used model adjusts finely French mortality rates from 1950 to 2016. Analysis of this model shows that demographic events are clearly visible in its coefficients. This provides perspectives for a detailed understanding of French mortality and its projection.

Finally, we propose models to project the coefficients of the adjusted parametric form over time. Applied to mortality, this approach is recent. It has been used to project mortality to the oldest age groups. We propose a generalization applied to adult mortality. However, we must project multidimensional time series, whose components are dependent on each other. Literature leads us to consider neural network models for this task. They make possible to take into account the potentially non-linear spatial and temporal dependencies between mortality rates. We use regularization methods to obtain consistent projections based on complex trends learned from past data.

In conclusion, we propose a data-centred approach to mortality analysis and projection that is an alternative to existing models.

---

*Keywords : Parametric Mortality Model; High Dimensional Time Series; Neural Network; Mortality Prediction.*

# Note de Synthèse

Ce travail a pour objet la modélisation de la mortalité dans les dimensions des âges et des années. À des fins de tarification, de provisionnement et de gestion de contrats d'assurance vie, les actuaires analysent la mortalité historique afin d'inférer la mortalité future. Certaines activités comme la retraite nécessitent des projections de mortalité à long terme. Or depuis les années 50 l'espérance de vie à la naissance augmente de 6h par jour en moyenne. Les risques financiers liés à la longévité encourus par les assureurs peuvent donc potentiellement être très importants. Les modèles existants ne permettent pas d'obtenir de projection fiables à long terme. Nous nous attachons donc dans ce travail à comprendre et projeter la mortalité de la population française de façon cohérente.

Pour cela nous partons des modèles existants, notamment du modèle Logit+ de Debonneuil (2014) qui caractérise la mortalité à tous les âges une année donnée. À partir de celui-ci, nous créons notre propre modèle paramétrique répondant à certaines exigences de calibrage et d'interprétation des coefficients. L'analyse de ce modèle montre que ses coefficients sont particulièrement adaptés pour décrire les événements démographiques de la population française et notamment pour caractériser les effets cohortes. Enfin nous projetons les coefficients obtenus avec un modèle RVFL qui permet de prendre en compte la structure de dépendance spatio-temporelle de ces derniers.

## I - Développement des modèles paramétriques de mortalité

Le modèle de mortalité le plus rencontré est celui de Lee & Carter (1992). Ces derniers ont proposé un modèle qui décompose la surface de mortalité en une tendance latente  $\kappa_t$  et deux paramètres sensibles à l'âge  $\alpha_x$  et  $\beta_x$  :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}.$$

Il présente plusieurs limites. Les résidus n'ont pas une variance constante (Lee & Miller (2001) et Gao & Hu (2009)). L'effet cohorte n'est pas pris en compte dans le modèle et apparaît donc dans les résidus (Willets 2004). Enfin les taux d'amélioration sont projetés de façon constante, ce qui est peu vraisemblable au vu des taux d'amélioration historiques.

Pour palier à ces problèmes d'autres modèles ont été développés comme les modèles Âge-Période-Cohorte et Lee-Carter-Cohort (Renshaw & Haberman 2006). Ces derniers ajustent mieux la mortalité passée que Lee & Carter (1992). En revanche ces modèles sur-apprennent les données d'apprentissage et leurs projections ne sont pas fiables.

Face à ces insuffisances une autre méthodologie a émergée. Bongaarts (2004) propose de ne caractériser la mortalité que dans l'espace des âges, pour une année donnée, puis de projeter un ou plusieurs paramètres de ce modèle par une méthode de série temporelle dans l'espace du temps. Il illustre sa démarche avec le modèle de Thatcher et al. (1998) en projetant deux de ses paramètres avec un modèle ARIMA. Mais le modèle de Thatcher et al. (1998) ne rend compte que de la mortalité aux grands âges. Devant l'incapacité de ce modèle à généraliser l'ensemble des formes de mortalité rencontrées, Debonneuil (2014) a créé et étudié un nouveau modèle paramétrique de mortalité plus souple appelé modèle *logit+*. Celui-ci est intéressant car il prend en compte la sur-mortalité précoce observée à plusieurs reprises autour de 60 ans dans la deuxième partie du vingtième siècle (composante *Mid*). Il s'exprime pour une année donnée pour tous les âges  $x$  par :

$$q(x) = Old(x)(1 + Mid(x) + Young(x)) + c_{young}.$$

Avec les composantes suivantes :

$$\begin{aligned} Young(x) &= e^{a_{young} - b_{young}x}, \\ Mid(x) &= a_{mid} e^{-\frac{(x - c_{mid})^2}{b_{mid}}}, \\ Old(x) &= q_{asympt} \frac{1}{1 + e^{-a_{old}x - b_{old}}}. \end{aligned}$$

Ce modèle étant multiplicatif, les trois composantes sont corrélées de manière significative entre elles. Son calibrage est donc en pratique délicat. Pour la même raison ses paramètres ne sont pas intuitifs à interpréter. C'est pourquoi nous avons choisi de retenir notre propre modèle présenté dans la partie suivante.

## II - Définition et calibrage d'un modèle de mortalité générique

Le modèle retenu est additif. Nous l'appelons Logit+ additif. Il est définie par :

$$\mu(x) = Old(x) + Mid(x) + Young(x).$$

Avec les composantes :

$$\begin{aligned} Young(x) &= e^{a_{young} + b_{young}x}, \\ Mid(x) &= \beta e^{-\frac{(x - m)^2}{\sigma}}, \\ Old(x) &= \mu_{asympt} \frac{1}{1 + e^{-ax - b}}. \end{aligned}$$

Un exemple d'ajustement est visible sur la figure 1. Il faut alors calibrer ce modèle chaque année sur les données historiques disponibles. Nous avons retenu un calibrage par maximum de vraisemblance. Notons  $E_x$  l'exposition à l'âge  $x$  et  $D_x$  le nombre de morts dans les 12 derniers mois à l'âge  $x$ .  $D_x$  peut donc être vu comme la somme de la réalisation de  $E_x$  variables aléatoires suivant une loi binomiale de paramètre  $q_x$ . On considérera les décès indépendants. La vraisemblance de l'échantillon est donc :

$$L(q_x) = \prod_{x=x_{min}}^{x_{max}} \binom{E_x}{D_x} q_x^{D_x} (1 - q_x)^{E_x - D_x}$$

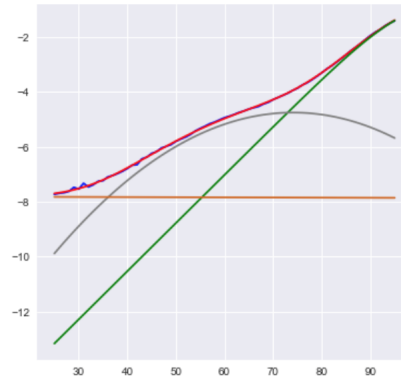


FIGURE 1 – Ajustement de la mortalité observée (en bleu) en fonction de l'âge par le modèle *Logit+additif* (en rouge) sur échelle logarithmique. Les trois composantes sont *Old* en vert, *Mid* en gris et *Young* en chocolat. Population française en 2015.

Le problème revient à trouver les huit coefficients qui déterminent  $q_x$  pour tous les âges observés tel que  $L(q_x)$  soit maximisé. Il est équivalent de maximiser la log-vraisemblance :

$$\begin{aligned} l(q_x) &= \sum_{x=x_{min}}^{x_{max}} \left[ \ln\left(\frac{E_x}{D_x}\right) + D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x) \right] \\ &= \sum_{x=x_{min}}^{x_{max}} [D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x)] + \text{constante} \end{aligned}$$

En pratique nous maximiserons  $\sum_{x=x_{min}}^{x_{max}} [D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x)]$ , avec  $q_x$  donné par le modèle retenu. Pour ce faire nous utilisons l'algorithme Adam (Kingma & Ba 2014), dérivé d'une méthode de gradient. Il présente l'avantage d'avoir un pas adaptatif pour chaque composante du gradient, condition indispensable dans notre problème sachant que les paramètres à ajuster n'ont pas la même échelle.

Pour stabiliser l'ajustement nous fixons  $b_{young}$  et  $\mu_{asympt}$ . Les coefficients obtenus sont présentés sur la figure 2.

Nous remarquons que ces coefficients reflètent l'histoire démographique française :

- Une tendance à la baisse de  $a_{young}$  signifie un abaissement générale de la courbe de mortalité dans le temps et donc une augmentation de l'espérance de vie à tous les âges.
- Une diminution de  $b$  ainsi qu'une augmentation de  $a$  reflète la rectangularisation de la fonction de survie.
- Les saisonnalités présentes dans  $m$  et  $\sigma$  reflètent le passage de certaines cohortes à la surmortalité précoce autour de 60 ans plus importante (Horiuchi 1983).

Dès lors que l'influence de chaque paramètre est comprise, il est possible pour un assureur d'adapter le modèle à son portefeuille et de ne choquer que certaines composantes. Cette approche peut être particulièrement utile pour évaluer ses propres risques de mortalité et de longévité.

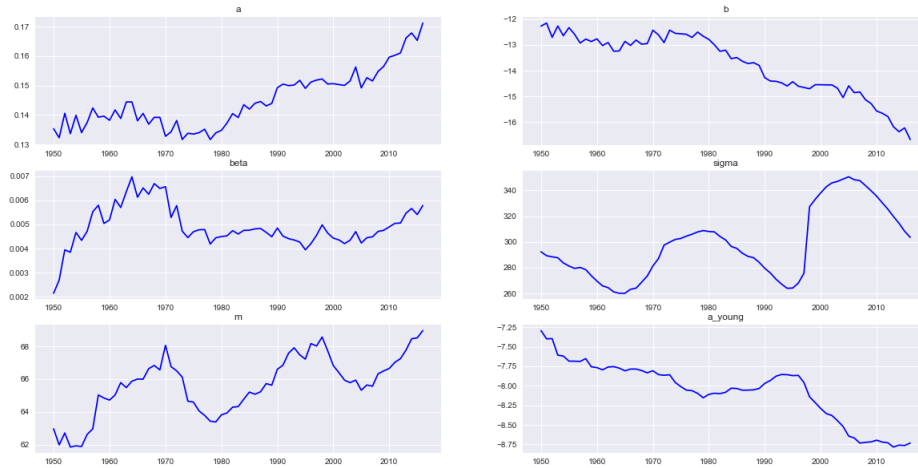


FIGURE 2 – Coefficients obtenus par maximum de log vraisemblance sur la population française entre 1955 et 2015 en fixant  $b_{young}$ .

### III - Projection temporelle de la mortalité par réseaux RVFL

Nous cherchons à projeter dans le temps les coefficients obtenus en partie précédente. L'approche décrite dans cette partie consiste à utiliser des modèles pour lesquels la prédiction d'un paramètre dépend des valeurs passées de celui-ci, ainsi que des valeurs passées des autres paramètres. Le problème se formule de la façon suivante :

- Soit  $p$  le nombre de séries à extrapoler, dans notre cas  $p = 6$ .
- $\forall j \in \llbracket 1, p \rrbracket$ ,  $(X_t^{(j)})_{t \geq 0}$  sont les  $p$  séries temporelles observées à  $n$  dates discrètes.
- Nous voulons prédire ces séries jusqu'au temps  $n + h$ ,  $h \in \mathbb{N}^*$ . Pour simplifier les modèles, nous effectuerons  $h$  prédictions successives de pas de temps de taille 1.
- Soit  $k$  l'ordre du modèle considéré, c'est à dire le nombre de périodes d'observation utilisées pour faire une prédiction.

Le problème de régression que nous voulons résoudre est alors le suivant, nous cherchons une fonction  $f$  tel qu'au temps  $i$  :

$$\begin{bmatrix} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{bmatrix} = f \left( \begin{bmatrix} X_{i-k}^{(1)} & \dots & X_{i-1}^{(1)} \\ \vdots & \ddots & \vdots \\ X_{i-k}^{(p)} & \dots & X_{i-1}^{(p)} \end{bmatrix} \right)$$

Pour entraîner les modèles de projection, nous utilisons les données sur la mortalité issues de la HMD (Shkolnikov et al. 2000) entre 1955 et 2015. Les données précédents 1950 sont perturbées par les guerres. Nous n'observons donc que 61 points en dimension 6 alors que les modèles utilisés auront plusieurs dizaines de paramètres à calibrer. Le protocole utilisé pour chaque modèle est alors le suivant :

- Les séries sont centrées et réduites.



- Les modèles sont calibrés à l'aide des points obtenus strictement avant 2006.
- Les points entre 2006 et 2015 servent de test afin de choisir les hyperparamètres. Nous mettons en place une recherche par grille : nous entraînerons  $r$  fois les modèles sur l'intervalle [1955, 2005] pour  $r$  jeux d'hyperparamètres différents, nous retenons le jeu ayant l'erreur la plus faible sur l'intervalle [2006, 2015]. Soient  $(q_x^y)_{x \in \llbracket 25, 95 \rrbracket, y \in \llbracket 2006, 2015 \rrbracket}$  et  $(\hat{q}_x^y)_{x \in \llbracket 25, 95 \rrbracket, y \in \llbracket 2006, 2015 \rrbracket}$  les probabilités observées et estimées entre 25 et 95 ans des années 2006 à 2015, l'erreur minimisée est alors :

$$e = \sum_{a=25}^{95} \frac{1}{2015 - 2006 + 1} \sum_{y=2006}^{2015} (\ln(q_x^y) - \ln(\hat{q}_x^y))^2.$$

- Nous sommes obligés de prendre un échantillon de test relativement petit. En effet, la bosse sur la mortalité autour de 50 ans décrite en partie 1 apparaît au début des années 2000. Si nous l'excluons de l'ensemble d'entraînement, le modèle ne la fera pas apparaître. Dès lors si la prédiction du modèle ne paraît pas assez stable, nous nous réservons le droit d'augmenter la régularisation des modèles pour lisser les projections.
- Une fois les hyperparamètres choisis, nous entraînons le modèle sur toutes les données pour projeter la mortalité au delà de 2015.

Nous commençons par projeter les séries avec un modèle linéaire : le modèle Vecteur Autoregressif (VAR) (Lütkepohl 2005) défini à l'ordre  $k$  par :

$$\forall i \in \llbracket k, n \rrbracket, \begin{bmatrix} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{bmatrix} = \sum_{t=1}^k A_t \begin{bmatrix} X_{i-t}^{(1)} \\ \vdots \\ X_{i-t}^{(p)} \end{bmatrix} + b + e_i$$

Avec :

- $(A_t)_{t \in \llbracket 1, k \rrbracket}$ , avec  $A_t \in \mathbb{R}^{p \times p}$  les matrices des coefficients,
- $b \in \mathbb{R}^{p \times 1}$  le vecteur des biais,
- $e_i \in \mathbb{R}^{p \times 1}$  le vecteur des erreurs.

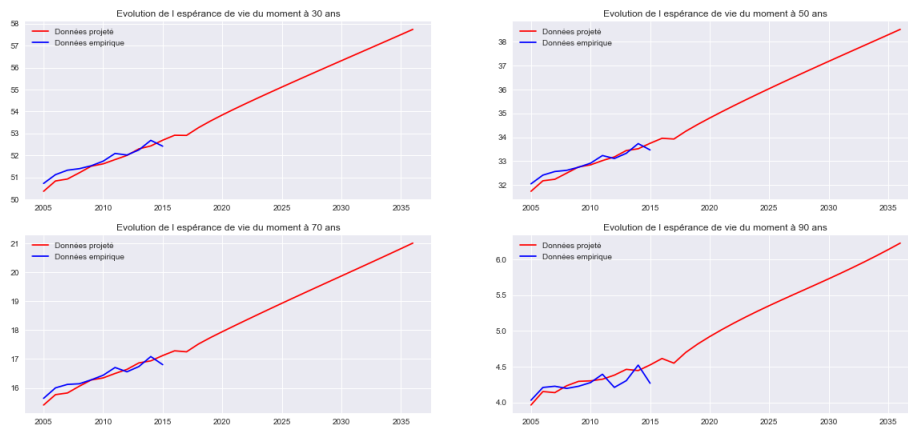


FIGURE 3 – Espérances de vie du moment empirique (en bleu) et projetée (en rouge) par VAR(1) à 30, 50, 70 et 90 ans.

Les espérances de vie du moment projetées sont visibles sur la figure 3. Le modèle anticipe une augmentation de l'espérance de vie plus rapide que par le passé principalement aux grands âges (la pente de la courbe est plus importante). Ces tendances, bien que plausibles, ne paraissent pas cohérentes avec la tendance actuelle des pays développés pour lesquels l'augmentation de l'espérance de vie tend à ralentir (Denney et al. 2013). Cela peut être un signe de sur-apprentissage. Nous ajustons en effet 61 points de dimension 6 avec un modèle à  $1 \times 6 \times 6 + 6 = 42$  paramètres sans régularisation. Il ne faut donc pas accorder trop de crédit à la projection réalisée par ce modèle.

Le modèle retenu est donc un réseau RVFL. Il est défini de la manière suivante :

- Soit  $p$  le nombre de séries à projeter,
- Soit  $k$  l'ordre du modèle, c'est à dire le nombre de pas de temps utilisés pour une prédiction,
- Soit  $L$  le nombre de neurones cachés.
- Soit  $\Phi(x)$  une fonction non linéaire de  $x$ .

On a alors :

$$\forall i \in \llbracket k, n \rrbracket, \forall j \in \llbracket 1, p \rrbracket, X_i^{(j)} = \beta_0 + \sum_{t=1}^k \sum_{v=1}^p \beta_{t,v} X_{i-t}^{(v)} + \sum_{l=1}^L \gamma_l \Phi \left( \sum_{t=1}^k \sum_{v=1}^p W_{t,v}^l X_{i-t}^{(v)} \right) + \epsilon_i^j.$$

Le schéma 4 illustre cette équation avec les valeurs  $p = 2$ ,  $k = 2$  et  $L = 3$ . Chaque cercle de la couche cachée et de la couche de sortie représente un perceptron : les sorties de chacun d'entre eux sont les sommes pondérées des entrées, passées dans une non linéarité pour la couche cachée. Nous utilisons la fonction tangente hyperbolique :  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , qui est bijective de  $\mathbb{R}$  dans  $[-1, 1]$ , comme fonction non linéaire dans la couche cachée. Ce choix est justifié par le fait que l'on centre et réduit les données.

Dans ce modèle il y a trois matrices de poids différents :

- $W = (w_{i,j})_{1 \leq i \leq k \cdot p, 1 \leq j \leq L}$  représente les poids permettant de passer de la couche d'entrée à la couche cachée.
- $\beta = (\beta_{i,j})_{1 \leq i \leq k \cdot p, 1 \leq j \leq p}$  représente les poids permettant de passer de la couche d'entrée à la couche de sortie.
- $\gamma = (\gamma_{i,j})_{1 \leq i \leq L, 1 \leq j \leq p}$  représente les poids permettant de passer de la couche cachée à la couche de sortie.

Les éléments de  $W$  sont tirés d'une séquence de Sobol déterministe quasi aléatoire sur  $[0, 1]$  qui est transformée dans  $[-1, 1]$ . L'idée est d'avoir des points plus ou moins équidistants dans  $[-1, 1]^{k \cdot p}$ . Ces poids sont tirés à la création du modèle. Il ne sont ensuite pas modifiés pour effectuer des prédictions. Le calibrage du modèle se fait alors comme une régression linéaire exprimant les sorties en fonction des entrées et d'une non linéarité de ces dernières.

En pratique le calibrage de tel modèle n'est pas évident. Le faible nombre de données nous oblige à trouver des astuces tel que la différentiation des séries, la réduction de dimension et la régularisation des modèles. On obtient alors les taux d'amélioration projetés visibles sur la figure 5. Ces derniers sont en continuité avec les taux d'amélioration passés. On observe notamment une projection des effets cohorte cohérentes.

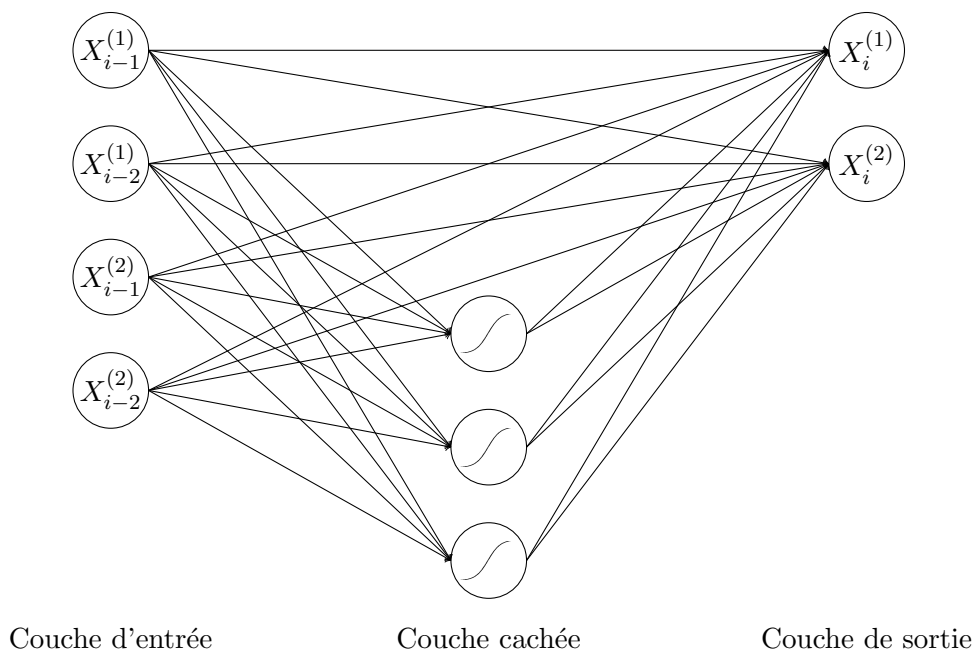


FIGURE 4 – Schéma d'illustration d'un réseau RVFL avec  $p = 2$ ,  $k = 2$  et  $L = 3$ .

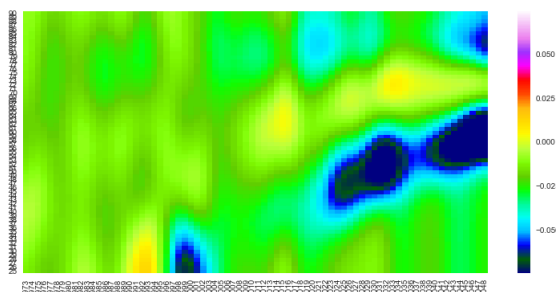


FIGURE 5 – Taux d'amélioration lissés calibrés (avant 2015) et projetés (à partir de 2015) par RVFL(1) sur les séries différenciées.

## Conclusion

Bien que le démographe puisse se servir de ces travaux afin de projeter la mortalité à l'échelle d'un pays, l'actuaire s'intéressera à des problématiques légèrement différentes :

- Il souhaitera positionner la mortalité de son portefeuille par apport à une mortalité de référence. Pour cela de futurs travaux peuvent reprendre l'étude menée en partie 2 sur des données plus fines, par exemple par sexe, pour se rapprocher de la mortalité assurée.
- L'actuaire pourra aussi à l'aide du modèles Logit+ additif construire des scénarios démographiques adaptés à son portefeuille. Des études plus poussées sur les interactions entre les coefficients du modèle seront alors nécessaires. Il serait pour cela intéressant d'étudier les calibrages des coefficients obtenus pour différents pays afin d'avoir plus d'exemples d'interaction.

De futurs travaux peuvent être menés afin de fiabiliser les prédictions obtenues en partie 3. Il est possible d'ajouter des variables explicatives aux modèles décrits dans cette partie. Enfin le modèle Logit + additif permet de faire une étude spécifique sur l'apparition et l'anticipation des cohortes aux mortalités particulières.



# Synthesis note

The purpose of this work is to model mortality. For the purpose of pricing, funding and managing life insurance contracts, actuaries analyze historical mortality in order to infer future mortality. Some businesses such as pensions require long-term mortality projections. However, since the 1950s, life expectancy at birth has increased by an average of 6 hours per day. Financial risks related to longevity incurred by insurers can therefore be potentially very significant. Existing models do not provide reliable long-term projections. In this work we try to understand and to project French population mortality in a coherent manner.

To do this, we rely on existing models, in particular Logit+ model from Debonneuil (2014) which characterizes mortality rate at all ages for a given year. From this we create our own parametric model that meets certain calibration and coefficient interpretation requirements. Analysis of this model shows that its coefficients are particularly well suited to describe the demographic events of the French population and in particular to characterize cohort effects. Finally, we project the coefficients obtained with an RVFL model to take into account the space-time dependence structure of the coefficients.

## I - Development of parametric mortality models

The most common mortality model is Lee & Carter (1992) model. It decomposes the mortality surface into a latent trend  $\kappa_t$  and two age-sensitive parameters  $\alpha_x$  and  $\beta_x$  :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}.$$

It has several limitations. Residuals do not have a constant variance (Lee & Miller (2001) and Gao & Hu (2009)). Cohort effect is not taken into account in the model and therefore appears in residuals (Willets 2004). Finally, improvement rates are projected consistently, which is unlikely given historical improvement rates.

To address these problems other models have been developed such as the Age-Period-Cohort and Lee-Carter-Cohort models (Renshaw & Haberman 2006). They adjust past mortality better than Lee & Carter (1992). On the other hand these models overfit training data and their projections are not reliable.

To address these weaknesses another methodology has emerged. Bongaarts (2004) proposes to characterize mortality only in the age space, for a given year, and then to project one or more parameters of this model by a time series method in the time space. He illustrates his approach with

the Thatcher et al. (1998) model by projecting two of his parameters with an ARIMA model. But the Thatcher et al. (1998) model only captures mortality in the old age group. Given the inability of this model to generalize all forms of mortality encountered, Debonneuil (2014) created and studied a new, more flexible parametric mortality model called the *logit+* model. It is interesting because it takes into account the early excess mortality observed several times around the age of 60 in the second half of the twentieth century (*Mid* component). It is expressed for a given year for all ages  $x$  per:

$$q(x) = Old(x)(1 + Mid(x) + Young(x)) + c_{young}.$$

With the following components:

$$\begin{aligned} Young(x) &= e^{a_{young} - b_{young}x}, \\ Mid(x) &= a_{mid} e^{-\frac{(x - c_{mid})^2}{b_{mid}}}, \\ Old(x) &= q_{asympt} \frac{1}{1 + e^{-a_{old}x - b_{old}}}. \end{aligned}$$

As this model is multiplicative, the three components are significantly correlated with each other. Its calibration is therefore in practice delicate. For the same reason its parameters are not intuitive to interpret. That is why we have chosen to use our own model presented in the next section.

## II - Definition and calibration of a generic mortality model

The model used is additive. We call it Logit+ additive. It is defined by:

$$\mu(x) = Old(x) + Mid(x) + Young(x).$$

With the following components:

$$\begin{aligned} Young(x) &= e^{a_{young} + b_{young}x}, \\ Mid(x) &= \beta e^{-\frac{(x - m)^2}{\sigma}}, \\ Old(x) &= \mu_{asympt} \frac{1}{1 + e^{-ax - b}}. \end{aligned}$$

An example of a fit is shown in figure 6. This model must then be calibrated each year on the basis of available historical data. We have chosen a maximum likelihood calibration. Let  $E_x$  be the exposure at age  $x$  and  $D_x$  the number of deaths in the last 12 months at age  $x$ .  $D_x$  can therefore be seen as the sum of  $E_x$  realization of a binomial random variables with parameter  $q_x$ . Independent deaths will be considered. The likelihood of the sample is therefore:

$$L(q_x) = \prod_{x=x_{min}}^{x_{max}} \binom{E_x}{D_x} q_x^{D_x} (1 - q_x)^{E_x - D_x}$$

The problem is to find the eight coefficients that determine  $q_x$  for all observed ages such that  $L(q_x)$  is maximized. It is equivalent to maximize the log-likelihood:

$$\begin{aligned} l(q_x) &= \sum_{x=x_{min}}^{x_{max}} \left[ \ln\left(\binom{E_x}{D_x}\right) + D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x) \right] \\ &= \sum_{x=x_{min}}^{x_{max}} [D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x)] + \text{constante} \end{aligned}$$

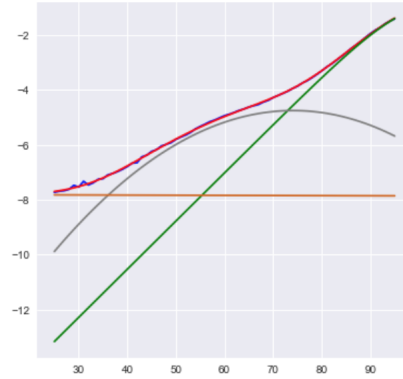


Figure 6 – Adjustment of observed mortality (in blue) as a function of age by the model *Logit+additive* (in red) on a logarithmic scale. The three components are *Old* in green, *Mid* in grey and *Young* in chocolate. 2015 French population.

In practice we maximize  $\sum_{x=x_{min}}^{x_{max}} [D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x)]$ , with  $q_x$  given by the selected model. To do this we use the Adam (Kingma & Ba 2014) algorithm, derived from a gradient method. It has the advantage of having an adaptive step for each component of the gradient, an essential condition in our problem knowing that the adjusted parameters do not have the same scale.

To stabilize the fit we fix  $b_{young}$  and  $\mu_{asympt}$ . The resulting coefficients are presented in figure 7.

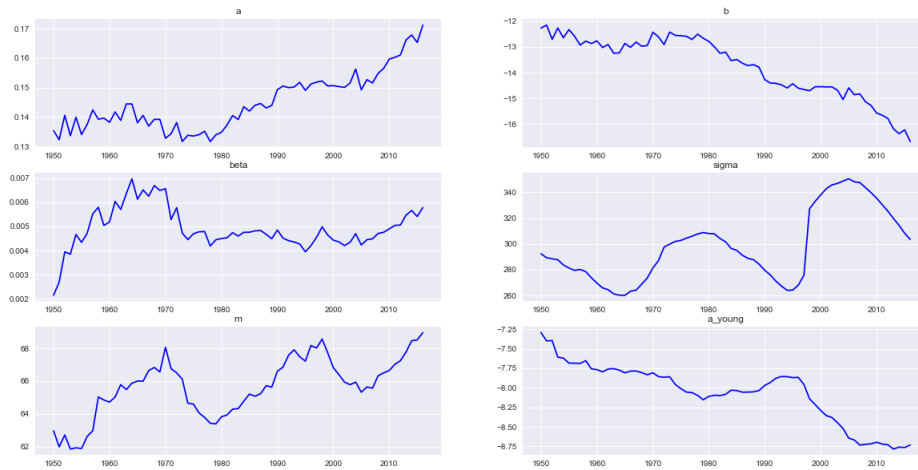


Figure 7 – Coefficients determined by maximum log likelihood on the French population between 1955 and 2015 by fixing  $b_{young}$ .

We note that these coefficients reflect French demographic history:

- A decreasing trend of  $a_{young}$  means a general lowering of the mortality curve over time and therefore an increase in life expectancy at all ages.

- A decrease of  $b$  as well as an increase of  $a$  reflects the rectangularization of the survival function.
- The seasonalities in  $m$  and  $\sigma$  reflect the transition around 60 from some cohorts with early over-mortality more significant (Horiuchi 1983).

Once each parameter influence is understood, it is possible for an insurer to adapt the model to its portfolio and shock only certain components. This approach can be particularly useful for assessing our own mortality and longevity risks.

### III - Time projection of mortality with RVFL networks

We aim to project in time the coefficients obtained in the previous part. Described approach consists in using models for which the prediction of a parameter depends on the past values of all the parameters. The problem is formulated as follows:

- Let  $p$  be the number of series to be extrapolated, here  $p = 6$ .
- $\forall j \in \llbracket 1, p \rrbracket$ ,  $(X_t^{(j)})_{t \geq 0}$  are the  $p$  observed time series at  $n$  discrete dates.
- We want to predict these series until time  $n + h$ ,  $h \in \mathbb{N}^*$ . To simplify the models, we will make  $h$  successive predictions of 1 time steps.
- Let  $k$  be the order of the considered model, i.e. the number of observation periods used to make a prediction.

The regression problem we want to solve is then the following, we are looking for a function  $f$  such that at time  $i$  :

$$\begin{bmatrix} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{bmatrix} = f \left( \begin{bmatrix} X_{i-k}^{(1)} & \cdots & X_{i-1}^{(1)} \\ \vdots & \ddots & \vdots \\ X_{i-k}^{(p)} & \cdots & X_{i-1}^{(p)} \end{bmatrix} \right)$$

To train projection models we use mortality data from the HMD (Shkolnikov et al. 2000) between 1955 and 2015. Previous 1950 data are disrupted by wars. We therefore only observe 61 points in dimension 6 while the models used will have several dozen parameters to calibrate. The protocol used for each model is then the following:

- The series are normalized.
- Models are calibrated using points obtained strictly before 2006.
- The points between 2006 and 2015 are used as test set to choose the hyperparameters. We set up a grid search method: we train  $r$  times the models on the interval  $[1955, 2005]$  for  $r$  sets of different hyperparameters. We retain the set with the lowest error on the interval  $[2006, 2015]$ . Let  $(q_x^y)_{x \in \llbracket 25, 95 \rrbracket, y \in \llbracket 2006, 2015 \rrbracket}$  and  $(\hat{q}_x^y)_{x \in \llbracket 25, 95 \rrbracket, y \in \llbracket 2006, 2015 \rrbracket}$  be the observed and estimated probabilities between 25 and 95 from 2006 to 2015, minimized error is then:

$$e = \sum_{a=25}^{95} \frac{1}{2015 - 2006 + 1} \sum_{y=2006}^{2015} (\ln(q_x^y) - \ln(\hat{q}_x^y))^2.$$

- We have to take a relatively small test sample. Indeed, the bump in mortality around age 60 described in part 1 appeared in the early 2000s. If we exclude it from the training set, the model will not show it. Therefore if the model prediction does not appear stable enough, we reserve the right to increase the regularization of the models to smooth the projections.



- Once the hyperparameters are chosen, we train the model on all the data to project mortality beyond 2015.

We start by projecting the series with a linear model: the Vector Autoregressive Model (VAR) (Lütkepohl 2005) defined at order  $k$  by:

$$\forall i \in \llbracket k, n \rrbracket, \begin{bmatrix} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{bmatrix} = \sum_{t=1}^k A_t \begin{bmatrix} X_{i-t}^{(1)} \\ \vdots \\ X_{i-t}^{(p)} \end{bmatrix} + b + e_i$$

Avec :

- $(A_t)_{t \in \llbracket 1, k \rrbracket}$ , avec  $A_t \in \mathbb{R}^{p \times p}$  coefficients matrices,
- $b \in \mathbb{R}^{p \times 1}$  bias vector,
- $e_i \in \mathbb{R}^{p \times 1}$  error vector.

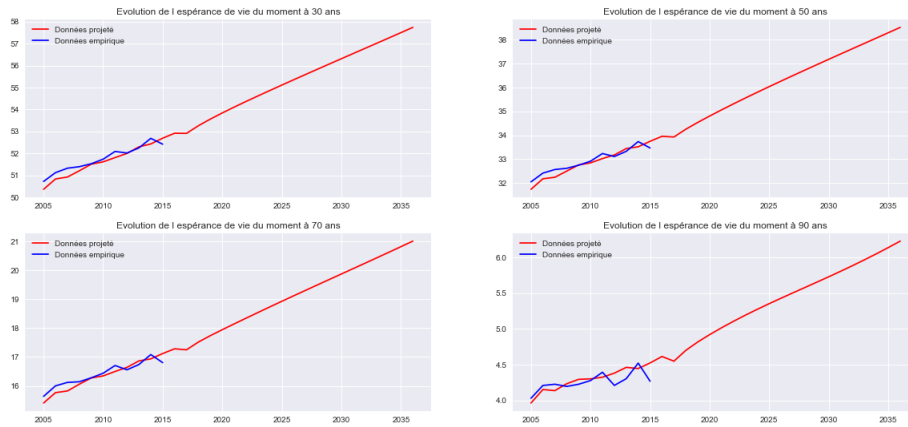


Figure 8 – Empirical (blue) and projected (red) current life expectancies by VAR(1) at 30, 50, 70 and 90.

The projected current life expectancies are visible in the figure 8. The model anticipates a faster increase in life expectancy than in the past, mainly at older ages (the slope of the curve is steeper). These trends, while plausible, do not appear to be consistent with the current trend in developed countries where increasing life expectancy tends to slow down (Denney et al. 2013). This can be a sign of overfitting. We adjust 61 points of dimension 6 with a model of  $1 \times 6 \times 6 \times 6 + 6 = 42$  parameters without regularization. Therefore, it is important not to give too much credit to the projection made by this model.

The model chosen is therefore an RVFL network. It is defined as follows:

- Let  $p$  be the number of times series,
- Let  $k$  be the model order,
- Let  $L$  be the number of hidden neurons,
- Let  $\Phi(x)$  be a non linear function of  $x$ .

Then we have:

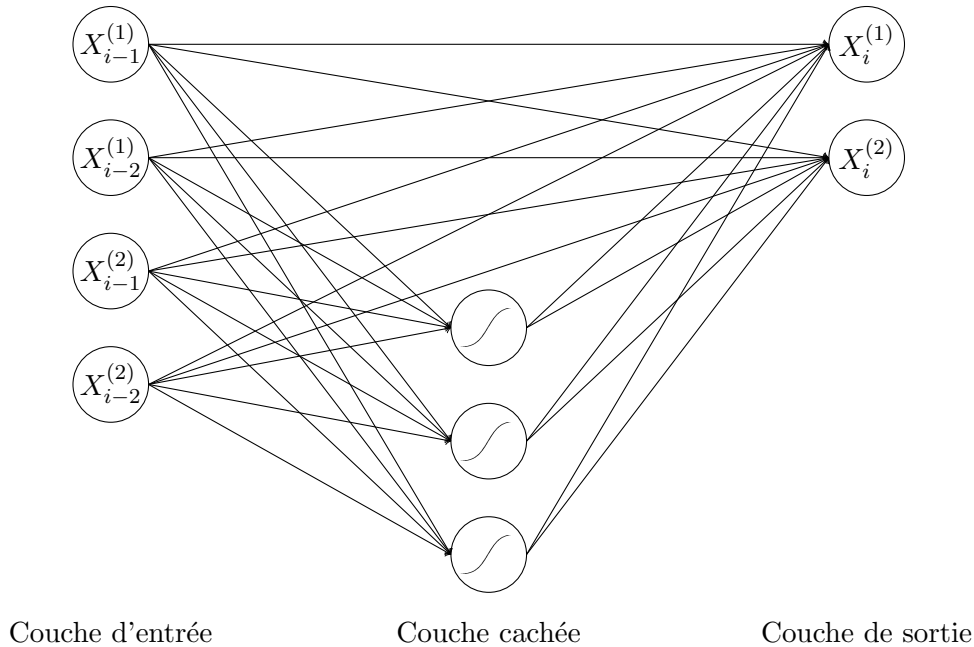


Figure 9 – Illustration diagram of an RVFL network with  $p = 2$ ,  $k = 2$  and  $L = 3$ .

$$\forall i \in \llbracket k, n \rrbracket, \forall j \in \llbracket 1, p \rrbracket, X_i^{(j)} = \beta_0 + \sum_{t=1}^k \sum_{v=1}^p \beta_{t,v} X_{i-t}^{(v)} + \sum_{l=1}^L \gamma_l \Phi \left( \sum_{t=1}^k \sum_{v=1}^p W_{t,v}^l X_{i-t}^{(v)} \right) + \epsilon_i^j.$$

The diagram 9 illustrates this equation with the values  $p = 2$ ,  $k = 2$  and  $L = 3$ . Each circle of the hidden layer and the output layer represents a perceptron: the outputs of each of them are the weighted sums of the inputs, passed in a non-linearity for the hidden layer. We use the hyperbolic tangent function:  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , which is bijective from  $\mathbb{R}$  to  $[-1, 1]$ , as a non-linear function in the hidden layer. This choice is justified by the fact that the data are normalised.

In this model there are three different weights matrices:

- $W = (w_{i,j})_{1 \leq i \leq k \cdot p, 1 \leq j \leq L}$  represents the weights used to pass from the input layer to the hidden layer.
- $\beta = (\beta_{i,j})_{1 \leq i \leq k \cdot p, 1 \leq j \leq p}$  represents the weights used to pass from the input layer to the output layer.
- $\gamma = (\gamma_{i,j})_{1 \leq i \leq L, 1 \leq j \leq p}$  represents the weights used to pass from the hidden layer to the output layer.

The  $W$  elements are extracted from a quasi random deterministic Sobol sequence on  $[0, 1]$  which is transformed into  $[-1, 1]$ . The idea is to have more or less equidistant points in  $[-1, 1]^{k \cdot p}$ . These weights are extracted when the model is created. They are then not modified to make predictions. The calibration of the model is then done as a linear regression expressing the outputs as a function of the inputs and a non-linearity of them.

In practice it is not easy to calibrate such model. The small amount of data requires us to find tricks such as series differentiation, size reduction and model regularization. The projected

improvement rates shown in Figure 10 are then obtained. These are in line with past improvement rates. In particular, a projection of coherent cohort effects is observed.

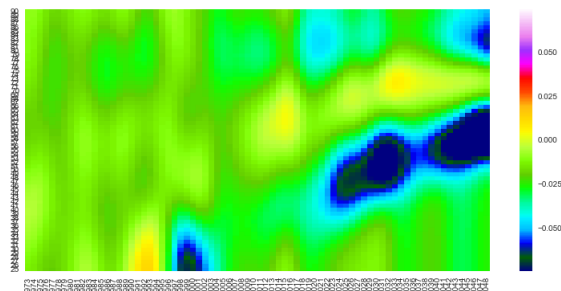


Figure 10 – Smoothed improvement rates calibrated (before 2015) and projected (from 2015) by RVFL(1) on differentiated series.

## Conclusion

Although the demographer can use this work to project mortality at a country level, the actuary will focus on slightly different issues:

- He wishes to position the mortality of his portfolio in relation to a benchmark mortality. For this reason, future work may repeat the study conducted in part 2 on more detailed data, for example by sex, to get closer to insured mortality.
- An actuary will also be able to use the additive Logit+ model to construct demographic scenarios adapted to his portfolio. Further studies on the interactions between the model's coefficients will then be required. It would therefore be interesting to study the calibrations of the coefficients obtained for different countries in order to have more examples of interaction.

Future work can be carried out to make the predictions obtained in part 3 more reliable. It is possible to add explanatory variables to the models described in this section. Finally, the additive Logit + model allows specific study to be carried out on the appearance and anticipation of cohorts with particular mortality rates.



# Remerciements

Je tiens à exprimer toute ma reconnaissance à Michael Donio qui m'a mis dans les meilleurs dispositions pour la réalisation de ce mémoire. Je souhaite aussi témoigner toute ma gratitude envers Santiago Fiallos, Frédéric Planchet et Julien Tomas qui m'ont guidé et éclairé de leurs précieux conseils et remarques tout au long de ce travail.

Je remercie toute l'équipe pédagogique de l'INSA Toulouse et tout particulièrement du département GMM pour leur pédagogie, leur patience et leur accompagnement tout au long de mon parcours. Ils m'ont permis de m'épanouir dans mes études et d'acquérir de nombreuses compétences techniques, mais avant tout de la rigueur, un esprit d'analyse critique ainsi que des qualités humaines. Je remercie également l'équipe enseignante de l'université Paris-Dauphine.

Je témoigne également toute ma reconnaissance à Cécile Stas, Philippe Rossignol et Rémi Chapelon qui de près ou de loin m'ont aidé à construire mon parcours professionnel.

Je remercie mes chers parents et mes soeurs, pour leur soutien constant et leurs encouragements.

Enfin un grand merci à Cédrik, Murielle, Jérémy, Gauthier, Damien, Nicolas et Andréa pour leur bonne humeur, leurs petits déjeuners et leur bienveillance.



# Table des matières

<b>Note de Synthèse</b>	<b>5</b>
<b>Synthesis note</b>	<b>13</b>
<b>Remerciements</b>	<b>21</b>
<b>Table des matières</b>	<b>23</b>
<b>Introduction</b>	<b>27</b>
<b>1 Développement des modèles paramétriques de mortalité</b>	<b>31</b>
1.1 Contexte de l'étude et problématiques liées à la mortalité . . . . .	31
1.2 Définitions, notations et présentation des données . . . . .	34
1.2.1 Définitions et notations utiles dans l'étude de la mortalité . . . . .	34
1.2.2 Présentation des données de la <i>Human Mortality Database</i> . . . . .	36
1.3 Présentation des modèles de référence et de leurs limites . . . . .	37
1.3.1 Le modèle de <i>Lee-Carter</i> . . . . .	37
1.3.2 Le modèle <i>Âge-Période-Cohorte</i> . . . . .	38
1.3.3 Le modèle <i>Lee-Carter-Cohorte</i> . . . . .	40
1.4 Présentation des modèles avancés existants . . . . .	42
1.4.1 Le modèle de Thatcher . . . . .	42

1.4.2	Le modèle <i>Logit+</i> . . . . .	43
1.5	Conclusion . . . . .	46
<b>2</b>	<b>Définition et calibrage d'un modèle de mortalité générique</b>	<b>47</b>
2.1	Modèle <i>Logit+</i> additif . . . . .	47
2.2	Méthodes d'ajustement du modèle <i>Logit+</i> additif . . . . .	49
2.2.1	Calibrage par maximum de vraisemblance . . . . .	49
2.2.2	Calibrage par minimisation d'une erreur d'ajustement . . . . .	49
2.2.3	Optimisateur utilisé . . . . .	50
2.2.4	Méthodologie générale employée . . . . .	53
2.3	Présentation de l'ajustement de la mortalité . . . . .	55
2.3.1	Ajustement initial . . . . .	55
2.3.2	Ajustement retenu . . . . .	56
2.4	Analyse des coefficients ajustés . . . . .	58
2.4.1	Une mortalité générale s'abaissant avec le temps . . . . .	58
2.4.2	Rectangularisation de la fonction de survie . . . . .	58
2.4.3	Influence de cohortes particulières . . . . .	59
2.5	Projection de la mortalité par composante . . . . .	61
2.5.1	Projection de la composante <i>Old</i> . . . . .	61
2.5.2	Projection de la composante <i>Young</i> . . . . .	61
2.5.3	Projection de la composante <i>Mid</i> . . . . .	61
2.5.4	Résultats de la projection . . . . .	62
2.5.5	Critiques de cette projection . . . . .	64
2.6	Conclusion . . . . .	66



<i>TABLE DES MATIÈRES</i>	25
<b>3 Projection temporelle de la mortalité par réseaux RVFL</b>	<b>67</b>
3.1 Protocole expérimental . . . . .	69
3.2 Projection par Vecteur Autoregressif . . . . .	71
3.2.1 Spécification du modèle . . . . .	71
3.2.2 Présentation des résultats obtenus par VAR(1) . . . . .	71
3.3 Projection par réseaux <i>RVFL</i> . . . . .	74
3.3.1 Présentation des modèles <i>RVFL</i> . . . . .	74
3.3.2 Projection par <i>RVFL(1)</i> . . . . .	77
3.3.3 Projection des séries différenciées par <i>RVFL(1)</i> . . . . .	79
3.3.4 Réduction de la dimension par ACP sur les séries différenciées . . . . .	82
3.4 Conclusion . . . . .	86
<b>Conclusion</b>	<b>87</b>
<b>Bibliographie</b>	<b>88</b>
<b>Annexes</b>	<b>93</b>
<b>A Ajustement de la mortalité française par modèle <i>Logit+ Additif</i></b>	<b>95</b>
<b>B Détails de la projection par VAR(1)</b>	<b>111</b>



# Introduction

Par définition l'assurance vie regroupe l'ensemble des contrats ayant pour sous-jacent la vie d'un individu ou d'un groupe d'individus. L'étude des durées de vie et de la mortalité est donc indispensable à la conception, à la tarification, au provisionnement et à la gestion de tels contrats. Certaines activités comme la retraite nécessitent des projections à long terme. En effet, une personne va généralement commencer à cotiser vers 25 ans. Or depuis les années 50 l'espérance de vie à la naissance augmente de 6h par jour en moyenne. Les risques financiers liés à la longévité encourus par les assureurs, les fonds de pension et les caisses de retraite peuvent donc potentiellement être très importants.

Identifier des structures dans les dynamiques de mortalité des populations est une tâche complexe à cause des nombreux phénomènes impactant les taux de mortalité. De nombreux modèles ont été développés depuis l'introduction du premier proposé par Lee & Carter (1992). Les projections à court terme reposent généralement sur une simple extrapolation des tendances historiques de mortalité, des espérances de vie ou des paramètres des modèles. Mais pour des périodes de plus d'une dizaine d'années, une projection linéaire peut mener à des résultats incohérents. Un jugement d'expert est souvent utile pour juger de la plausibilité des tendances à long terme ainsi obtenues.

Les modèles courant échouent à projeter la mortalité de manière efficace pour plusieurs raisons :

- Les modèles traditionnels, même ceux où l'effet cohorte\* est pris en compte, ajustent relativement bien les données historiques mais ont de faibles capacités de projection. Ils sur-apprennent les données.
- Une autre approche a émergé avec Bongaarts (2004) qui propose d'ajuster la courbe des taux de mortalité par année avec un modèle paramétrique, puis d'extrapoler les coefficients obtenus par des méthodes de série temporelle. Dans son étude Bongaarts utilise un modèle exponentiel pour ajuster la mortalité aux âges élevés. Il ne projette qu'un seul paramètre avec une méthode ARIMA. Son modèle n'ajuste que les grands âges. Il est trop simple pour être appliqué à toute la population adulte.

Récemment Debonneuil (2014) a caractérisé de façon très fine toute la mortalité adulte avec une forme paramétrique plus complexe. La problématique de ce mémoire est donc de reprendre les travaux de Debonneuil (2014) afin de projeter les coefficients d'un modèle proche du sien avec la

---

\*. Voir définition 1.3.1

méthodologie de Bongaarts (2004) :

- Dans la première partie nous posons le contexte de l'étude. Nous décrivons la mortalité française récente. Nous présentons les définitions et les données utilisées. Nous montrons les limites des modèles traditionnels. Puis nous décrivons le modèle d'ajustement proposé par Debonneuil (2014).
- Dans la deuxième partie nous développons un modèle dont le calibrage est plus facile. En effet pour notre travail nous avons besoin de calibrer le modèle de mortalité sur de nombreuses années afin d'en extraire les séries temporelles des coefficients. Nous développons toute une méthodologie pour le calibrer de manière efficace. Nous analysons ensuite les coefficients obtenus. Nous remarquons qu'ils caractérisent de manière explicite les différents évènements et tendances démographiques survenus sur la population française. Cela ouvre la voie à l'utilisation de ce modèle et de cette méthodologie à des fins de construction de scénarios démographiques et de gestion de risque de longévité.
- Les séries de coefficients étant fortement dépendantes les unes des autres, nous utilisons des modèles avancés afin de les projeter dans le temps en partie trois. Nous utilisons d'abord un modèle linéaire VAR. Puis pour tenir compte des relations non linéaires entre coefficients nous utilisons un réseau de neurones RVFL régularisé. Pour avoir des résultats cohérents et éviter le sur-apprentissage nous retraits les séries afin de les rendre plus stationnaires et de réduire la dimension du problème.

En conclusion cette étude apporte de nouvelles méthodes pour analyser et projeter la mortalité à long terme. Les modèles sont appliquées aux données issues de la population française, mais la méthodologie générale est utilisable sur n'importe quelle population à condition de disposer d'assez de données. Ce mémoire ouvre donc des perspectives sur la gestion à long terme des portefeuilles retraites.





# Chapitre 1

## Développement des modèles paramétriques de mortalité

Dans cette première partie, nous nous appuyons sur une analyse fine des durées de vie passées pour souligner l'intérêt et l'utilité des modèles paramétriques de mortalité. Nous posons les définitions et les notations nécessaires à l'étude de la mortalité, puis nous présentons de manière critique les principaux modèles existants.

### 1.1 Contexte de l'étude et problématiques liées à la mortalité

En France, l'espérance de vie à la naissance a triplé de 1800 à 2017. Depuis 1950, elle est passée de 66 à 82,5 ans (voir figure 1.1), soit une augmentation moyenne de 6 heures par jour (Tekian 2018). La principale cause de l'augmentation de la durée de vie au cours des siècles derniers était la baisse de la mortalité infantile (voir figure 1.2).

Mais depuis une soixantaine d'année, les avancées médicales augmentent l'espérance de vie des personnes plus âgées (voir figure 1.3). Ces améliorations sont dues en premier lieu à la lutte contre les maladies infectieuses et respiratoires, puis aux progrès quant aux maladies cardiovasculaires et enfin plus récemment aux meilleurs traitements contre les cancers. En parallèle, la science contre le vieillissement progresse de manière fulgurante sur les animaux en laboratoire. Des scientifiques ont en effet réussi à augmenter l'espérance de vie à la naissance de souris de 70% à l'aide d'une seule modification génétique et d'un régime adapté. (Bartke et al., 2008)

Ces observations soulèvent plusieurs questions du point de vue de l'assurance : la longévité va-t-elle continuer à augmenter ? Si oui, à quel rythme ? Ces questions sont primordiales pour les actuaires. La durée de vie des individus est un sous-jacent de nombreux produits d'assurance. On peut citer les produits d'épargne, les temporaires décès, les garanties décès, les capitaux différés etc... Mais les plus sensibles à la longévité sont les produits de retraites à cause de leurs encours importants et de leurs maturités très longues de plusieurs dizaines d'années. Zhavoronkov et al. (2012) ont estimé que si les maladies cardiovasculaires et les cancers étaient éliminés, la Sécurité Sociale et le programme Medicare aux États-Unies serraient en déficits de plus de 80 000 milliard

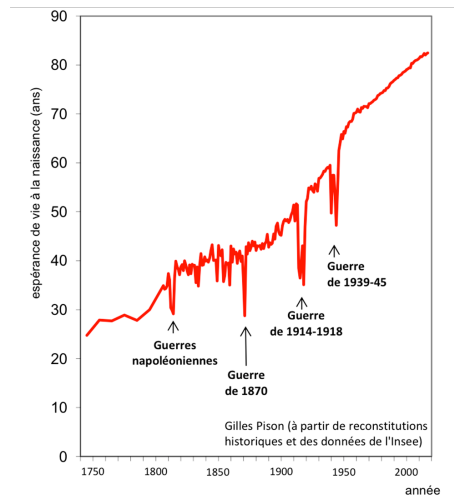


FIGURE 1.1 – Évolution de l'espérance de vie à la naissance en France.

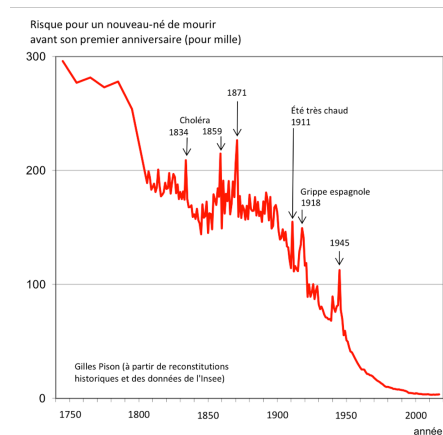


FIGURE 1.2 – Évolution de la mortalité infantile en France.

de dollars. Ces produits sont de plus très sensibles car ils sont souvent la seule source de revenus des personnes âgées. Les actuaires ont donc voulu modéliser la mortalité pour comprendre et anticiper la longévité humaine.

Les projections d'espérance de vie réalisées par le passé ont été démenties par la réalité (voir figure 1.4). Face aux difficultés de modélisation de la mortalité la réglementation prudentielle s'est adaptée de manière simpliste. Actuellement le choc de longévité de la Formule Standard de la réglementation Solvabilité II impose aux assureurs de provisionner assez de fonds pour faire face à une baisse relative de 20% de la probabilité de décès à tous les âges\*. La réglementation se veut générale pour être facilement applicable, mais elle ne rend pas compte des risques précis encourus par les assureurs.

\*. Article 138 du Règlement Délégué Solvabilité II au 22 Juin 2018.



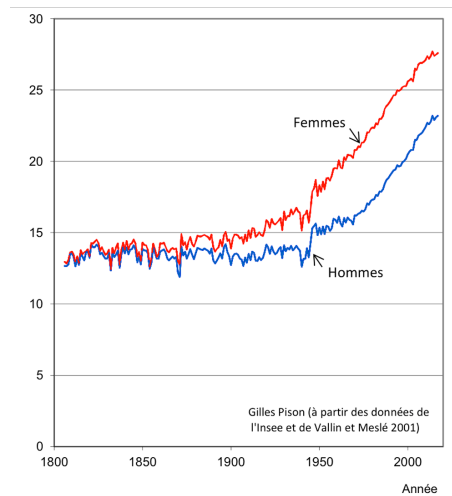


FIGURE 1.3 – Évolution de l'espérance de vie à 60 ans en France.

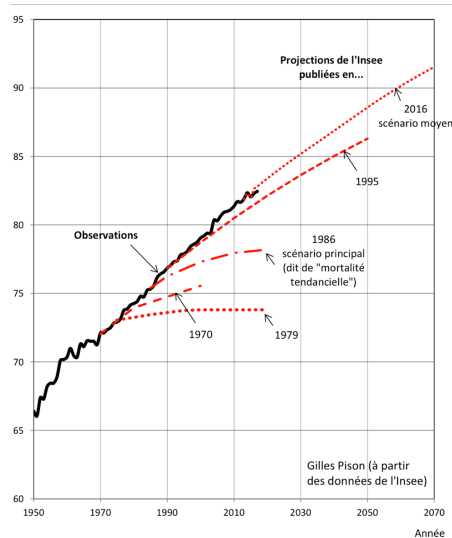


FIGURE 1.4 – Évolution de l'espérance à la naissance en France, observations et projections.

Dans l'optique d'une meilleure maîtrise des risques, des modèles d'évolution de la mortalité ont été développés. Dans la suite de cette partie, après avoir défini les outils mathématiques et présenté les données utilisées par la suite, nous détaillons les modèles existants en soulignant leurs forces et leurs faiblesses.

## 1.2 Définitions, notations et présentation des données

Pour étudier la mortalité les démographes se sont dotés d'outils mathématiques présentés ci-dessous. Il ont aussi recensés des données précises disponibles en libre accès que nous utiliserons pour développer, tester et comparer plusieurs modèles de mortalité.

### 1.2.1 Définitions et notations utiles dans l'étude de la mortalité

Afin de définir des modèles de mortalité, nous introduisons plusieurs notations classiquement utilisées pour l'étude des populations et des durées de vie.

**Définition 1.2.1** Soit la variable aléatoire  $T_x(t)$  représentant la durée de vie restante d'un individu d'âge  $x$  dans l'année calendaire  $t$ .

**Définition 1.2.2** On note  $q_x(t)$  la probabilité annuelle de décès d'un individu d'âge  $x$  dans l'année calendaire  $t$  :

$$q_x(t) = \mathbb{P}[T_x(t) \leq 1].$$

**Définition 1.2.3** On note  $p_x(t)$  la probabilité annuelle de survie d'un individu d'âge  $x$  dans l'année calendaire  $t$  :

$$p_x(t) = \mathbb{P}[T_x(t) > 1] = 1 - q_x(t).$$

**Définition 1.2.4** Le nombre d'individus vivants à l'âge  $x$  au début de l'année calendaire  $t$  est noté  $L_{x,t}$ .

**Définition 1.2.5** Le nombre de décès à l'âge  $x$  dans l'année calendaire  $t$  est noté  $D_{x,t}$ .

**Définition 1.2.6** Le temps durant lequel les individus sont exposés au risque de décès est noté :

$$E_{x,t} = \int_{\nu=0}^1 L_{x+\nu,t+\nu} d\nu.$$

Il correspond à la durée totale vécue par les individus durant la période d'observation.

**Définition 1.2.7** Le risque de décès instantané ou force de mortalité à l'âge  $x$  et à l'année  $t$  est noté  $\mu_x(t)$ . Il est défini par :

$$\mu_{x+\epsilon}(t) = \lim_{\Delta\epsilon \rightarrow 0^+} \frac{\mathbb{P}[\epsilon < T_x(t) \leq \epsilon + \Delta\epsilon \mid T_x(t) > \epsilon]}{\Delta\epsilon}.$$

Dans la suite de ce mémoire nous ferons l'hypothèse de constance des forces de mortalité à un âge donnée  $x$  sur toute la durée d'une année. Sous cette hypothèse :

$$p_x(t) = 1 - q_x(t) = e^{-\int_0^1 \mu_{x+\nu}(t+\nu) d\nu} = e^{-\mu_x(t)}. \quad (1.1)$$

Il est à noter que sous une approximation d'ordre 1 :  $\mu_x(t) = q_x(t)$ .

**Définition 1.2.8** Avec les notations précédentes on estime l'espérance de vie à l'âge  $x$  par :

$$e_x = \frac{\sum_{k=1}^{\infty} L_{x+k}}{L_x}.$$

On considérera pour cela l'espérance de vie finie. De ce fait, les données utilisées et présentées ci-dessous conciderent un âge limite de 110 ans, c'est à dire :  $\forall x > 110, L_x = 0$ .

**Définition 1.2.9** On appelle taux d'amélioration, noté  $IR$ , le pourcentage de baisse de mortalité à l'âge  $x$  entre l'année  $t$  et  $t + 1$  :

$$IR_x(t) = 1 - \frac{q_x(t-1)}{q_x(t)}.$$

Ces taux serviront à juger de la plausibilité et de la cohérence des tendances des modèles extrapolant la mortalité.

Toutes ces valeurs sont définies pour un âge  $x$  et une année  $t$ . La modélisation et la prédiction de la mortalité est alors un problème en deux dimensions : selon l'âge et l'année, on parlera de surface de mortalité (figure 1.5).

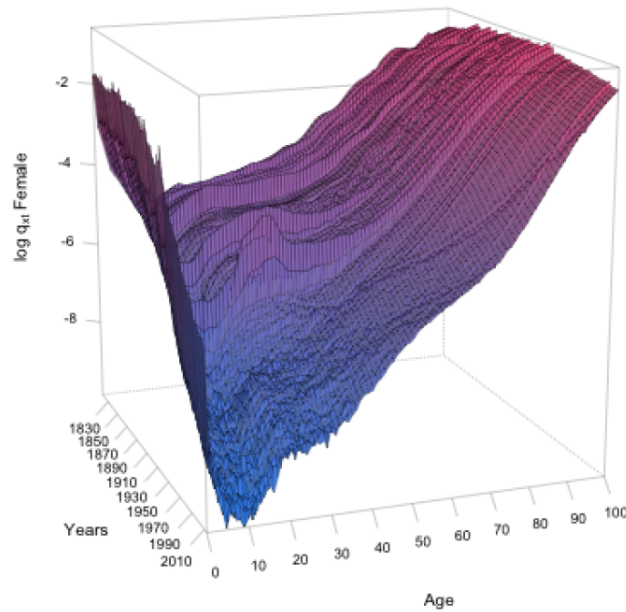


FIGURE 1.5 – Surface de mortalité ( $q_x(t)$ ) de la population féminine française sur une échelle logarithmique de 0 à 100 ans entre 1830 et 2010 (Thomas 2018).

**Hypothèse 1.2.1** On considérera la mortalité assez régulière dans l'espace du temps et des âges, "Natura non agit per saltum" (Leibniz).

### 1.2.2 Présentation des données de la *Human Mortality Database*

La base de données sur la mortalité humaine (HMD) (Shkolnikov et al. 2000) a été créée pour fournir des données détaillées sur la mortalité et la population aux chercheurs, étudiants, journalistes, analystes politiques et autres personnes intéressées par l'histoire de la longévité humaine. Ce projet est le travail de deux équipes de chercheurs aux Etats-Unis et en Allemagne, avec l'aide de bailleurs de fonds et de collaborateurs scientifiques du monde entier. Au moment de la rédaction de ce mémoire, cette base contient les données détaillées sur la mortalité de 39 pays.

Dans ce mémoire nous utiliserons les variables suivantes pour la population française :

- *Exposure-to-risk* : Estimations de la population exposée au risque de décès sur un certain intervalle d'âge :  $E_{x,t}$ . Elles sont fondées sur les estimations démographiques annuelles au 1er janvier, avec une petite correction qui reflète le moment des décès dans l'intervalle.
- *Deaths* : Décomptes des décès. Si les données brutes sont agrégées, des méthodes uniformes sont utilisées pour estimer les décomptes de décès en fonction de l'âge au moment du décès, de l'année civile du décès et de l'année civile de naissance.
- *Death rates* : Les forces de mortalité sont estimées par le ratio du nombre de morts dans un intervalle d'âge donnée (*Deaths*) sur une estimation de l'exposition au risque (*Exposure-to-risk*) dans le même intervalle :  $\hat{\mu}_x(t) = \frac{\hat{D}_{x,t}}{\hat{E}_{x,t}}$ .

Il est à noter que par convention les tables s'arrêtent à 110 ans. Toutes les personnes encore en vie à cet âge sont comptées comme décédées. De plus, à cause de la faible exposition aux âges élevés, les taux de mortalité au delà de 95 ans pour toutes les années sont lissés par le modèle logistique suivant inspiré de Thatcher et al. (1998) :

$$\mu_x(a, b) = \frac{ae^{b(x-80)}}{1 + ae^{b(x-80)}}.$$

Ces ajustements sont visibles sur la figure 1.6.

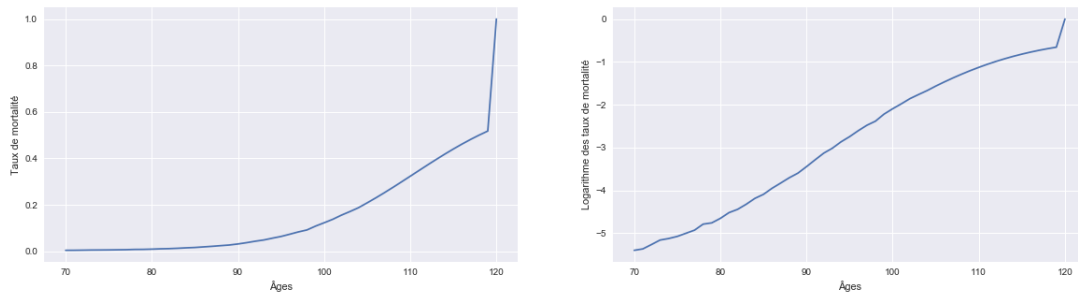


FIGURE 1.6 – Taux de mortalité issus de la HMD observés sur la population féminine française en 2005 aux âges élevés, sur une échelle logarithmique à droite.

Maintenant que nous disposons des outils mathématiques et des données relatives à la mortalité, nous allons présenter les modèles de mortalité existants ainsi que leurs avantages et leurs limites.

## 1.3 Présentation des modèles de référence et de leurs limites

L'étude de la mortalité consiste à trouver des modèles caractérisant  $\mu_x(t)$  ou  $q_x(t)$  afin de les projeter dans le temps. Il est à noter que l'équation 1.1 implique que l'étude d'une de ces deux grandeurs est équivalente à l'étude de l'autre.

Dans cette section nous présentons trois modèles de mortalité connus. En pratique les modèles les plus utilisés par les professionnels sont des modèles paramétriques reposant sur une décomposition en différents facteurs de la surface de mortalité. Ces facteurs ont pour but de capter l'évolution complexe de la mortalité à travers le temps et les âges.

### 1.3.1 Le modèle de *Lee-Carter*

Le modèle le plus connu et servant de référence est celui de Lee & Carter (1992). Ces derniers ont proposé un modèle qui décompose la surface de mortalité en une tendance latente  $\kappa_t$  et deux paramètres sensibles à l'âge  $\alpha_x$  et  $\beta_x$  :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}.$$

Pour rendre le modèle identifiable il faut ajouter deux contraintes, généralement :  $\sum_x \beta_x = 1$  et  $\sum_t \kappa_t = 0$ .  $\alpha_x$  s'interprète comme la moyenne temporelle des  $\ln(\mu_{x,t})$  alors que  $\beta_x$  traduit la sensibilité des forces de mortalité à l'âge  $x$  par rapport à l'évolution générale  $\kappa_t$ . Des travaux récents ont montré les limites de ce modèle, par exemple Giacommetti et al. (2012), Chai et al. (2013), Hunt & M.Villegas (2015) et Mavros et al. (2016).

Premièrement un de ses principaux désavantages est l'hypothèse d'homoscédasticité des résidus  $\epsilon_{x,t}$ , leur variance n'est pas constante. En pratique, la variance de ces derniers varie avec le temps et l'âge. En effet, à cause notamment de la baisse des effectifs, la variance du taux de décès croît aux âges élevés, voir Lee & Miller (2001) et Gao & Hu (2009).

Deuxièmement lorsque l'on affiche les résidus du modèle de Lee & Carter (1992) qui repose sur les facteurs *âge* et *période*, on observe une structure diagonale reflétant l'effet cohorte (voir définition 1.3.1) comme le montre la figure 1.7. L'évolution de la mortalité est connue pour être liée à l'année de naissance (Willets 2004). C'est ce que l'on appelle l'effet cohorte. Les personnes ayant été exposées aux mêmes facteurs historiques comme les guerres, les épidémies, les mêmes habitudes alimentaires ou tabagiques sont susceptibles d'avoir des formes de mortalité proches. Ceci est dû à la persistance de certains chocs sur la mortalité au sein d'une même génération.

**Définition 1.3.1** *On appelle cohorte un ensemble d'individus nés au même moment, c'est à dire à peu d'année d'intervalle.*

Troisièmement la figure 1.8 montre que les taux d'amélioration sont projetés par ce modèle pour un âge donnée de façon constante. Les projections de ce modèle sont réalisées par l'extrapolation de

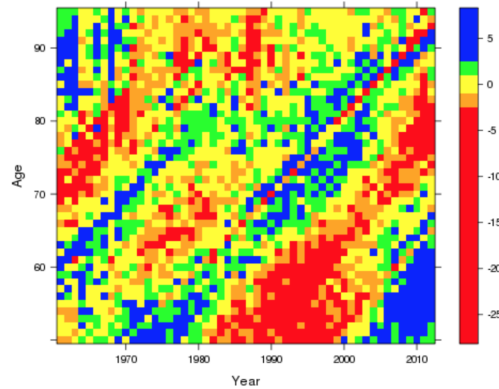


FIGURE 1.7 – Résidus de Pearson pour le modèle de Lee-Carter calibré sur la population anglaise masculine de 50 à 95 ans entre 1960 et 2012 (Currie 2018).

$\kappa_t$  dans le temps à l'aide de modèle de projection de série temporelle de type *ARIMA*. Ce dernier n'a donc que peu de pouvoir prédictif et n'est pas capable d'extrapoler des tendances dans le futur.

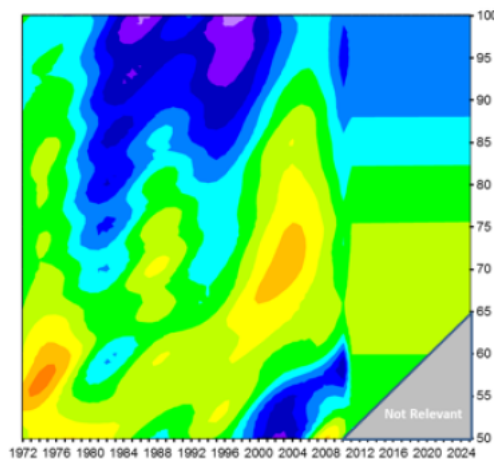


FIGURE 1.8 – Taux d'amélioration empiriques et projetés par le modèle de Lee-Carter sur la population américaine masculine de 1972 à 2024 entre 50 et 100 ans (Thomas 2018).

### 1.3.2 Le modèle *Âge-Période-Cohorte*

Afin de palier aux problèmes de Lee & Carter (1992), d'autres modèles s'en sont inspirés en ajoutant des composantes qui capturent des effets d'âge, de période et de cohorte. Le modèle Age-Période-Cohorte (APC) en est un premier exemple (Renshaw & Haberman 2006). Pour résoudre le problème d'homoscédasticité des résidus, ce dernier modélise le nombre de décès par une loi de Poisson plutôt que les forces de mortalité :

$$D_{x,t} \sim \mathcal{P}(E_{x,t}\mu_{x,t}), \text{ avec } \ln(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} + \epsilon_{x,t}.$$

Pour rendre le modèle identifiable il faut ajouter trois contraintes, généralement :  $\sum_t \kappa_t = 0$ ,  $\sum_{t-x} \gamma_{t-x} = 0$  et  $\sum_{t-x} (t-x)\gamma_{t-x} = 0$ . Le modèle permet de capturer les influences distinctes des effets liés à l'âge avec  $\alpha_x$ , à la période avec  $\kappa_t$  et à la cohorte avec  $\gamma_{t-x}$ .

Ainsi l'incorporation de l'effet cohorte  $\gamma_{t-x}$  dans le modèle APC réduit la non stationnarité des résidus (figure 1.9) en comparaison avec le modèle de Lee & Carter (1992). De plus la figure 1.10 montre certaines tendances dans la projection des taux de mortalité, notamment avec la prise en compte des effets cohortes. Cependant toutes les projections ne sont pas cohérentes notamment aux grands âges. Cela peut être dû au fait que ce modèle est sur-paramétré et sur-apprend les données d'entraînement.

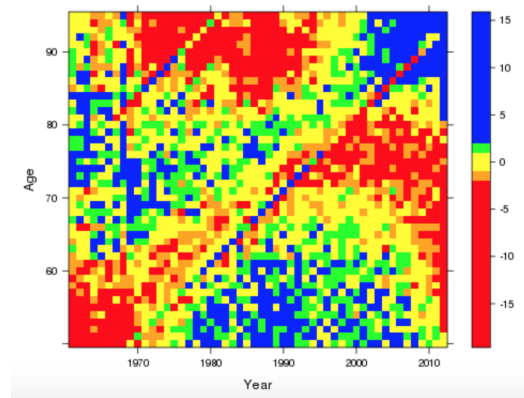


FIGURE 1.9 – Résidus de Pearson pour le modèle APC calibré sur la population anglaise masculine de 50 à 95 ans entre 1960 et 2012 (Currie 2018).

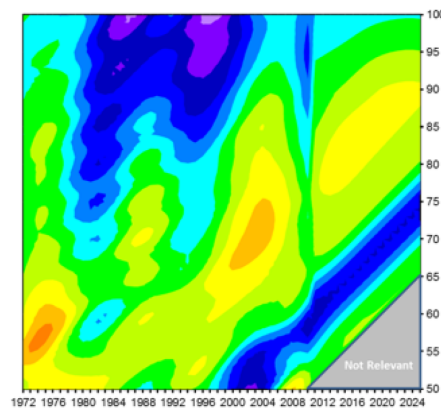


FIGURE 1.10 – Taux d'amélioration empiriques et projetés par le modèle APC sur la population américaine masculine de 1972 à 2024 entre 50 et 100 ans (Thomas 2018).

Cependant ce modèle suppose des effets de période et de cohorte indépendants de l'âge. En pratique ces effets peuvent dépendre de l'âge.

### 1.3.3 Le modèle *Lee-Carter-Cohorte*

Le dernier modèle présenté dans cette partie est plus complet car il combine les modèles APC et Lee-Carter. Il s'agit du modèle *Lee-Carter Cohort* (LCC) de Renshaw & Haberman (2006) :

$$D_{x,t} \sim \mathcal{P}(E_{x,t}, \mu_{x,t}), \text{ avec } \ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \gamma_{t-x} + \epsilon_{x,t}.$$

Avec les contraintes  $\sum_x \beta_x = 1$ ,  $\sum_t \kappa_t = 0$  et  $\sum_{t-x} \gamma_{t-x} = 0$ . Le modèle LCC donne de meilleurs résultats que les précédents en combinant les effets liés à l'âge, à la période, à la cohorte et à l'évolution générale. La structure des résidus de ce modèle est observable sur la figure 1.11. Les effets cohorte ont presque disparu. Le tableau suivant (Currie 2018) présente une comparaison des qualités d'ajustement des différents modèles par critères de Déviance et d'AIC sur la population anglaise masculine. Il montre que ce dernier modèle reflète beaucoup mieux les données observées en comparaison avec les modèles précédents.

Modèle	Déviance	AIC
Lee-Carter	17,057	17,341
APC	15,907	16,291
LCC	7,202	7,678

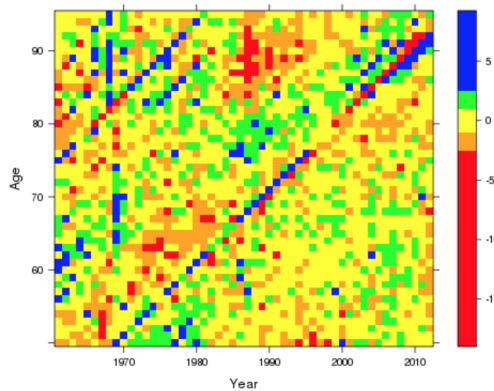


FIGURE 1.11 – Résidus de Pearson pour le modèle LCC calibré sur la population anglaise masculine de 50 à 95 ans entre 1960 et 2012 (Currie 2018).

Cependant la figure 1.12 montrent qu'il y a peu de cohérence entre les observations et les projections des taux d'amélioration. Ce modèle est sur-paramétré et sur-apprend les données d'entraînement. Pour cette raison il ajuste très bien la mortalité observée mais n'est pas capable de la projeter de façon consistante dans le temps.

Néanmoins on ne sait pas très bien comment l'effet cohorte doit être interprété et identifié. Ceci est mis en lumière par certains résultats empiriques récents (Hunt & M.Villegas 2015). Bien que les performances d'ajustement des modèles APC et LCC soient meilleures que celles du modèle de Lee-Carter, leurs prévisions sont instables. Ces modèles sont sur-paramétrés et ont tendance à conduire au sur-apprentissage. Leurs prévisions ne sont donc pas fiables.



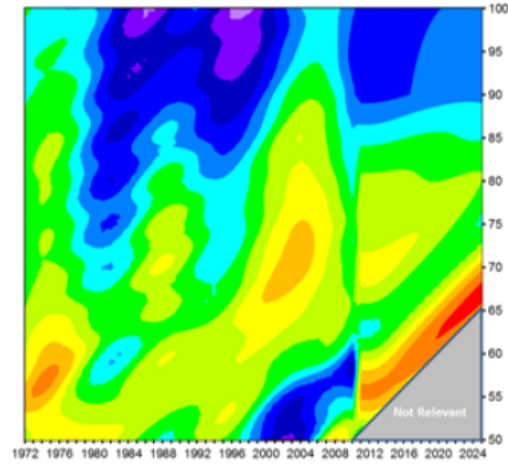


FIGURE 1.12 – Taux d’amélioration empiriques et projetés par le modèle LCC sur la population américaine masculine de 1972 à 2024 entre 50 et 100 ans (Thomas 2018).

## 1.4 Présentation des modèles avancés existants

Face à la constatation de ces insuffisances un autre méthodologie a émergé. Dans un travail réalisé pour la Banque Mondiale, Bongaarts (2004) propose de ne caractériser la mortalité que dans l'espace des âges, pour une année donnée, puis de projeter un ou plusieurs paramètres du modèle par une méthode de série temporelle dans l'espace du temps. Dans cette partie nous présentons quelques uns des modèles existants utilisables pour caractériser la courbe de mortalité pour une seule année donnée.

### 1.4.1 Le modèle de Thatcher

Le modèle de Thatcher et al. (1998) est une forme paramétrique de la mortalité pour une année donnée. Il est inspiré d'une forme logistique et s'exprime de la manière suivante :

$$\mu_x = \frac{\alpha e^{\beta x}}{1 + \alpha e^{\beta x}} + \gamma.$$

$\gamma$  s'interprète comme le niveau minimum de mortalité (il peut actuellement être négligé Bongaarts (2004), et Debonneuil (2014)),  $\alpha$  régit la position de la courbe dans la dimension des âges et  $\beta$  dirige la pente de la fonction logistique.

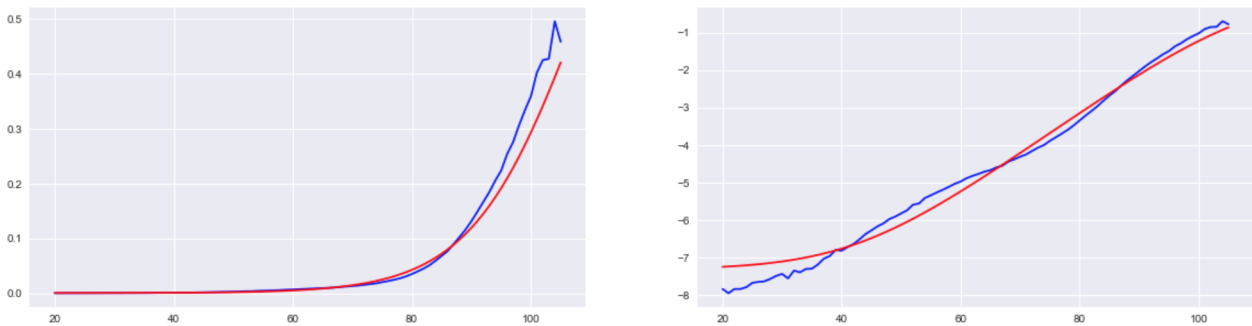


FIGURE 1.13 – Ajustement (par maximum de la log vraisemblance, voir partie 2) de la mortalité observée (en bleu) en fonction de l'âge par le modèle de Thatcher (en rouge), sur échelle logarithmique à droite. Population française en 2014.

Ce modèle a deux inconvénients majeurs :

- Il ne s'adapte qu'aux grands âges. Sur la figure 1.13, nous pouvons clairement observer que la bosse autour de 55 ans empêche le modèle de s'ajuster correctement sur toute la courbe. C'est le modèle utilisé pour reconstituer les taux de mortalité aux âges supérieurs à 95 ans dans la HMD (voir partie 1.2.2).
- Il considère que  $\mu_\infty = 100\%$ , c'est à dire  $q_\infty = 1 - e^{-1} = 63\%$ . Cette valeur est discutable. Les études empiriques tendent à montrer que la mortalité limite se situe plutôt autour de 50% (Debonneuil 2014). La faible exposition aux grands âges complique l'étude de la mortalité des personnes âgées, il peut être nécessaire de faire des hypothèses pour fermer les tables de mortalité.

### Fermeture des tables de mortalité

Les expositions sont très faibles aux grands âges, ce qui rend les modèles instables. Dans cette sous-partie nous testons une méthode de fermeture brutale afin de vérifier que l'espérance de vie du moment n'est que marginalement influencée par la fermeture de la table.

Pour cela nous calculons l'espérance de vie du moment à 75 ans (âge approximatif moyen d'un portefeuille retraite en phase de rente) de la population française en 2015 empirique (population A) et sur la même population avec les individus qui meurent au maximum à 95 ans (population B). Les résultats sont présentés dans le tableau suivant :

Population	Espérance de vie du moment à 75 ans
A	13.04
B	12.56

La différence relative entre ces deux scénarios est de moins de 4%. Compte tenu de la sévérité du scénario B, nous pouvons affirmer qu'une variation de moins de 20% du paramètre  $q_\infty$  n'a que peu d'influence sur l'espérance de vie. Par la suite nous fixerons la limite de mortalité à  $q_\infty = 50\%$  (Debonneuil 2014) lorsque les modèles le permettront.

#### 1.4.2 Le modèle *Logit+*

Devant l'incapacité du modèle de Thatcher à généraliser l'ensemble des formes de mortalité rencontrées, Debonneuil (2014) a créé et étudié un nouveau modèle paramétrique de mortalité plus souple appelé modèle *logit+*. Son approche consiste à s'intéresser à la mortalité adulte (la mortalité des personnes âgées de moins de 25 ans n'est pas un sujet d'étude pour des applications retraites) en s'inspirant de la méthodologie de Heligman & Pollard (1980). Ces derniers ont décomposé la courbe de mortalité en trois composantes :

$$\frac{q_x}{1 - q_x} = ae^{(x+b)^c} + de^{-e(\ln(x)-\ln(f))} + gh^x.$$

$$\text{Avec : } \frac{q_x}{1 - q_x} = gh^x \iff q_x = \frac{ge^{x\ln(h)}}{1 + ge^{x\ln(h)}}.$$

La figure 1.14 tirée de leur article donne une représentation visuelle de l'effet de chaque composante sur le modèle. Chacune d'entre elles caractérise la mortalité sur une plage d'âge donnée, à savoir la mortalité infantile de la naissance à 10 ans, la mortalité des jeunes adultes jusqu'à 30 ans, puis la mortalité du reste de la population.

Néanmoins la courbe actuelle ne ressemble plus à celle des années 1970, et une nouvelle modélisation est nécessaire. On remarque en effet sur la figure 1.15 que les taux de mortalité ont baissé, mais qu'une bosse est apparue autour de 55 ans. Ce phénomène est observable dans plusieurs pays (Debonneuil 2014), notamment en Allemagne, Belgique, Espagne, mais ni au Royaume-Uni, ni aux États-Unis. Cette bosse peut s'expliquer par une baisse de la mortalité générale qui fait ressortir la mortalité précoce due à certaines maladies. Il est à noter que cette bosse est plus faible pour les populations assurées (Debonneuil 2014).

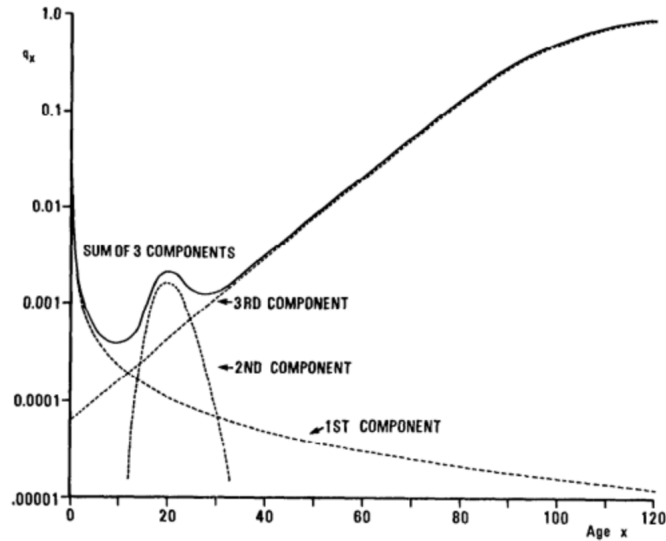


FIGURE 1.14 – Courbe de mortalité et ses trois composantes. Population masculine australienne de 1970 à 1972 sur une échelle logarithmique. Illustration tirée de Heligman & Pollard (1980).

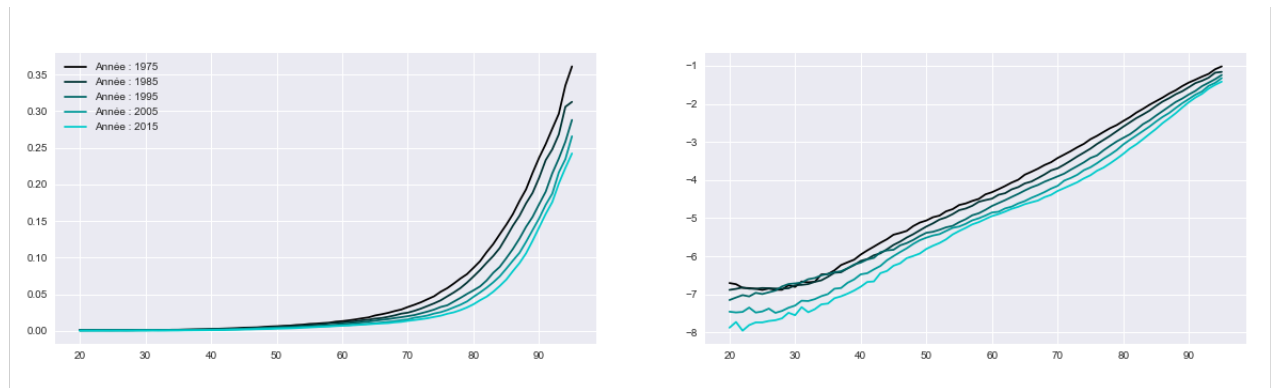


FIGURE 1.15 – Évolution de la mortalité en France de 1985 à 2015, sur une échelle logarithmique à droite.

Pour prendre en compte cette spécificité aux âges qui nous intéressent, le modèle *logit+* s'exprime pour une année  $t$  donnée par :

$$q(x) = Old(x)(1 + Mid(x) + Young(x)) + c_{young}.$$

Avec les composantes suivantes :

$$\begin{aligned} Young(x) &= e^{a_{young} - b_{young}x}, \\ Mid(x) &= a_{mid} e^{-\frac{(x - c_{mid})^2}{b_{mid}}}, \\ Old(x) &= q_{asymp} \frac{1}{1 + e^{-a_{old}x - b_{old}}}. \end{aligned}$$

- La composante *Old* suit une loi logistique paramétrée à 50% ( $q_{asympt} = 50\%$ ). Elle représente l'aplanissement progressif des taux de mortalité aux âges supérieurs à 75 ans.
- La composante *Mid* est inspirée d'une densité gaussienne. Elle représente la surmortalité autour de 55 ans.
- La composante *Young* est une exponentielle décroissante. Elle modélise le plateau de mortalité entre 20 et 30 ans. Le paramètre  $c_{young}$  est similaire à  $\gamma$  dans le modèle de Thatcher (partie 1.4.1). Il est facultatif au vu des niveaux de mortalité qu'il caractérise qui sont des centaines de fois plus petits que la mortalité aux âges élevés.

Il y a 7 coefficients à calibrer :  $a_{old}$ ,  $b_{old}$ ,  $a_{mid}$ ,  $b_{mid}$ ,  $c_{mid}$ ,  $a_{young}$  et  $b_{young}$ .  $q_{asympt}$  est en effet considéré comme un paramètre fixé à 50%. Les coefficients  $a_{old}$  et  $b_{old}$  sont à calibrer sur des âges supérieurs à 75 ans,  $a_{mid}$ ,  $b_{mid}$  et  $c_{mid}$  sur des âges autour de 50 ans et  $a_{young}$  et  $b_{young}$  sur des âges inférieurs à 40 ans.

Ce modèle étant multiplicatif, les trois composantes sont corrélées de manière significative entre elles : un ajustement de la composante *Old* devra être suivi par un ajustement des composantes *Mid* et *Young*. L'ajustement est donc délicat et la variance des coefficients obtenus est grande.

De plus les paramètres ne sont pas intuitifs à interpréter. Par exemple  $c_{mid}$  sensé représenter le point culminant de la surmortalité adulte sera en réalité autour de 25, et non de 50 comme attendu à cause de la composante *Old* qui est multipliée par la *Mid*. Pour ces raisons nous avons choisi de retenir le modèle additif présenté dans la partie suivante.

## 1.5 Conclusion

Nous avons vu que les projections de mortalité passées ont toutes été démenties par la réalité. Les insuffisances des modèles classiques de type Lee-Carter nous ont amené à changer de méthodologie de modélisation. Au lieu de décomposer la surface de mortalité en différents facteurs, nous ajustons une forme paramétrique pour tous les âges pour une année donnée. Le chapitre 2 détaille la méthode utilisée pour ajuster cette forme paramétrique sur toutes les années disponibles. Les coefficients de cette dernière sont extrapolés dans le temps dans le chapitre 3.

## Chapitre 2

# Définition et calibrage d'un modèle de mortalité générique

Suite aux difficultés de calibrage et d'interprétation rencontrées avec le modèle *Logit+*, nous présentons ici le modèle *Logit+ additif*. Nous détaillons ensuite la procédure utilisée pour calibrer ce modèle sur des dizaines d'années différentes. Nous remarquons que les coefficients du modèle reflètent de façon très lisibles les tendances et événements démographiques de la population française sur la deuxième moitié du vingtième siècle. Nous montrons donc en quoi cette étude permet de construire des scénarios démographiques prospectifs afin de maîtriser au mieux les risques de mortalité et de longévité. Les coefficients issus de ces calibrages seront ensuite extrapolés en partie 3 pour inférer la mortalité future.

### 2.1 Modèle *Logit+* additif

Suite à plusieurs expérimentations le modèle retenu est le suivant, pour une année  $t$  donnée :

$$\mu(x) = Old(x) + Mid(x) + Young(x).$$

Avec les composantes :

$$\begin{aligned} Young(x) &= e^{a_{young} + b_{young}x}, \\ Mid(x) &= \beta e^{-\frac{(x-m)^2}{\sigma}}, \\ Old(x) &= \mu_{asympt} \frac{1}{1 + e^{-ax-b}}. \end{aligned}$$

Les composantes sont donc les mêmes que celles du modèle *Logit+*. On ignore le paramètre  $c_{young}$  du modèle précédent (voir partie 1.4.2). On fixera aussi  $\mu_{asympt}$  pour faciliter le calibrage. La figure 2.1 présente le modèle calibré et ses trois composantes.

Nous avons choisi de travailler directement sur les forces de mortalité  $\mu_x$ . Celles-ci sont en effet estimables directement à partir du décompte des décès  $D_x$  et de l'exposition au risque  $E_x$  disponible dans la HMD. On a  $\hat{\mu}_x = \frac{D_x}{E_x}$ .

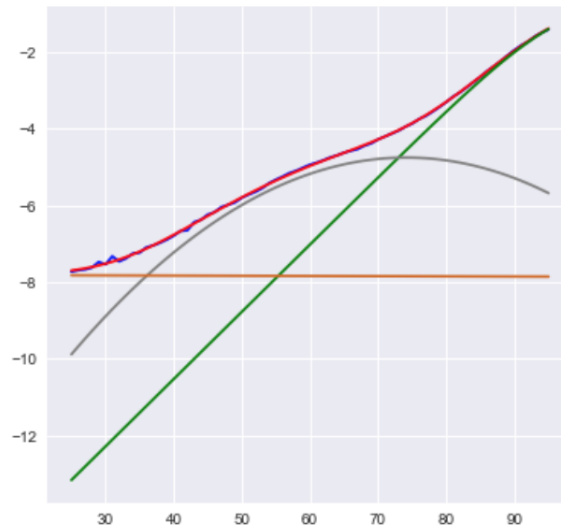


FIGURE 2.1 – Ajustement de la mortalité observée (en bleu) en fonction de l'âge par le modèle *Logit+ additif* (en rouge) sur échelle logarithmique. Les trois composantes sont *Old* en vert, *Mid* en gris et *Young* en chocolat. Population française en 2015.

Les paramètres sont cette fois plus faciles à interpréter. Dans l'exemple de la figure 2.1, les valeurs obtenues pour les coefficients sont les suivantes :

Coefficient	Valeur
$a_{young}$	-7.814
$b_{young}$	$-5.08 \cdot 10^{-4}$
$\beta$	$8.57 \cdot 10^{-3}$
$m$	74.118
$\sigma$	470.379
$a$	0.175
$b$	-17.038

Les coefficients  $\mu_{asympt}$ ,  $a$  et  $b$  s'interprètent comme dans les modèles de *Thatcher* et *Logit+*. Ils régissent la forme de la mortalité aux âges élevés.  $\sigma$  caractérise l'étendu de la bosse,  $m$  le mode de celle-ci. Il est intéressant de noter que le mode ne sera pas forcément situé à l'endroit où la bosse semble la plus marquée sur une échelle logarithmique.  $\beta$  représente la hauteur maximale de la bosse.  $a_{young}$  et  $b_{young}$  dirigent la droite des âges jeunes visibles sur la figure 2.1. Cette composante sert à aplatir la mortalité des jeunes. Il est important de noter que sur la figure 2.1 les différentes composantes semblent très proches les unes des autres. Mais du fait de l'échelle logarithmique la mortalité à un âge donné est en réalité presque intégralement représentée par la composante la plus importante.

Il faut désormais à calibrer ce modèle sur toutes les données historiques disponibles.



## 2.2 Méthodes d'ajustement du modèle Logit+ additif

Nous décrivons ici une méthode efficace de calibrage du modèle. Nous décrivons ensuite la stratégie employée pour le calibrer sur plusieurs dizaines d'années.

### 2.2.1 Calibrage par maximum de vraisemblance

Notons  $E_x$  l'exposition à l'âge  $x$  et  $D_x$  le nombre de morts dans les 12 derniers mois à l'âge  $x$ .  $D_x$  peut donc être vu comme la somme de la réalisation de  $E_x$  variables aléatoires suivant une loi binomiale de paramètre  $q_x$ . On considérera les décès indépendants. La vraisemblance de l'échantillon est donc :

$$L(q_x) = \prod_{x=x_{min}}^{x_{max}} \binom{E_x}{D_x} q_x^{D_x} (1 - q_x)^{E_x - D_x}$$

Le problème revient à trouver les huit coefficients qui déterminent  $q_x$  pour tous les âges observés tel que  $L(q_x)$  soit maximisé. Il est équivalent de maximiser la log-vraisemblance :

$$\begin{aligned} l(q_x) &= \sum_{x=x_{min}}^{x_{max}} \left[ \ln\left(\binom{E_x}{D_x}\right) + D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x) \right] \\ &= \sum_{x=x_{min}}^{x_{max}} [D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x)] + \text{constante} \end{aligned}$$

En pratique nous maximiserons  $\sum_{x=x_{min}}^{x_{max}} [D_x \ln(q_x) + (E_x - D_x) \ln(1 - q_x)]$ , avec  $q_x$  donné par le modèle retenu, dans notre cas le modèle Logit+ additif.

### 2.2.2 Calibrage par minimisation d'une erreur d'ajustement

Une autre solution est de minimiser une métrique d'erreur. Il en existe plusieurs candidates. Soient  $q_x$  le vecteur des taux de mortalité observés et  $\hat{q}_x$  le vecteur de leurs estimations par le modèle retenu.

On peut chercher à minimiser l'erreur des moindres carrés  $\|q_x - \hat{q}_x\|_{I_d}^2$  avec  $\|\alpha\|^2$  la norme euclidienne du vecteur  $\alpha$  pondérée par la matrice identité  $I_d$ . La mortalité aux âges inférieurs à 70 ans étant très faible devant la mortalité aux âges plus élevés, cette erreur privilégiera l'ajustement aux âges élevés qui sont susceptibles de produire des erreurs élevées.

Une deuxième solution est d'utiliser l'erreur des moindres carrés pondérés  $\|q_x - \hat{q}_x\|_M^2$ , avec  $M$  la matrice des pondérations. Bien choisir la pondération  $M$  n'est pas évident. Nous pouvons privilégier les âges jeunes pour corriger le biais de la méthode précédente. Une pondération par l'exposition  $E_x$  peut aller dans ce sens. A cause du problème délicat de la pondération, nous n'utiliserons pas cette métrique.

Une troisième méthode consiste à minimiser l'erreur des moindres carrés du logarithme des taux de mortalité :  $\| \ln(q_x) - \ln(\hat{q}_x) \|_{Id}^2$ . Nous contournerons alors le problème de la première erreur. Avec cette méthode, nous remarquons empiriquement que les taux de mortalité aux âges élevés ne sont pas bien ajustés. Cela peut être dû aux données qui ont une variance intrinsèque plus importante à ces âges là à cause de l'exposition qui est plus faible. Nous pouvons alors pondérer cette métrique aux grands âges pour corriger ce biais. Néanmoins une pondération par l'exposition fait aussi sens, au détriment de l'ajustement aux âges élevés. Nous nous rapprochons alors des résultats obtenus par maximum de la log vraisemblance.

Ces méthodes ont été expérimentées mais ont montrés des résultats moins stables et/ou beaucoup plus long à obtenir qu'un ajustement par maximum de log-vraisemblance. De part son approche plus naturelle, nous utilisons ce dernier par la suite.

### 2.2.3 Optimisateur utilisé

Quelque soit la méthode de calibrage retenue, il n'existe pas de solution explicite évidente pour maximiser la log-vraisemblance ou minimiser une erreur pour le modèle Logit+ additif. L'expression du gradient est complexe. Nous avons donc eu recours aux algorithmes de différenciation automatique et aux optimisateurs implémentés dans la bibliothèque TensorFlow de Abadi et al. (2015) développée par Google afin d'implémenter un algorithme de descente de gradient. Il est important de noter que la différenciation automatique n'est pas une différenciation numérique basée sur le calcul de la limite d'un taux d'accroissement qui peut être délicat en pratique. Cette technique manipule des opérations. Lors de la phase de déclaration, les dérivés de chaque opération doivent être renseignés. Le différenciateur utilise ces dérivés ainsi que la formule de dérivation en chaîne pour calculer le gradient d'une expression complexe. Il est alors nécessaire de ne calculer l'expression du gradient qu'une seule fois au début de l'algorithme, et de s'y référer à chaque itération en une opération lors de la minimisation.

Cette partie traite en détaille les avantages et les inconvénients de différents algorithmes afin de sélectionner le plus adapté à notre problème.

#### Algorithme du gradient

Une fois que le gradient a été calculé par différenciation automatique, il se pose la question de la mise à jour des paramètres. Soit  $f(\theta)$  la fonction à minimiser dépendante des  $d$  paramètres  $\theta \in \mathbb{R}^d$  et  $\nabla_{\theta} f(\theta)$  son gradient par rapport aux paramètres  $\theta$ . Le premier choix à l'itération  $t$  est de prendre :

$$\theta_t = \theta_{t-1} - \eta \nabla_{\theta} f(\theta_t)$$

Avec  $\eta$  un paramètre généralement choisi par validation croisée. Il faudra le faire décroître au fil de l'apprentissage pour converger vers un minimum. En fonction de la vitesse de décroissance, cette méthode pourra soit converger vers un minimum local, soit ne pas converger, soit converger en un temps très long. Plusieurs méthodes ont été développées pour limiter ces risques.

## Momentum

Cette méthode (Qian 1999) introduit un nouveau vecteur  $v_t$  qui garde en mémoire les mises à jour du gradient aux itérations précédentes. On a alors à l'itération  $t$  :

$$\begin{aligned}v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} f(\theta_t) \\ \theta_t &= \theta_{t-1} - v_t\end{aligned}$$

$\gamma$  est un nouveau paramètre de l'algorithme qui est généralement fixé à 0.9 (Qian 1999), mais peut être choisi par validation croisée. Empiriquement cette méthode est plus efficace car l'algorithme "slalome" moins dans l'espace.

## Gradient accéléré de Nesterov

L'algorithme du gradient accéléré de Nesterov (Nesterov 1983) consiste en une petite modification de la méthode du Momentum qui permet de s'extraire de minimums locaux. À l'itération  $t$  :

$$\begin{aligned}v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} f(\theta_{t-1} - \gamma v_{t-1}) \\ \theta_t &= \theta_{t-1} - v_t\end{aligned}$$

Jusqu'ici  $\eta$  contrôle la magnitude de la mise à jour du gradient. Il serait plus optimal d'avoir un contrôle fin de ce paramètre afin que chaque paramètre du vecteur  $\theta$  soit mis à jour avec une magnitude différente. Ceci est primordiale pour notre problème sachant que les paramètres du modèle *Logit+ additif* n'ont pas tous la même échelle.

## Adagrad

Adagrad (John Duchi & Singer 2011) autorise une calibration fine des différentes composantes du gradient en ayant des magnitudes individuelles pour chaque paramètre. Cet algorithme modifie la vitesse d'apprentissage  $\eta$  à chaque itération  $t$  pour tous les paramètres  $\theta_i$  selon les gradients précédents qui ont été calculés pour  $\theta_i$ . À l'itération  $t$  :

$$\begin{aligned}v_t &= \sum_{t'=0}^{t-1} (\nabla_{\theta} f(\theta_{t'}))^2 \\ \theta_t &= \theta_{t-1} - \frac{\eta}{\sqrt{v_t + \epsilon}} \nabla_{\theta} f(\theta_t)\end{aligned}$$

Ainsi si une des composantes du gradient est faible, le pas suivant celle-ci sera plus grande et inversement. Un des avantages de cet algorithme est que le pas d'apprentissage  $\eta$  peut être fixé une fois pour toute, habituellement à  $10^{-2}$  (John Duchi & Singer 2011) et ne pas être choisi par validation croisée.  $\epsilon$  est fait pour éviter une division par 0, il est habituellement fixé à  $10^{-8}$ .

## RMSprop

Adagrad somme les gradients depuis la première itération. L'amplitude des mises à jour est donc forcée à diminuer. Ce comportement peut être limité avec la technique RMSprop\* qui calcule une moyenne exponentielle des gradients précédents :

$$v_t = \gamma v_{t-1} + (1 - \gamma) \nabla_{\theta} f(\theta_t)$$

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{v_t + \epsilon}} \nabla_{\theta} f(\theta_t)$$

$\gamma$  et  $\eta$  sont généralement fixés à 0.9 et  $10^{-3}$ . Ces paramètres n'ont pas besoin d'être modifiés d'une itération sur l'autre.

## Adadelta

Adadelta (Zeiler 2012) est une extension de RMSprop qui a pour but de se passer du paramètre  $\eta$ . Pour ce faire, un nouveau vecteur  $m$  est introduit :

$$v_t = \gamma v_{t-1} + (1 - \gamma) \nabla_{\theta} f(\theta_t)$$

$$\theta_t = \theta_{t-1} - \frac{\sqrt{m_{t-1} + \epsilon}}{\sqrt{v_t + \epsilon}} \nabla_{\theta} f(\theta_t)$$

$$m_t = \gamma m_{t-1} + (1 - \gamma) \left( \frac{\sqrt{m_{t-1} + \epsilon}}{\sqrt{v_t + \epsilon}} \nabla_{\theta} f(\theta_t) \right)^2$$

Le paramètre  $\eta$  est complètement supprimé, mais la procédure pour faire cela ne semble pas naturelle.

## Adam

La dernière technique présentée dans ce mémoire est Adam (Kingma & Ba 2014). Elle garde en mémoire la trace du gradient et de son carré à l'aide de deux vecteurs  $m$  et  $v$  mis à jour à chaque pas de temps :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} f(\theta_t)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} f(\theta_t))^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t$$

---

\*. Cette technique n'a pas été publiée, mais a été proposée par Geoff E. Hinton dans un cours : [www.cs.toronto.edu/~tijmen/csc321/slides/lecture\\_slides.lec6.pdf](http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides.lec6.pdf)

Avant la mise à jour des paramètres,  $m_t$  et  $v_t$  sont normalisés pour éviter les problèmes numériques lors des premières itérations.  $\eta$  est alors le pas maximal que l'algorithme pourra effectuer.

En pratique les méthodes RMSprop, Adadelta et Adam sont adaptées à notre problème car le pas d'une itération s'adapte à chaque composante. Nous implémentons cette dernière car c'est l'une des plus utilisée en pratique. Nous l'emploierons implicitement dans la suite de ce mémoire pour chaque étape d'optimisation. Les paramètres initiaux seront ceux de la publication originale (Kingma & Ba 2014).

#### 2.2.4 Méthodologie générale employée

Il est nécessaire de disposer d'une méthodologie efficace afin de calibrer le modèle *Logit+ Additif* sur toutes les années. Ce calibrage devra être fait plusieurs dizaines de fois pour une population donnée.

Nous commençons par calibrer le modèle pour une année donnée en initialisant les paramètres avec une solution proche de l'optimale trouvée de manière empirique. Comme le problème n'est pas convexe, le point de départ de l'algorithme est important sinon la méthode peut converger vers un minimum local non satisfaisant.

Nous commençons le processus de calibration dans les années 1950. La mortalité avant ces années là est fortement perturbée par les guerres. Puis la solution trouvée une année donnée donnera le point de départ du calibrage de l'année suivante, et ce jusqu'en 2015. Cette méthodologie est motivée par l'hypothèse 1.2.1.

Certains paramètres seront projetés à chaque itération de l'algorithme du gradient pour qu'ils ne caractérisent que les formes voulues :

- $\beta$  qui représente l'amplitude de la gaussienne autour de 60 ans sera maintenu supérieur ou égale à 0. Sans cela, si aucune bosse n'est présente dans les données ce paramètre pourra devenir négatif et caractériser du bruit ou un creux dans la mortalité (Debonneuil 2014).
- Dans un premier temps  $b_{young}$  qui représente la pente de la composante *Young* est maintenu entre 0 et 0.038 afin d'empêcher la composante *Young* de représenter l'information de *Old* ou de *Mid*. Ces valeurs ont été trouvées de manière empirique. La figure 2.2 montre un exemple d'ajustement obtenu sans projection des paramètres. On remarque que la composante *Young* représente presque en intégralité la mortalité. La composante *Old* est alors complètement décalée et ne caractérise plus la mortalité aux grands âges. Nous réalisons cela dans un souci d'interprétabilité des coefficients obtenus.

Nous utiliserons une combinaison de deux critères d'arrêt classiques pour l'algorithme du gradient :

- Un nombre d'itérations maximum,
- Une métrique d'erreur qui ne diminue plus à un  $\epsilon = 10^{-9}$  près depuis un certains nombre d'itérations.

Nous validerons les ajustements année par année de façon visuelle. Il est en effet difficile de se fier uniquement à la log vraisemblance, celle-ci étant influencée par l'exposition qui évolue au cours du temps. Nous ne pouvons pas non plus nous fier à une métrique d'erreur tel que les moindres

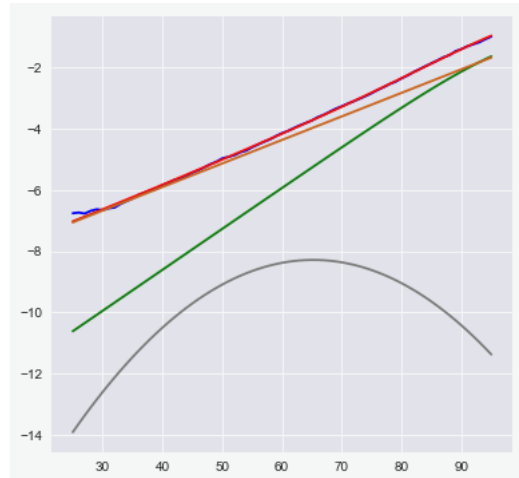


FIGURE 2.2 – Exemple d’ajustement de la mortalité française en 1970 sans projection des paramètres.

carrés ou les moindres carrés du logarithme de la mortalité car les résidus d’une année à l’autre ne possèdent pas la même variance.

## 2.3 Présentation de l'ajustement de la mortalité

Nous présentons dans cette partie les résultats obtenus par maximisation de la log-vraisemblance. Ces résultats sont utilisés dans la partie 3 pour projeter la mortalité.

### 2.3.1 Ajustement initial

Trois exemples d'ajustements obtenus avec la méthode décrite précédemment sont présentés en figure 2.3. Les coefficients obtenus sur la période 1955-2015 sont présentés sur la figure 2.4.

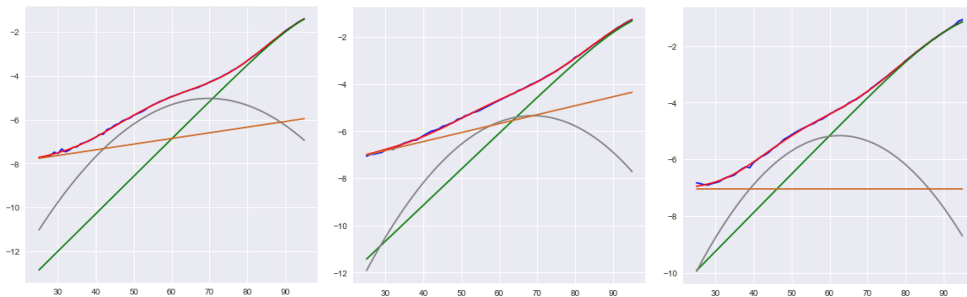


FIGURE 2.3 – Ajustement de la mortalité de la population française en 2015, 1996 et 1981.

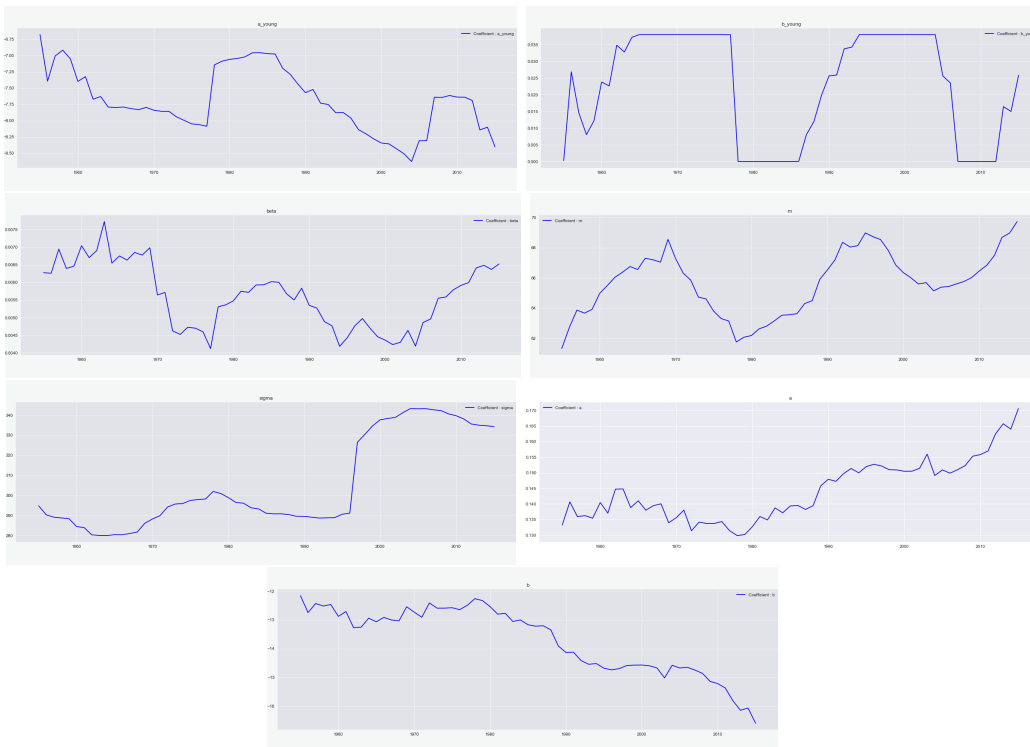


FIGURE 2.4 – Coefficients obtenus par maximum de vraisemblance sur la population française entre 1955 et 2015.

Nous remarquons qu'il y a du bruit dans les séries temporelles induites par les coefficients

calibrés chaque année. Le modèle extrapolant ces séries devra les lisser.

De plus le coefficient  $b_{young}$  évolue dans l'intervalle  $[0, 0.038]$  car on le projète à chaque itération du gradient. Il est la plus part des années proche d'une de ses bornes. Il correspond à la pente de la composante *Young* et agit comme un filtre : on observe une bosse dans les données autour de 55 ans lorsque  $b_{young}$  est proche de 0.

Les paramètres  $m$ ,  $b_{young}$ ,  $a_{young}$  présentent une saisonnalité qui correspond à priori à l'apparition ou non d'une bosse autour de 55 ans. Celle-ci est sera difficile à capturer car nous n'avons que deux périodes.

Les coefficients sont dépendants les uns des autres. Certains sauts ou changements de tendances sont synchronisés comme en 1978 ou encore en 2005. Ce comportement était attendu car les composantes se positionnent les unes par apport aux autres, si l'une bouge, les autres sont modifiées.

De plus nous observons un saut de  $\sigma$  en 1997. Ce changement brutal influe sur la mortalité. À des fins de prédiction et au vu de l'historique assez court des données, il conviendrait de trouver un autre ajustement plus régulier.

### 2.3.2 Ajustement retenu

Pour contourner le problème soulevé au paragraphe précédent, nous avons décidé de fixer le paramètre  $b_{young}$ . Cette solution présente l'avantage de stabiliser les autres coefficients et de réduire le nombre de valeurs à projeter. Empiriquement nous l'avons fixé à 0.038. Cette valeur a été déterminée de manière à réduire au maximum les changements bruts dans les autres coefficients tout en conservant une qualité d'ajustement suffisante. Tous les ajustements sont présentés en annexe A. Les séries obtenues ainsi sont visibles en figure 2.5.

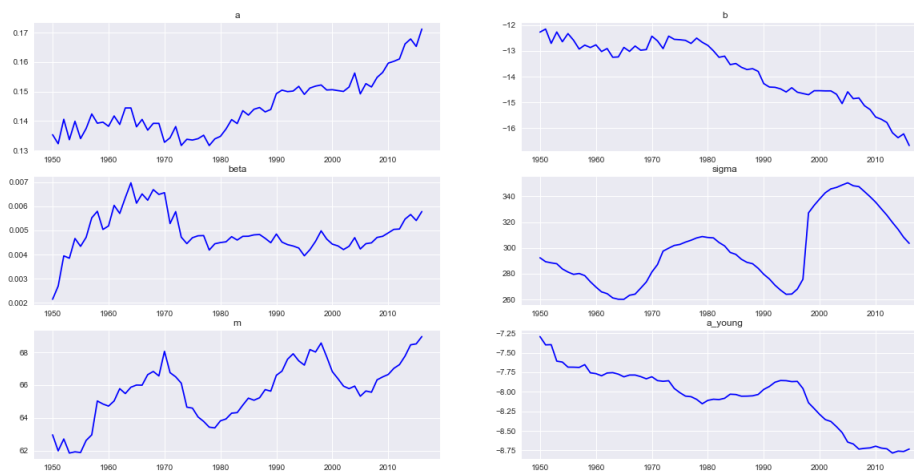


FIGURE 2.5 – Coefficients obtenus par maximum de log vraisemblance sur la population française entre 1955 et 2015 en fixant  $b_{young}$ .



Ce modèle ajuste la mortalité française de manière satisfaisante. Afin de comprendre et de maîtriser la sens de chaque coefficient, nous allons analyser l'ajustement obtenu au regard de l'histoire démographique française.

## 2.4 Analyse des coefficients ajustés

Les différents coefficients présentent plusieurs tendances que nous pouvons décomposer en trois catégories :

- Le coefficient  $a_{young}$  diminue avec une tendance linéaire depuis les années 1950. On notera une légère bosse au milieu des années 1990.
- Les coefficients  $a$  et  $b$  sont constants puis présentent des tendances linéaires à partir de 1980 respectivement croissante et décroissante.
- Les coefficients  $m$  et  $\sigma$  présentent une saisonnalité.

Nous effectuons donc l'analyse précise de ces trois points.

### 2.4.1 Une mortalité générale s'abaissant avec le temps

Dans la décomposition réalisée par le modèle Logit+ additif,  $a_{young}$  caractérise l'abscisse à l'origine de la droite ajustant les âges jeunes sur une échelle logarithmique. Or la composante *Young* représente un niveau minimum de mortalité observé. Une tendance à la baisse de  $a_{young}$  signifie donc un abaissement générale de la courbe de mortalité dans le temps et donc une augmentation de l'espérance de vie à tous les âges. Le tableau suivant présente les corrélations entre  $a_{young}$  et les espérances de vie du moment à plusieurs âges. Il montre que ce coefficient est effectivement très corrélé négativement avec la hausse de l'espérance de vie à tous les âges.

	$e_{30}$	$e_{50}$	$e_{70}$	$e_{90}$
$a_{young}$	-0.85970867	-0.8555384	-0.8553134	-0.86009551

Néanmoins la corrélation n'est pas parfaite, ce qui indique que d'autres phénomènes ont eu lieu dans l'histoire démographique française qui expliquent l'augmentation générale de l'espérance de vie.

### 2.4.2 Rectangularisation de la fonction de survie

Les coefficients caractérisant les âges élevés,  $a$  et  $b$  définissant la composante *Old*, sont constants jusqu'à la fin des années 1970, puis ont une tendance linéaire par la suite. Pour rappel cette composante s'exprime par :  $Old(x) = \mu_{asympt} \frac{1}{1+e^{-ax-b}}$ . À partir de 1975  $b$  diminue. Cela signifie que la mortalité des âges élevés diminue aussi. À partir de la même période  $a$  augmente. Cela se traduit par la pente de la composante *Old* qui devient plus importante. Ce modèle capte donc le phénomène de rectangularisation de la fonction de survie : avec le temps les personnes meurent moins de manière précoce et l'âge moyen des décès hors décès précoce augmente. Cela est notamment mis en lumière par Fiallos & Noel (2018).

### 2.4.3 Influence de cohortes particulières

Le coefficient  $m$  caractérise le pic de sur-mortalité précoce autour de 65 ans. Il peut se décomposer en un effet de saisonnalité et une tendance linéaire. La tendance linéaire à la hausse indique qu'avec le temps cette sur-mortalité précoce avance vers les âges élevés. La saisonnalité traduit le passage de certaines cohortes arrivant à des âges proches de 60 ans.

La diminution de  $m$  d'une année sur l'autre reflète le passage d'une génération dont la sur-mortalité précoce survient plus tôt que pour la générations la précédant. On identifie ainsi deux générations remarquables visibles sur la figure 2.6 :

- La génération correspondant au premier rectangle rouge sur la figure 2.6. Elle correspond aux personnes nées entre 1902 et 1916. Nous l'appellerons cohorte A.
- La génération correspondant au deuxième rectangle rouge sur la figure 2.6. Elle correspond aux personnes nées entre 1930 et 1940. Nous l'appellerons cohorte B.

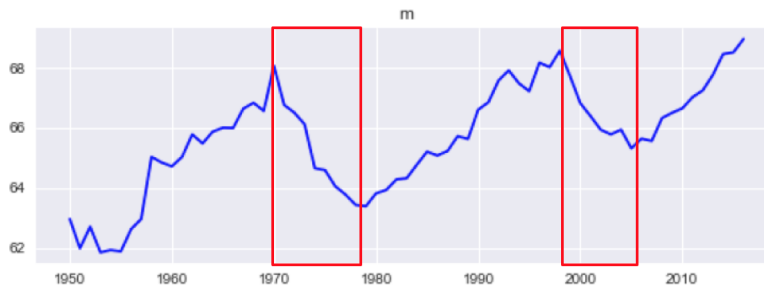


FIGURE 2.6 – Coefficients  $m$  obtenus par maximum de log vraisemblance sur la population française entre 1955 et 2015 avec deux générations remarquables.

Ces générations ont pour point commun d'avoir vécu la guerre pendant leur enfance et adolescence. Une étude de Horiuchi (1983) a mis en lumière l'impacte de certaines déficiences alimentaires sur ces générations créant un terrain favorable aux maladies de dégénérescence cardio-vasculaire.

Certaines causes de sur-mortalité sont différentes entre ces générations. Les générations de 1916, 1917 et 1918 dont la sur-mortalité est la plus précoce ont été frappées par la grippe espagnole de l'hiver 1918-1919. Cette maladie peut entraîner des complications dont les nouveaux nés, population particulièrement vulnérable, ont gardés des séquelles. Cette génération est aussi combattante pendant la seconde guerre mondiale. Par ailleurs la génération des années 1930 à 1940 a été très touchée par l'alcoolisme et les accidents de la route se généralisant au cours des années 1950 et 1960 (Vallin & Lopez (1985) et Vallin & Meslé (1988)). Nous remarquons que  $m$  diminue plus pour A que pour B, preuve que cette cohorte a été plus marquée.

De plus lorsque  $m$  atteint un minimum local, en 1980 et 2005,  $\sigma$  atteint un maximum local. Pour rappelle la composante  $Mid$  s'exprime par :  $Mid(x) = \beta e^{-\frac{(x-m)^2}{\sigma}}$ .  $\sigma$  correspond à la variance de la gaussienne représentant la surmortalité aux âges intermédiaires. Il reflète donc l'étendu de la cohorte à la mortalité intermédiaire plus précoce.

Lorsque  $m$  est important  $\sigma$  diminue. Or lorsque la surmortalité précoce intervient à un âge assez élevé la bosse n'est plus significative dans le modèle et les composantes  $Old$  et  $Young$  prennent le

dessus sur la composante *Mid*. Les deux premières illustrations de la figure 2.7 montre en effet que pour  $m$  grand, comme en 1970 et 1996, la bosse de surmortalité n'apparaît pas. Au contraire cette bosse est clairement visible en 1980 et 2005 comme le montre les deux dernières illustrations de la figure 2.7. Donc lorsque  $m$  est grand,  $\sigma$  n'influe plus sur la forme générale de la mortalité.

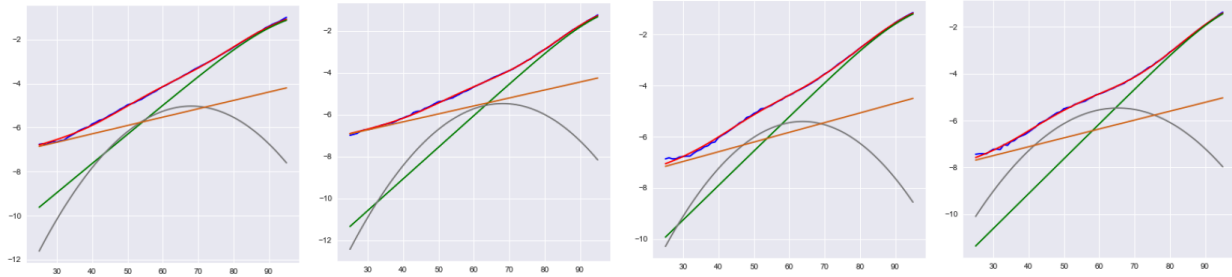


FIGURE 2.7 – Ajustement du modèle Logit+ additif et de ses composantes sur la population française en 1970, 1996, 1980 et 2005, échelle logarithmique

## 2.5 Projection de la mortalité par composante

À partir de ces observations il est intéressant de projeter la mortalité de façon à définir un scénario démographique de référence. Ce scénario général pourra être utilisé par des assureurs en ajustant la mortalité de leur portefeuille d'assurés par rapport à la population française sur des données historiques. Puis ils pourront inférer la mortalité future de leurs assurés en fonction de ce scénario de référence.

### 2.5.1 Projection de la composante *Old*

Nous avons vu que les coefficients  $a$  et  $b$  définissant *Old* évoluent indépendamment des autres coefficients. Or depuis les années 1980 ils suivent une tendance linéaire reflétant la rectangularisation de la fonction de survie. Pour établir un scénario de référence nous prolongerons ces tendances linéaires. Nous faisons donc l'hypothèse que la rectangularisation de la fonction de survie va se poursuivre au même rythme.

### 2.5.2 Projection de la composante *Young*

Le coefficient  $a_{young}$  reflète l'abaissement générale de la courbe de mortalité. Or nous remarquons que depuis 2006 ce dernier reste constant. Nous prolongerons ce coefficient de façon constante suivant la moyenne de ses valeurs au cours des 10 dernières années d'observation. Nous faisons donc l'hypothèse que la courbe de mortalité ne va plus s'abaisser de façon général. Cette projection est donc prudente pour l'estimation du risque de mortalité, mais imprudente pour du risque de longévité.

### 2.5.3 Projection de la composante *Mid*

Le coefficient  $\beta$  reste constant depuis les années 1980. Nous le projeterons donc de manière constante suivant la moyenne de ses valeurs sur les 35 dernières années d'observation.

La bosse de sur-mortalité aux âges intermédiaires n'est présente que lorsque qu'une cohorte à la mortalité précoce plus importante atteint l'âge de 64 ans environ. Si la tendance linéaire de  $m$  se poursuit sans le passage d'une de ces cohortes, la composante *Mid* restera inférieurs aux composantes *Old* et *Young*. Si nous considérons qu'aucune cohorte ne passera dans les prochaines années nous pouvons projeter  $m$  en suivant sa tendance linéaire.

Il reste alors à projeter  $\sigma$ . Or à première vue ce paramètre est très lié à la saisonnalité de  $m$ . Pour mieux comprendre l'interaction entre ces deux paramètres nous pouvons ajuster le modèle Logit+ additif en considérant uniquement la tendance linéaire de  $m$ . La figure 2.8 présente cette ajustement. La figure 2.9 présente les coefficients ainsi obtenus.

Nous remarquons que seul  $\sigma$  a un comportement différent par rapport au précédent calibrage.

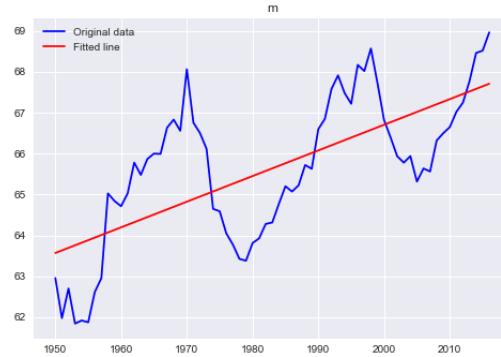


FIGURE 2.8 – Coefficient  $m$  obtenu par maximum de log vraisemblance sur la population française entre 1955 et 2015 en bleu et sa tendance linéaire en rouge.

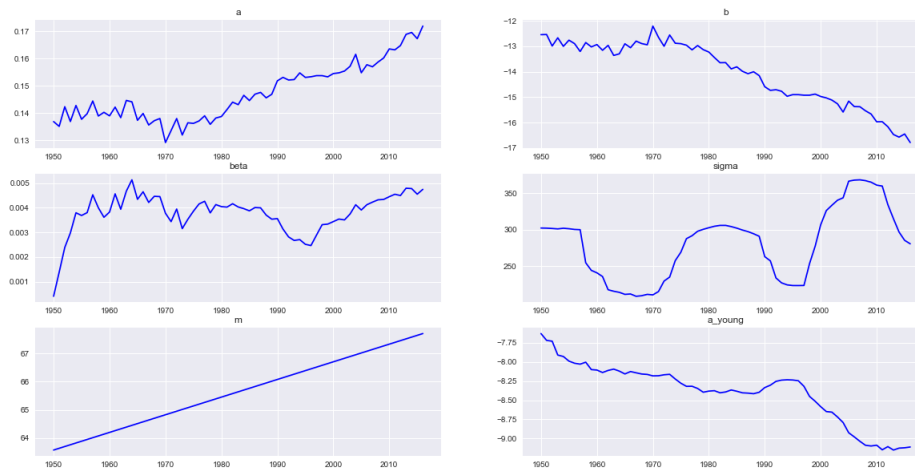


FIGURE 2.9 – Coefficients obtenus par maximum de log vraisemblance sur la population française entre 1955 et 2015 en fixant  $b_{young}$  et  $m$ .

Il garde sa saisonnalité mais ne présente plus vraiment de tendance comme on pouvait l'observer auparavant. Lorsque  $m$  est grand, il marque toujours des paliers, à 260 auparavant et à 210 ici. Nous projetterons donc de façon linéaire les 5 dernières années de  $\sigma$  jusqu'à 210, valeurs minimale obtenue pour  $m$  ajustée avec une tendance linéaire.

## 2.5.4 Résultats de la projection

Les coefficients initiaux et projetés selon les indications des sous parties précédentes sont visibles sur la figure 2.10. Ces projections donnent les projections d'espérance de vie du moment de la figure 2.11. Nous pouvons remarquer les espérances de vie du moment à tous les âges augmentent de façon presque linéaire en suivant la tendances de ses dernières années. Le ralentissement de leurs augmentations est certainement du au coefficient  $a_{young}$  qui ne décroît plus contrairement au passé. Il est intéressant de noter que cette projection est linéaire aux âges élevés car le phénomène de

rectangularisation de la fonction de survie prend le dessus sur la mortalité générale qui ne s'abaisse plus.

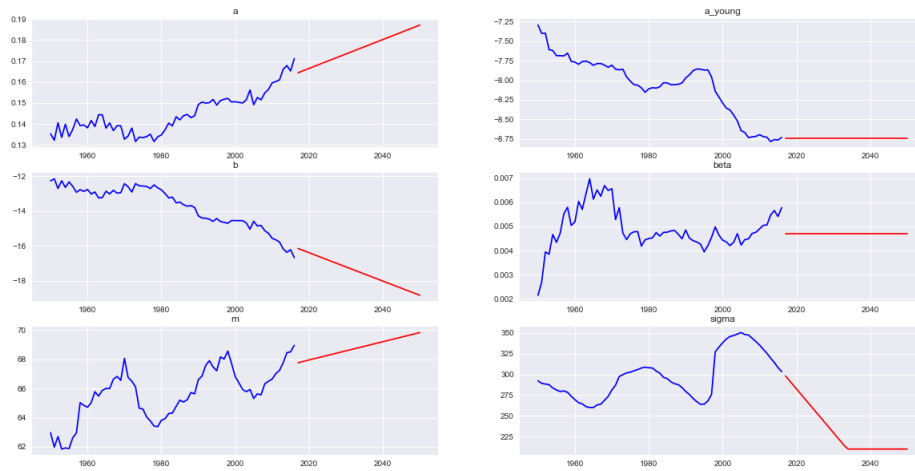


FIGURE 2.10 – Séries initiales en bleu et projetées manuellement en rouge entre 1955 et 2050.

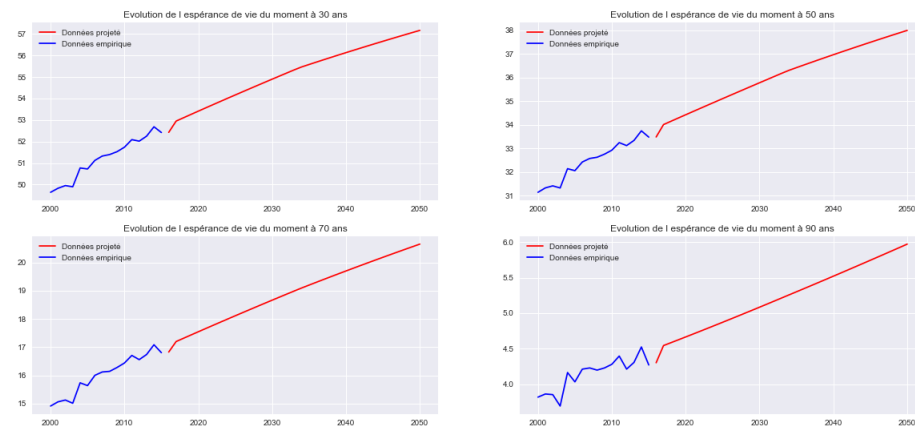


FIGURE 2.11 – Espérances de vie du moment empirique en bleu et obtenues à partir des coefficients projetées en rouge à 30, 50, 70 et 90 ans.

La figure 2.12 présente les taux d'amélioration obtenus avec cette projection. Ces taux d'amélioration sont en continuité avec les données observées. Ils contiennent des tendances qui n'étaient pas extrapolées avec le modèle de Lee & Carter (1992). Néanmoins cette projection ne fait pas apparaître d'effet période. En effet, les coefficients extrapolés de façon linéaire ne subissent aucun choc.

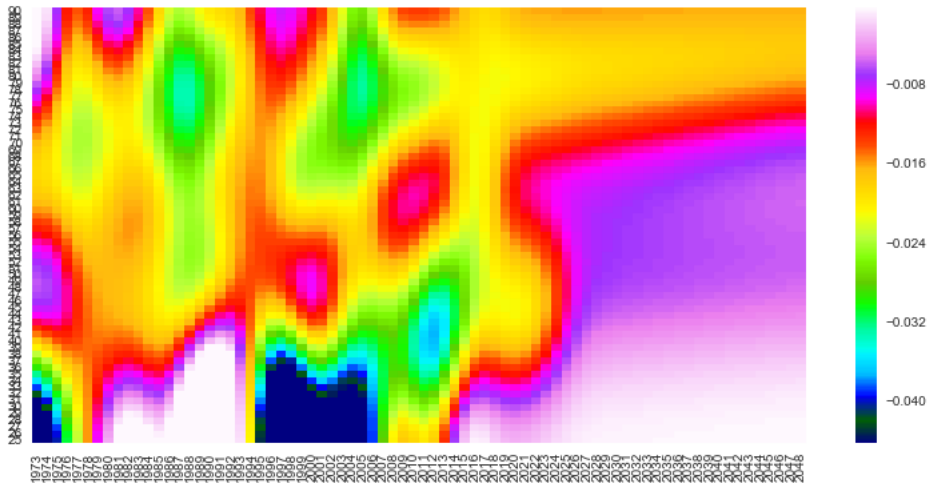


FIGURE 2.12 – Taux d’amélioration lissés ajustés (avant 2015) et projetés manuellement (à partir de 2015).

### 2.5.5 Critiques de cette projection

Cette projection est basée sur plusieurs hypothèses, notamment concernant la projection de tendances de façon linéaires. Nous enlevons complètement la saisonnalité des coefficients de la composante *Mid*. Or Wilmoth et al. (1989) soulèvent l’apparition d’une autre cohorte née dans les années 50 à la mortalité plus importante que ses voisines. Celle-ci n’a pas vécu de guerre sur le territoire français mais est marquée par plusieurs facteurs :

- Une généralisation de l’accouchement en milieu hospitalier. Il a permis de réduire la mortalité infantile mais les interventions obstétricales ont laissé des séquelles aux survivants.
- Une hausse des accidents pendant les années 1950 et 1960. Ces derniers peuvent en effet laisser des séquelles sur les victimes.
- Une hausse des dégénérescences.

Les premiers individus de cette cohorte ont eu 65 ans en 2015. Le coefficient  $m$  ne les a donc pas encore captés. Nous observons deux cohortes dans les données passées, nous suspectons l’arrivée d’une troisième cohorte très prochainement. Les causes de chacune de ces cohortes étant différentes, il est difficile de savoir si cette alternance de comportement de mortalité entre génération va se poursuivre ou non. L’arrivée et l’anticipation d’une cohorte aux âges intermédiaires critiques relève d’un jugement d’expert en fonction des évènements survenus et des données disponible sur la population jeune. Ce travail de caractérisation des cohortes aux âges jeunes à des fins d’anticipations de la mortalité aux âges intermédiaires peut faire l’objet de futurs travaux.

Néanmoins cette projection permet d’avoir une référence ancrée sur des tendances passées. Il est possible d’en interpréter chaque composante. Cette transparence du modèle permet d’en adapter tous les paramètres aux spécificités de chaque population assurée.

De plus il est possible d’établir de façon explicite certains scénarios stressés cohérents. Certaines interactions entre coefficients sont maîtrisés car rencontrés dans certaines circonstances passées.



Il est par exemple possible de simuler le passage d'une cohorte à la mortalité plus importante ou encore un comportement spécifique de la mortalité des personnes âgées. Il faut néanmoins prendre des précautions en réalisant de tels chocs. Toutes les situations de stress imaginables n'ont pas été rencontrées dans les données. Dès lors que l'on veut choquer un paramètre, dans de nombreux cas nous ne savons pas comment ajuster les autres. Les chocs des paramètres du modèle Logit+ additif sont donc une piste de recherche pour de futurs travaux.

## 2.6 Conclusion

Nous disposons maintenant des séries temporelles de chaque coefficient du modèle *Logit+Additif*. En les analysant nous avons établi une projection de référence. Cette projection se veut relativement simpliste pour être en mesure de maîtriser et de comprendre chaque composante mais elle n'est pas exhaustive. Elle ne prend pas en compte les dépendances fines entre tous les paramètres. La prochaine partie détaille un ensemble de méthodes plus avancées pour projeter les coefficients dans le temps.

## Chapitre 3

# Projection temporelle de la mortalité par réseaux RVFL

Bongaarts (2004) a utilisé des modèles ARIMA pour projeter dans le temps les deux paramètres de son modèle. Avec son approche l'espérance de vie des femmes devient significativement plus faible que celle des hommes au bout d'une vingtaine d'année. Or si la différence d'espérance de vie entre les sexes tend à se réduire, nous n'avons pour le moment aucune raison de penser que l'espérance de vie des femmes va devenir inférieure à celle des hommes. Les modèles ARIMA ne reflètent pas les dépendances entre les paramètres qui, comme nous l'avons montré en partie 2, sont bien présentes entre les coefficients. L'approche décrite dans cette partie consiste à utiliser des modèles pour lesquels la prédiction d'un paramètre dépend des valeurs passées de celui-ci, ainsi que des valeurs passées des autres paramètres.

Le problème se formule de la manière suivante :

- Soit  $p$  le nombre de séries à extrapoler, dans notre cas  $p = 6$ .
- $\forall j \in \llbracket 1, p \rrbracket$ ,  $(X_t^{(j)})_{t \geq 0}$  sont les  $p$  séries temporelles observées à  $n$  dates discrètes.
- Nous voulons prédire ces séries jusqu'au temps  $n + h$ ,  $h \in \mathbb{N}^*$ . Pour simplifier les modèles, nous effectuerons  $h$  prédictions successives de pas de temps de taille 1.
- Soit  $k$  l'ordre du modèle considéré, c'est à dire le nombre de périodes d'observation utilisées pour faire une prédiction.

Le problème de régression que nous voulons à résoudre est alors le suivant, nous cherchons une fonction  $f$  tel qu'au temps  $i$  :

$$\begin{bmatrix} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{bmatrix} = f \left( \begin{bmatrix} X_{i-k}^{(1)} & \dots & X_{i-1}^{(1)} \\ \vdots & \ddots & \vdots \\ X_{i-k}^{(p)} & \dots & X_{i-1}^{(p)} \end{bmatrix} \right)$$

L'objet de cette partie est alors de trouver une fonction  $f$  et une valeur de  $k$  de manière à obtenir un modèle au pouvoir prédictif satisfaisant, sans pour autant sur-apprendre les données d'entraînement. Nous utiliserons pour cela des méthodes de régularisation qu'il faudra ajuster.

Un des avantages de cette formalisation est qu'il est possible d'intégrer des variables extérieures aux prédicteurs afin d'affiner le résultat. On peut imaginer utiliser des données concernant par exemple le nombre de personnes atteintes de certaines maladies. Cette méthodologie s'inspire de Planchet et al. (2018) et Qiu et al. (2016) qui ont projeté dans le temps les coefficients d'une forme paramétrique ajustant respectivement une courbe de taux et la demande en puissance électrique de systèmes d'alimentation modernes.

### 3.1 Protocole expérimental

Nous avons à disposition les séries des coefficients d’ajustement de la mortalité de la population française par le modèle *Logit+ additif* de 1955 à 2015. Il faut séparer cet ensemble en deux sous-ensembles, l’un pour entraîner les différents modèles de projection et l’autre pour les tester et les comparer entre eux. L’idéal serait d’avoir un troisième ensemble de validation non utilisé pendant les phases d’entraînement et de tests afin de valider la qualité de la méthode retenue. En effet, si l’on compare de nombreux modèles sur un même ensemble de test, le modèle retenu sera influencé par cet ensemble et en aura appris certaines caractéristiques.

Mais nous n’observons que 61 points en dimension 6. Les modèles utilisés auront plusieurs dizaines de paramètres à calibrer. La solution la plus simple serait d’augmenter le nombre d’observations. Nous avons plusieurs solutions pour ce faire :

- Nous pouvons augmenter la taille de la plage d’observation et prendre en compte des données plus anciennes. Mais la mortalité avant les années 50 est perturbée par les guerres. Les tendances et les dynamiques ne sont pas comparables avec ce que l’on observe actuellement. Les modèles risqueraient d’apprendre des comportements non applicables à la population présente.
- Il est possible d’agréger les données de plusieurs pays dont les dynamiques de mortalité sont similaires pour calibrer un même modèle. Mais chaque pays a son histoire démographique propre, ce qui rend difficile la tâche de trouver un modèle simple applicable à tous.

Nous nous passerons d’ensemble de validation pour maximiser la taille de l’échantillon d’entraînement. La validation des modèles se fera alors par analyse de cohérence de l’espérance de vie du moment et des taux d’amélioration projetés.

Les modèles testés possèdent des hyperparamètres à fixer a priori. Ils peuvent correspondre à des coefficients de régularisation ou caractériser l’architecture des modèles. Ces paramètres sont habituellement choisis par validation croisée. Comme nous travaillons sur des séries temporelles, il n’est pas possible d’appliquer les méthodes de type ”k-fold cross-validation” ou ”leave-one-out cross-validation” qui ne conservent pas la succession temporelle entre les ensembles d’entraînement et de test (Arlot 2016). Il serait théoriquement possible d’utiliser une fenêtre glissante pour multiplier les couples entraînement/validation, mais encore une fois il est impossible de faire plusieurs échantillons d’entraînement avec seulement 61 points pour calibrer des modèles avec plus d’une dizaine de paramètres.

Le protocole utilisé pour chaque modèle est alors le suivant :

- Les séries sont centrées et réduites.
- Les modèles sont calibrés à l’aide des points obtenus strictement avant 2006.
- Les points entre 2006 et 2015 servent de test afin de choisir les hyperparamètres. Nous mettons en place une recherche par grille : nous entraînerons  $r$  fois les modèles sur l’intervalle [1955, 2005] pour  $r$  jeux d’hyperparamètres différents, nous retenons le jeu ayant l’erreur la plus faible sur l’intervalle [2006, 2015]. Soient  $(q_x^y)_{x \in [25, 95], y \in [2006, 2015]}$  et  $(\hat{q}_x^y)_{x \in [25, 95], y \in [2006, 2015]}$  les probabilités observées et estimées entre 25 et 95 ans des années 2006 à 2015, l’erreur

minimisée est alors :

$$e = \sum_{a=25}^{95} \frac{1}{2015 - 2006 + 1} \sum_{y=2006}^{2015} (\ln(q_x^y) - \ln(\hat{q}_x^y))^2.$$

- Nous sommes obligés de prendre un échantillon de test relativement petit. En effet, la bosse sur la mortalité autour de 50 ans décrite en partie 1 apparaît au début des années 2000. Si nous l'excluons de l'ensemble d'entraînement, le modèle ne la fera pas apparaître. Dès lors si la prédiction du modèle ne paraît pas assez stable, nous nous réservons le droit d'augmenter la régularisation des modèles pour lisser les projections.
- Une fois les hyperparamètres choisis, nous entraînons le modèle sur toutes les données pour projeter la mortalité au delà de 2015.

Cette méthodologie est appliquée à plusieurs modèles décrits ci-dessous.

## 3.2 Projection par Vecteur Autoregressif

Nous commençons par lier les entrées et les sorties par un modèle linéaire afin d'obtenir une projection que les modèles plus complexes devront améliorer : le modèle Vecteur Autoregressif (VAR) (Lütkepohl 2005).

### 3.2.1 Spécification du modèle

Ce modèle est une généralisation des processus autorégressifs en plusieurs dimensions. Sa spécification à l'ordre  $k$  est la suivante :

$$\forall i \in \llbracket k, n \rrbracket, \begin{bmatrix} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{bmatrix} = \sum_{t=1}^k A_t \begin{bmatrix} X_{i-t}^{(1)} \\ \vdots \\ X_{i-t}^{(p)} \end{bmatrix} + b + e_i$$

Avec :

- $(A_t)_{t \in \llbracket 1, k \rrbracket}$ , avec  $A_t \in \mathbb{R}^{p \times p}$  les matrices des coefficients,
- $b \in \mathbb{R}^{p \times 1}$  le vecteur des biais,
- $e_i \in \mathbb{R}^{p \times 1}$  le vecteur des erreurs.

Il y a  $k \times p \times p + p$  paramètres à calibrer. Nous utilisons pour cela l'implémentation Python de ce modèle issue du package StatsModels\*. Nous commençons par calibrer un modèle VAR(1), c'est à dire avec  $k = 1$ .

### 3.2.2 Présentation des résultats obtenus par VAR(1)

Les détails du modèle sont visibles en annexe B. La figure 3.1 montrent les résidus obtenus en entraînant ce modèle sur toutes les années disponibles. Les résidus affichés sont calculés avec la formule  $\epsilon_{x,t} = \ln(q_x(t)) - \ln(\hat{q}_x(t))$ . Contrairement aux modèles de Lee & Carter (1992), APC et LCC développés au chapitre 1, on observe l'apparition de structures verticales dans les résidus. Le modèle apprend les coefficients d'une forme paramétrique reproduisant les forces de mortalité. Une erreur dans un coefficient se traduira alors par une mauvaise estimation des forces de mortalité à plusieurs âges proches. Nous n'observons pas les structures diagonales comme avec le modèle de Lee & Carter (1992). Ce modèle approche donc la mortalité apprise de manière correcte.

Les figures 3.2 et 3.3 montrent que les projections des coefficients et des taux de mortalité sont cohérentes avec les valeurs passées. Nous n'observons pas de transition brutale dans les taux de mortalité comme avec les modèles présentés au chapitre 1. Ici ces derniers sont lissés avec un noyau

---

\*. <http://www.statsmodels.org/stable/index.html>

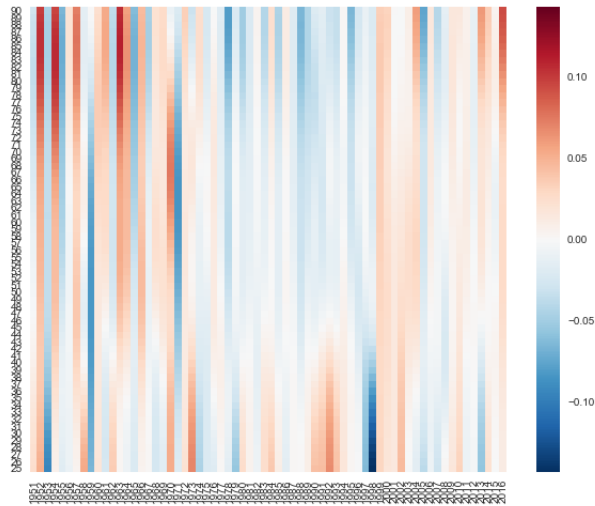


FIGURE 3.1 – VAR(1) : résidus issus de l'apprentissage sur les données de 1951 à 2016 entre 25 et 95 ans.

gaussien de variance 2. Des méthodes de lissage plus efficaces existent (méthodes de vraisemblance locale (Tibshirani & Hastie 1987) ou encore méthodes adaptatives de vraisemblance locale (Thomas & Planchet 2013)), mais n'ont pas été implémentées dans le cadre de ces travaux.

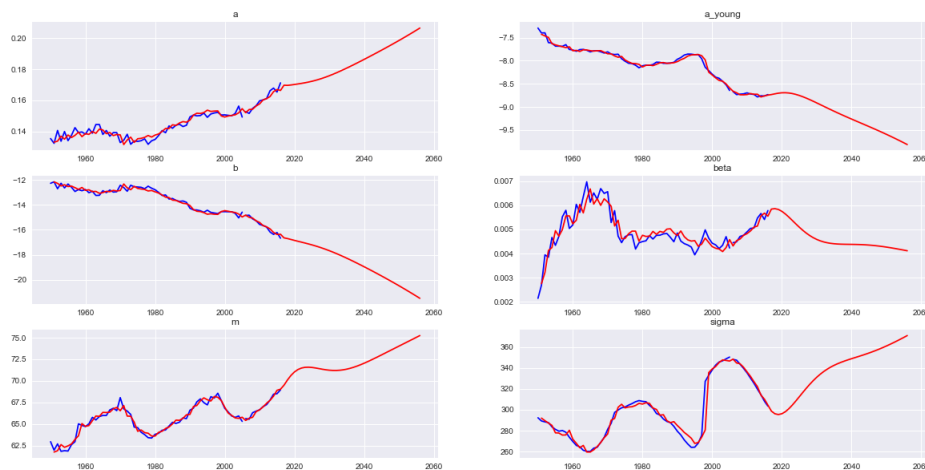


FIGURE 3.2 – Séries projetées par VAR(1) (ensembles d'entraînement en bleu) et prédictions (en rouge) entre 1955 et 2015.

Lorsque l'on entraîne le modèle sur toutes les données disponibles, l'espérance de vie, visible sur la figure 3.4, est projetée de façon linéaire à chaque âge. Le modèle anticipe une augmentation de l'espérance de vie plus rapide que par le passé principalement aux grands âges (la pente de la courbe est plus importante). Ces tendances, bien que plausibles, ne paraissent pas cohérentes



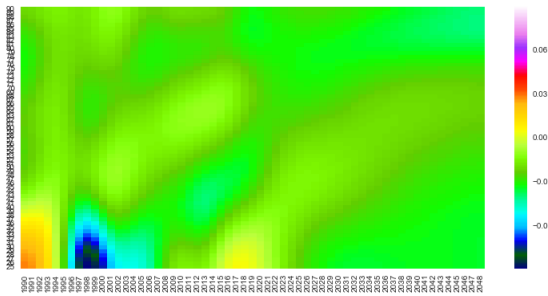


FIGURE 3.3 – Taux d'amélioration lissés calibrés (avant 2015) et projetés (à partir de 2015) par VAR(1).

avec la tendance actuelle des pays développés pour lesquels l'augmentation de l'espérance de vie tend à ralentir (Denney et al. 2013). La figure 3.2 montre que le modèle ajuste les données de très près. Par exemple les coefficients  $\beta$ ,  $b$  et  $a$  ne sont pas lissés par la prédiction. Cela peut être un signe de sur-apprentissage. Nous ajustons en effet 61 points de dimension 6 avec un modèle à  $1 \times 6 \times 6 + 6 = 42$  paramètres sans régularisation. Il ne faut donc pas accorder trop de crédit à la projection réalisée par ce modèle.

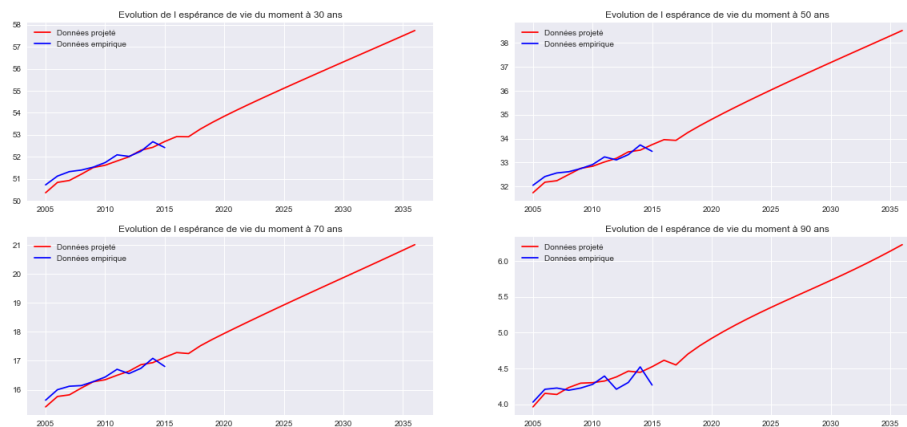


FIGURE 3.4 – Espérances de vie du moment empirique (en bleu) et projetée (en rouge) par VAR(1) à 30, 50, 70 et 90 ans.

L'objectif de la suite de ce chapitre est donc de trouver un modèle étant capable de comprendre de manière plus fine les dynamiques de mortalité à l'aide des données disponibles pour effectuer des prédictions plus cohérentes. Afin d'améliorer ce modèle, nous introduisons une couche intermédiaire. Elle permet de prendre en compte d'éventuelles relations non linéaires entre les prédicteurs. C'est le modèle RVFL présenté ci-dessous.

### 3.3 Projection par réseaux *RVFL*

Pour améliorer le modèle VAR, nous utilisons d'autres modèles plus complexes. Le premier est un modèle RVFL (*Random Vector Functional Link*). La méthodologie employée est alors la même que dans Planchet et al. (2018). Dans cette publication les auteurs appliquent ce modèle aux coefficients d'une forme paramétrique caractérisant une courbe de taux afin de la projeter dans le temps.

#### 3.3.1 Présentation des modèles *RVFL*

Ces modèles sont des réseaux de neurones exprimant les sorties comme une relation linéaire entre les entrées et une transformation non linéaire de celles-ci. Un de leurs avantages est le faible nombre de paramètres à calibrer en comparaison avec un réseau de neurones classique. En effet, les poids correspondant à la partie non linéaire sont fixés à l'aide d'une séquence quasi aléatoire. Dans notre cas nous utiliserons une séquence de Sobol, cela dans le but de couvrir l'espace de façon homogène.

#### Spécification du modèle

Mathématiquement le modèle se définit de la manière suivante :

- Soit  $p$  le nombre de séries à projeter,
- Soit  $k$  l'ordre du modèle, c'est à dire le nombre de pas de temps utilisés pour une prédiction,
- Soit  $L$  le nombre de neurones cachés.
- Soit  $\Phi(x)$  une fonction non linéaire de  $x$ .

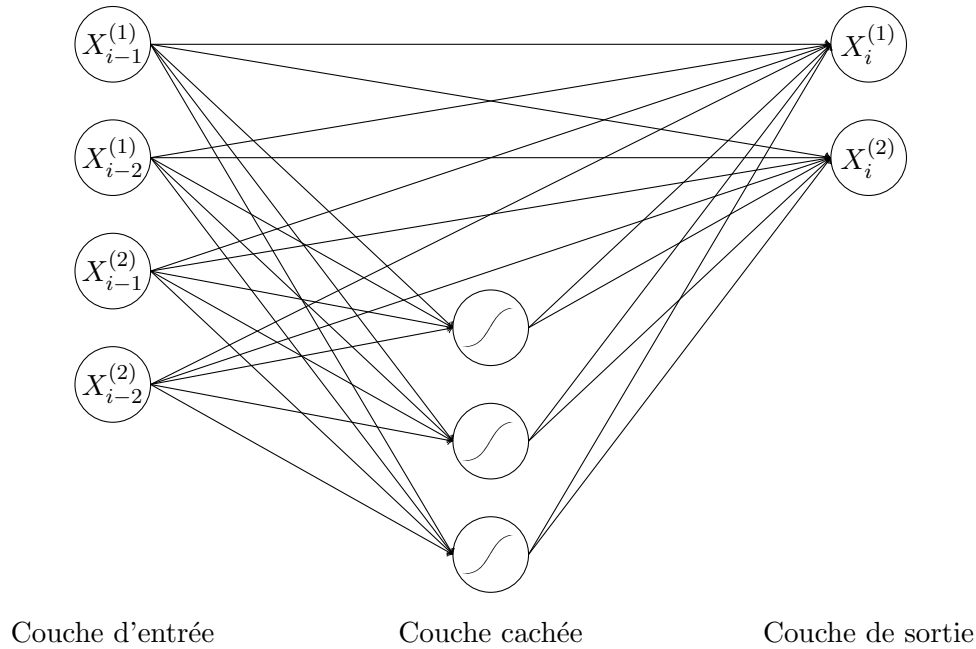
On a alors :

$$\forall i \in \llbracket k, n \rrbracket, \forall j \in \llbracket 1, p \rrbracket, X_i^{(j)} = \beta_0 + \sum_{t=1}^k \sum_{v=1}^p \beta_{t,v} X_{i-t}^{(v)} + \sum_{l=1}^L \gamma_l \Phi \left( \sum_{t=1}^k \sum_{v=1}^p W_{t,v}^l X_{i-t}^{(v)} \right) + \epsilon_i^j.$$

Le schéma 3.5 illustre cette équation avec les valeurs  $p = 2$ ,  $k = 2$  et  $L = 3$ . Chaque cercle de la couche cachée et de la couche de sortie représente un perceptron : les sorties de chacun d'entre eux sont les sommes pondérées des entrées, passées dans une non linéarité pour la couche cachée. Nous utilisons la fonction tangente hyperbolique :  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , qui est bijective de  $\mathbb{R}$  dans  $[-1, 1]$ , comme fonction non linéaire dans la couche cachée. Ce choix est justifié par le fait que l'on centre et réduit les données.

Dans ce modèle il y a trois matrices de poids différentes :

- $W = (w_{i,j})_{1 \leq i \leq k \cdot p, 1 \leq j \leq L}$  représente les poids permettant de passer de la couche d'entrée à la couche cachée.

FIGURE 3.5 – Schéma d'illustration d'un réseau RVFL avec  $p = 2$ ,  $k = 2$  et  $L = 3$ .

- $\beta = (\beta_{i,j})_{1 \leq i \leq k * p, 1 \leq j \leq p}$  représente les poids permettant de passer de la couche d'entrée à la couche de sortie.
- $\gamma = (\gamma_{i,j})_{1 \leq i \leq L, 1 \leq j \leq p}$  représente les poids permettant de passer de la couche cachée à la couche de sortie.

Les éléments de  $W$  sont tirés d'une séquence de Sobol déterministe quasi aléatoire sur  $[0, 1]$  qui est transformée dans  $[-1, 1]$ . L'idée est d'avoir des points plus ou moins équidistants dans  $[-1, 1]^{k * p}$ . Ces poids sont tirés à la création du modèle. Il ne sont ensuite pas modifiés pour effectuer des prédictions. Des traitements plus complets de ces séquences peuvent être trouvés dans Niederreiter (1992), Boyle & Tan (1997), Joe & Kuo (2008) et Planchet et al. (2018). Dans notre cas nous nous limiterons à une seule couche cachée. Nous utiliserons l'implémentation Python de Chisari (2016) pour générer ces nombres. Les résultats pourront être influencés par l'initialisation de la séquence. Dans nos simulations nous nous sommes assurés que l'initialisation de celle-ci n'avait que peu d'influence sur les projections réalisées dès lors que le nombre de neurones cachés était assez important.

La couche cachée aide à capturer des relations non linéaires entre les séries. Ces modèles ont été appliqués avec succès à différents problèmes de classification et de régression (Dehuri & Cho 2010). Ils ont aussi été utilisés pour projeter des séries temporelles (Rena et al. (2016) et Planchet et al. (2018)).

Le problème est alors de trouver les coefficients des  $p$  régressions linéaires liants les  $p$  sorties avec les  $k * p$  entrées et les  $L$  transformations non linéaires de celles-ci.

### Calibrage des poids du modèle

On rappelle que l'on dispose de  $p$  séries  $(X_t^{(j)})_{1 \leq j \leq p, 1 \leq t \leq n}$  à projeter dans le temps. On pose  $\mathbf{W} \in \mathbb{R}^{(k \times p) \times L}$  la matrice des coefficients de la couche cachée. Soit  $\mathbf{Y} \in \mathbb{R}^{(n-k) \times p}$  la matrice des réponses :

$$\mathbf{Y} = \begin{bmatrix} X_n^{(1)} & \dots & X_n^{(p)} \\ \vdots & \ddots & \vdots \\ X_{n-k+1}^{(1)} & \dots & X_{n-k+1}^{(p)} \end{bmatrix}.$$

Soit  $\mathbf{X} \in \mathbb{R}^{(n-k) \times (k \times p)}$  la matrice des prédicteurs :

$$\mathbf{X} = \begin{bmatrix} X_{n-1}^{(1)} & \dots & X_{n-k}^{(1)} & \dots & X_{n-1}^{(p)} & \dots & X_{n-k}^{(p)} \\ \vdots & & \vdots & & \vdots & & \vdots \\ X_k^{(1)} & \dots & X_1^{(1)} & \dots & X_k^{(p)} & \dots & X_1^{(p)} \end{bmatrix}.$$

On a par exemple pour  $p = 2$  séries temporelles observées à  $n = 5$  dates  $t_1 < \dots < t_5$ , avec un modèle d'ordre  $k = 2$  possédant  $L = 3$  neurones cachés :

$$\mathbf{Y} = \begin{bmatrix} X_{t_5}^{(1)} & X_{t_5}^{(2)} \\ X_{t_4}^{(1)} & X_{t_4}^{(2)} \\ X_{t_3}^{(1)} & X_{t_3}^{(2)} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} X_{t_4}^{(1)} & X_{t_3}^{(1)} & X_{t_4}^{(2)} & X_{t_3}^{(2)} \\ X_{t_3}^{(1)} & X_{t_2}^{(1)} & X_{t_3}^{(2)} & X_{t_2}^{(2)} \\ X_{t_2}^{(1)} & X_{t_1}^{(1)} & X_{t_2}^{(2)} & X_{t_1}^{(2)} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} W_{(1,1)} & W_{(1,2)} & W_{(1,3)} \\ W_{(2,1)} & W_{(2,2)} & W_{(2,3)} \\ W_{(3,1)} & W_{(3,2)} & W_{(3,3)} \\ W_{(4,1)} & W_{(4,2)} & W_{(4,3)} \end{bmatrix}.$$

Nous cherchons les matrices  $\hat{\beta}$  et  $\hat{\gamma}$  qui minimisent la somme pénalisée du carré des erreurs. Pour cela notons  $y$  la  $j^{\text{ème}}$  colonne ( $1 \leq j \leq p$ ) de la matrice des réponses  $\mathbf{Y}$  et  $\Phi(\mathbf{X})$  la matrice des prédicteurs  $\mathbf{X}$  transformés par la couche cachée. On note aussi  $\beta_m^{(j)} = \beta_m$  et  $\gamma_l^{(j)} = \gamma_l$  l'ensemble des paramètres de la régression associée à la  $j^{\text{ème}}$  série temporelle pour  $1 \leq m \leq k \times p$  et  $1 \leq l \leq L$ . Nous utilisons une pénalisation  $L^2$  des poids. Les contraintes de pénalisation s'expriment alors de la manière suivante :

$$\sum_{m=1}^{k \times p} \beta_m^2 \leq u, \quad (3.1)$$

$$\sum_{l=1}^L \gamma_l^2 \leq v. \quad (3.2)$$

Avec  $u \in \mathbb{R}$  et  $v \in \mathbb{R}$ . Ainsi pour les vecteurs  $\beta$  et  $\gamma$  contenant les paramètres de la régression, nous obtenons le Lagrangien suivant :

$$\mathcal{L}(\mathbf{X}, \beta, \gamma) = (y - \mathbf{X}\beta - \Phi(\mathbf{X})\gamma)^t (y - \mathbf{X}\beta - \Phi(\mathbf{X})\gamma) + \lambda_1 \beta^t \beta + \lambda_2 \gamma^t \gamma.$$

Avec  $\lambda_1$  et  $\lambda_2$  les multiplicateurs de Lagrange associés aux contraintes 3.1 et 3.2. Dériver cette expression par rapport à  $\beta$  et  $\gamma$  nous amène à :

$$\begin{aligned}\frac{\partial \mathcal{L}(\mathbf{X}, \beta, \gamma)}{\partial \beta} &= -y\mathbf{X} - \mathbf{X}^t y + 2(\mathbf{X}^t \mathbf{X})\beta + \mathbf{X}^t \Phi(\mathbf{X})\gamma + (\Phi(\mathbf{X})\gamma)^t \mathbf{X} + 2\lambda_1 \beta \\ &= 2(\mathbf{X}^t \mathbf{X} + \lambda_1 I_{k \times p})\beta - y^t \mathbf{X} - \mathbf{X}^t y + \mathbf{X}^t \Phi(\mathbf{X})\gamma + (\Phi(\mathbf{X})\gamma)^t \mathbf{X} \\ &= 2(\mathbf{X}^t \mathbf{X} + \lambda_1 I_{k \times p})\beta - 2\mathbf{X}^t y + 2\mathbf{X}^t \Phi(\mathbf{X})\gamma. \\ \frac{\partial \mathcal{L}(\mathbf{X}, \beta, \gamma)}{\partial \gamma} &= 2(\Phi(\mathbf{X})^t \Phi(\mathbf{X}) + \lambda_2 I_L)\gamma - 2\Phi(\mathbf{X})^t y + 2\Phi(\mathbf{X})^t \mathbf{X}\beta.\end{aligned}$$

Avec  $I_a$  la matrice identité de dimension  $a \times a$ . En fixant ces dérivés à 0 nous obtenons :

$$\begin{cases} (\mathbf{X}^t \mathbf{X} + \lambda_1 I_{k \times p})\beta + \mathbf{X}^t \Phi(\mathbf{X})\gamma = \mathbf{X}^t y \\ (\Phi(\mathbf{X})^t \Phi(\mathbf{X}) + \lambda_2 I_L)\gamma + \Phi(\mathbf{X})^t \mathbf{X}\beta = \Phi(\mathbf{X})^t y \end{cases}$$

Sous forme matricielle nous obtenons :

$$\begin{bmatrix} \mathbf{X}^t \mathbf{X} + \lambda_1 I_{k \times p} & \mathbf{X}^t \Phi(\mathbf{X}) \\ \Phi(\mathbf{X})^t \mathbf{X} & \Phi(\mathbf{X})^t \Phi(\mathbf{X}) + \lambda_2 I_L \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \mathbf{X}^t y \\ \Phi(\mathbf{X})^t y \end{bmatrix}. \quad (3.3)$$

Posons :

$$\begin{aligned}A &= \begin{bmatrix} \mathbf{X}^t \mathbf{X} + \lambda_1 I_{k \times p} & \mathbf{X}^t \Phi(\mathbf{X}) \\ \Phi(\mathbf{X})^t \mathbf{X} & \Phi(\mathbf{X})^t \Phi(\mathbf{X}) + \lambda_2 I_L \end{bmatrix} = \begin{bmatrix} B & C^t \\ C & D \end{bmatrix}, \\ S &= D - CB^+C^t.\end{aligned}$$

Avec la notation  $M^+$  matrice pseudo-inverse de Moore-Penrose de  $M$  (Penrose 1955). Alors en utilisant l'algorithme d'inversion d'une matrice par bloc décrit dans Cormen (2009) nous obtenons :

$$A^+ = \begin{bmatrix} B^+ + B^+C^tS^+CB^+ & -B^+C^tS^+ \\ -S^+CB^+ & S^+ \end{bmatrix} = \begin{bmatrix} A_1^+ & A_2^+ \\ A_3^+ & A_4^+ \end{bmatrix}. \quad (3.4)$$

À partir de l'équation 3.3 et de l'expression 3.4 on obtient alors les solutions suivantes :

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{bmatrix} = \begin{bmatrix} A_1^+ & A_2^+ \\ A_3^+ & A_4^+ \end{bmatrix} \begin{bmatrix} \mathbf{X}^t y \\ \Phi(\mathbf{X})^t y \end{bmatrix}.$$

L'ensemble des paramètres pour les  $p$  séries observées est donné par :

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{bmatrix} = \begin{bmatrix} A_1^+ & A_2^+ \\ A_3^+ & A_4^+ \end{bmatrix} \begin{bmatrix} \mathbf{X}^t \mathbf{Y} \\ \Phi(\mathbf{X})^t \mathbf{Y} \end{bmatrix}.$$

L'implémentation informatique de cette solution est réalisée avec la bibliothèque de calcul scientifique Numpy\* en Python qui permet de manipuler facilement les matrices.

### 3.3.2 Projection par *RVFL(1)*

Nous commençons par utiliser un réseau RVFL d'ordre 1. Les hyperparamètres trouvés par le protocole décrit au début du chapitre sont :  $L = 20$ ,  $\lambda_1 = 10^{-8}$  et  $\lambda_2 = 0.6$ . Les résidus de

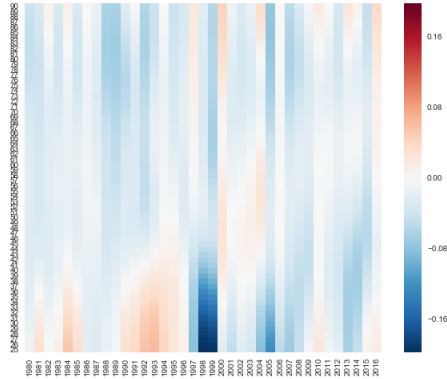


FIGURE 3.6 – Résidus issus de l’ajustement d’un modèle RVFL(1) sur la population française entre 1951 et 2016.

l’ajustement sont présentés sur la figure 3.6. Nous n’observons pas de tendance claire dans ces résidus signe que l’ajustement est de qualité sur les données d’entraînement.

Les espérances de vie du moment projetées sont visibles en figure 3.7. Ces projections montrent un changement de tendance brusque au début de la projection, puis un ralentissement de l’augmentation de l’espérance de vie à tous les âges. Ce changement brusque ne paraît pas naturel. Il peut être dû au sur-apprentissage des données d’entraînement. Dans ce cas les tendances projetées ne sont pas fiables. Nous pouvons en effet voir sur la figure 3.8 présentant les taux d’amélioration que bien que leur projection soit dans la continuité des données empiriques, les effets cohortes ne sont pas extrapolés. Nous n’observons pas de projection des structures diagonales dans le temps.

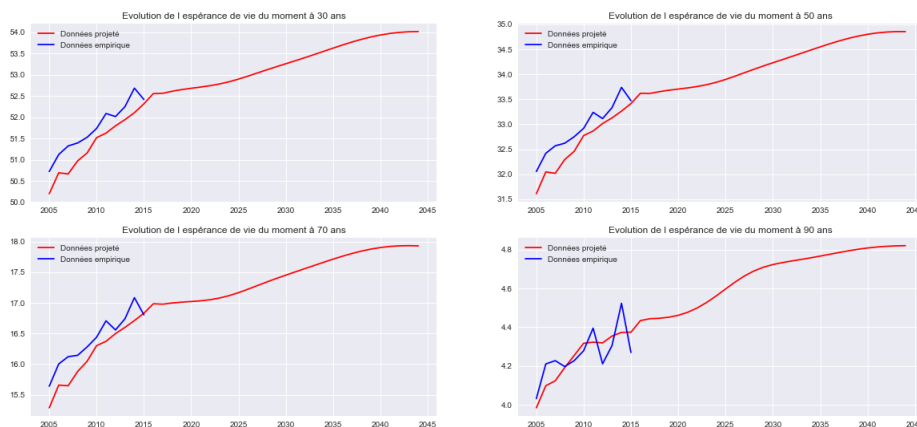


FIGURE 3.7 – Espérances de vie du moment empirique (en bleu) et projetée (en rouge) par RVFL(1) à 30, 50, 70 et 90 ans.

Les séries projetées sont visibles en figure 3.9. Nous remarquons que le modèle a du mal à projeter les coefficients en dehors des valeurs vues dans l’ensemble de test. Il faudrait travailler sur des séries stationnaires afin que nos modèles soient efficaces. Nous allons donc travailler sur les séries différenciées par la suite.

\*. <http://www.numpy.org/>

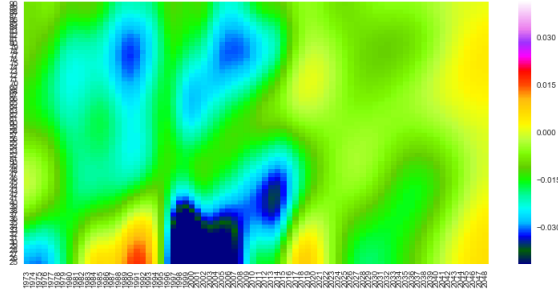


FIGURE 3.8 – Taux d’amélioration lissés calibrés (avant 2015) et projetés (à partir de 2015) par RVFL(1).



FIGURE 3.9 – Séries projetées par RVFL(1) (ensemble d’entraînement en bleu et prédictions en rouge) entre 1955 et 2055.

### 3.3.3 Projection des séries différenciées par *RVFL(1)*

Au vu des graphiques de la figure 2.5, certains coefficients ne sont pas stationnaires. Nous observons des tendances et des saisonnalités. Pour contourner ce problème, nous pouvons différencier les séries afin d’apprendre les accroissements de celles-ci. Les séries obtenues (figure 3.10) ne sont pas toutes stationnaires, mais elles s’en rapprochent. Il faudra faire attention à la variance de la prédiction qui augmentera à cause de l’intégration à réaliser pour reconstruire les séries. Pour cette raison, nous les différencions uniquement à l’ordre 1.

Les paramètres trouvés lors de la recherche par grille sont  $L = 20$ ,  $\lambda_1 = 3$  et  $\lambda_2 = 10^{-4}$ . Les espérances de vie projetées sont présentées sur la figure 3.11. Ces projections linéaires sont dans la continuité des données empiriques. La figure 3.12 des projections des taux d’amélioration montre que les effets cohortes sont pris en compte par le modèle. Nous remarquons de plus que les coefficients projetés peuvent l’être en dehors des valeurs de l’ensemble d’entraînement (figure 3.13), comportement que l’on attendait en différenciant les séries.

Cependant les résidus sont plus importants avant 2005 (figure 3.14). Cela montre que la recherche par grille décrite précédemment a permis au modèle de sur-apprendre les données après

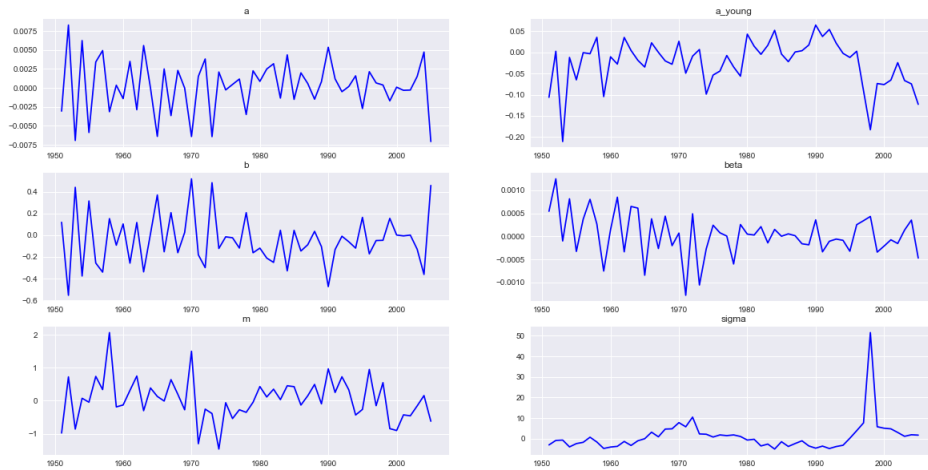


FIGURE 3.10 – Séries différenciées à l'ordre 1 entre 1955 et 2015.

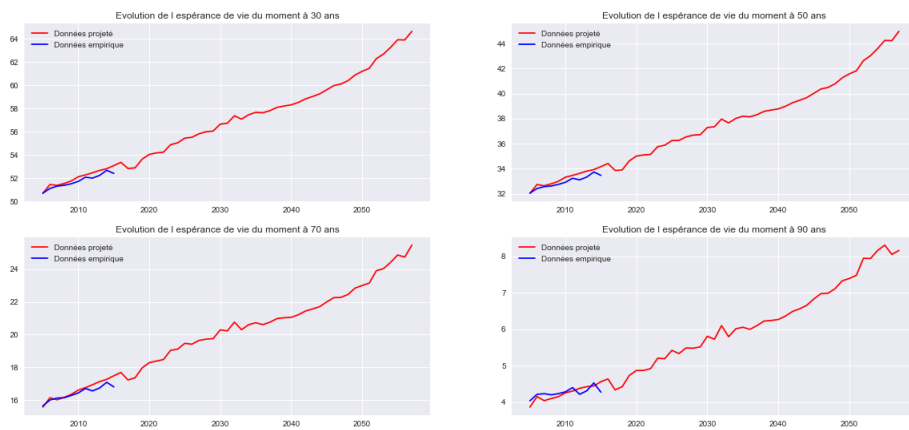


FIGURE 3.11 – Espérances de vie du moment empirique (en bleu) et projetée par RVFL(1) (en rouge) sur les séries différenciées à 30, 50, 70 et 90 ans.

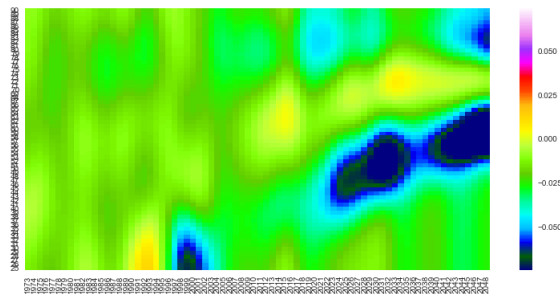


FIGURE 3.12 – Taux d'amélioration lissés calibrés (avant 2015) et projetés (à partir de 2015) par RVFL(1) sur les séries différenciées.

2005. Nous voulons éviter ce comportement pour avoir des prédictions exploitables. Pour ce faire nous pouvons augmenter la régularisation de modèle. On perd alors en qualité d'ajustement, mais les



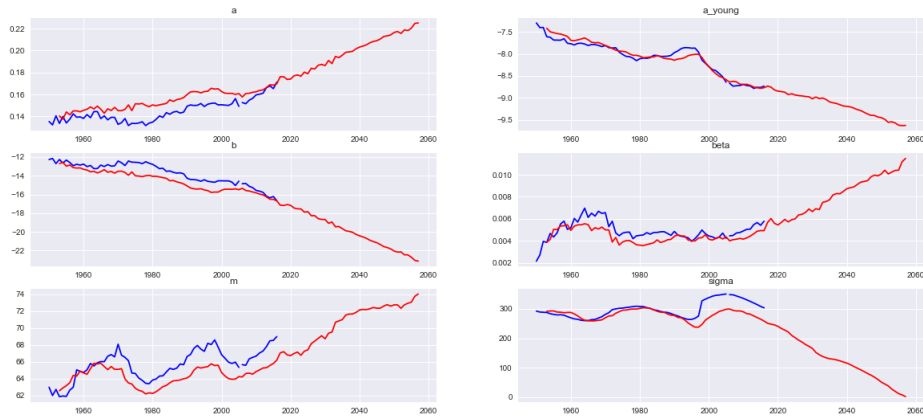


FIGURE 3.13 – Séries projetées par RVFL(1) (ensemble d’entraînement en bleu et prédictions en rouge) entre 1955 et 2055.

résultats sont plus fiables. On obtient par exemple les taux d’amélioration de la figure 3.15. Comme attendu ces derniers sont plus lisses que les précédents. Ils conservent aussi les effets cohortes. Une deuxième solution est de réduire la dimension du problème. Si nous réduisons le nombre de séries à projeter, les matrices des poids  $W$ ,  $\beta$  et  $\gamma$  seront plus petites. Nous explorons cette solution dans le paragraphe suivant.

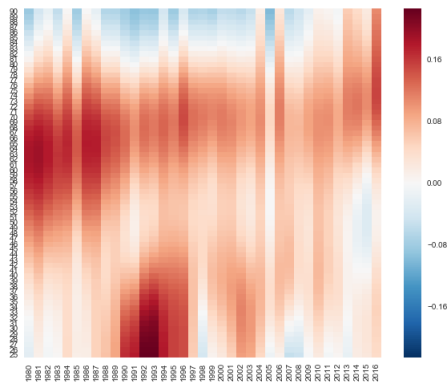


FIGURE 3.14 – Résidus issu de l’ajustement d’un modèle RVFL(1) sur séries différenciées sur la population française entre 1951 et 2016.

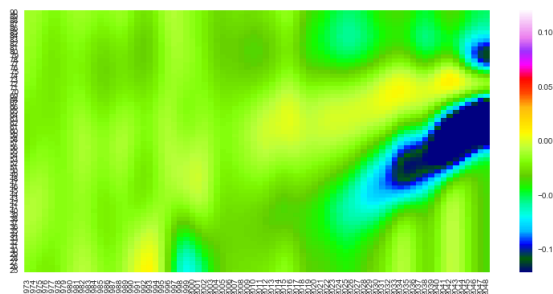


FIGURE 3.15 – Taux d’amélioration lissés calibrés (avant 2015) et projetés (à partir de 2015) par RVFL(1) sur les séries différenciées. Régularisation élevée.

### 3.3.4 Réduction de la dimension par ACP sur les séries différenciées

Afin de réduire la dimension du problème nous appliquons une ACP sur les séries différenciées. Nous projetons ensuite les premières composantes, puis nous reconstruisons les séries.

#### Analyse en Composantes Principales

Nous réalisons une ACP sur les séries différenciées centrées réduites. La figure 3.16 présente les variances cumulées des six composantes. Nous gardons les trois premières composantes qui représentent plus de 80% de la variance. Les séries à projeter sont présentées en figure 3.17. Leurs comportements est proche d'être stationnaire. La figure 3.18 affiche les courbes initiales et les courbes reconstruites. Nous remarquons que nous perdons très peu d'information en conservant uniquement les trois premières composantes de la projection.

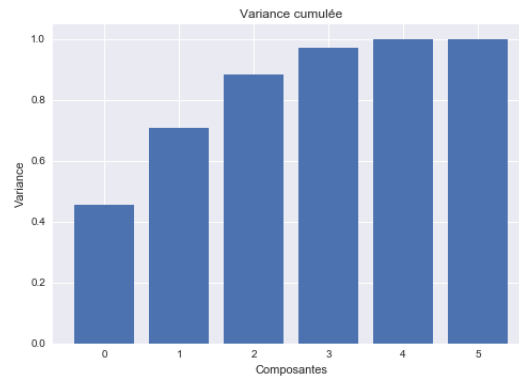


FIGURE 3.16 – Variances cumulées de l'ACP sur séries différenciées.



FIGURE 3.17 – Séries obtenues par projection du nuage sur les trois premiers axes factoriels.



FIGURE 3.18 – Séries initiales (en bleu) et reconstruites (en rouge) par ACP.

## Résultats

Il est possible d'augmenter l'ordre du modèle car nous avons moins de séries, et donc moins de paramètres à calibrer. Les hyperparamètres trouvés sont  $L = 20$ ,  $k = 3$ ,  $\lambda_1 = 0.001$  et  $\lambda_2 = 1$ . Nous obtenons alors les séries projetées visibles sur la figure 3.19. Nous avons alors un modèle capable de prendre en compte les saisonnalités et les tendances de chaque coefficient. Le modèle projette bien l'arrivée de la cohorte des années 1950 à 1960 aux âges intermédiaires où la sur-mortalité peut être plus élevée. Nous observons néanmoins une erreur de niveau assez importante sur la variable  $m$ . Cette erreur entraîne une sous estimation systématique des espérances de vie du moment (figure 3.20).

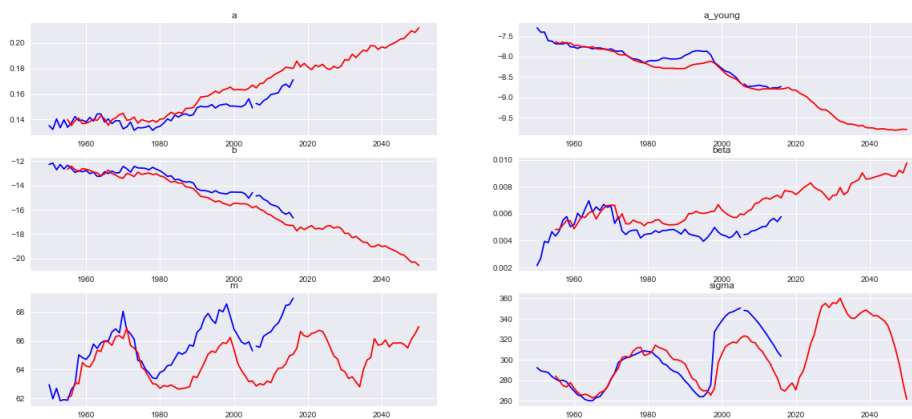


FIGURE 3.19 – Coefficients projetés par RVFL(3) avec ACP (ensembles d'entraînement en bleu) et prédictions (en rouge) entre 1955 et 2015.



FIGURE 3.20 – Espérances de vie du moment empirique et projetée par RVFL(3) sur les séries différenciées avant ACP à 30, 50, 70 et 90 ans.

Il est possible de corriger cette erreur en inversant l'ordre du traitement des données, c'est à dire en commençant par appliquer l'ACP aux séries brutes, puis en différenciant les composantes retenues. Nous obtenons alors les espérances de vie du moment visibles à la figure 3.21. Pour obtenir ces dernières courbes nous ne conservons que les deux premières composantes principales qui expliquent 79% de la variance. Les hyperparamètres retenus sont  $L = 40$ ,  $k = 10$ ,  $\lambda_1 = 0.5$  et  $\lambda_2 = 0.8$ . Les taux de mortalité réels et projetés sont visibles sur la figure 3.22. Les effets cohortes sont projetés et les taux sont en continuité avec les données observées.



FIGURE 3.21 – Espérances de vie du moment empirique et projetée par RVFL(3) sur les séries différenciées après ACP à 30, 50, 70 et 90 ans.

Nous obtenons alors des projections en cohérence avec les tendances passées. Il est possible de reconstruire des tables de mortalité à partir de ces extrapolations.

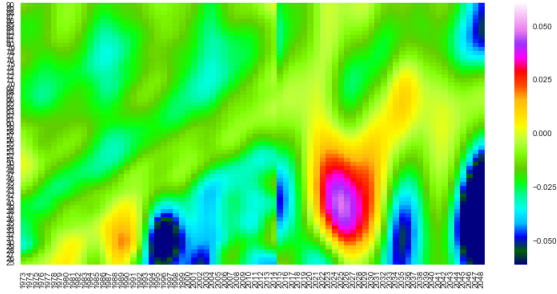


FIGURE 3.22 – Taux d'amélioration lissés réels (avant 2015) et projetés (à partir de 2015) par RVFL(10) sur les séries différenciées après ACP.

### 3.4 Conclusion

Nous remarquons qu'en pratique le calibrage de tel modèle n'est pas évident. Le faible nombre de données nous oblige trouver des astuces tel que la réduction de dimension et la régularisation des modèles. Pour palier en partie à ce problème il est possible d'augmenter l'ensemble d'apprentissage avec les données d'autres pays. Mais il faut pour cela bien en comprendre les dynamiques de mortalité de chacun d'entre eux. Il serait alors nécessaire de refaire l'étude menée en fin de partie 2 pour chaque pays utilisé afin de s'assurer que des phénomènes trop différents ne viennent pas complexifier les modèles.

Nous avons montré que ces modèles projettent des tendances cohérentes. Mais il semble pour le moment difficile d'utiliser ces résultats à des fins de tarification ou de provisionnement de contrat d'assurance vie. Il est en effet délicat de justifier de façon fine les projections des différents paramètres, notamment la périodicité et les niveaux atteints par certains d'entre eux. Ils proposent néanmoins une projection intéressante des taux de mortalité et des espérances de vie du moment plus cohérente que la projection obtenue en partie 2.

# Conclusion

Nous avons développé une nouvelle méthodologie d'analyse de la mortalité par étude des coefficients du modèle Logit+ additif. Cette étude permet de caractériser de manière fine la mortalité à tous les âges adultes. Elle présente une nouvelle approche quant à la façon de définir les effets cohortes qui ne sont pas faciles à capter et à projeter.

Nous avons ensuite présenté une méthodologie complète afin de répondre au problème de la projection à long terme de la mortalité. La principale difficulté a été de calibrer des modèles complexes avec peu de données. Nous obtenons des résultats cohérents avec les tendances passées.

Bien que le démographe puisse se servir de ces travaux afin de projeter la mortalité à l'échelle d'un pays, l'actuaire s'intéressera à des problématiques légèrement différentes :

- Il souhaitera positionner la mortalité de son portefeuille par rapport à une mortalité de référence. Pour cela de futurs travaux peuvent reprendre l'étude menée en partie 2 sur des données plus fines, par exemple par sexe, pour se rapprocher de la mortalité des assurés.
- L'actuaire pourra aussi à l'aide du modèle Logit+ additif construire des scénarios démographiques adaptés à son portefeuille. Des études plus poussées sur les interactions entre les coefficients du modèle seront alors nécessaires. Il serait pour cela intéressant d'étudier les calibrages des coefficients obtenus pour différents pays afin d'avoir plus d'exemples d'interaction.

De futurs travaux peuvent être menés afin de fiabiliser les prédictions obtenues en partie 3. Il est possible d'ajouter des variables explicatives aux modèles décrits dans cette partie, par exemple à propos des conditions sociales et économiques, de données médicales sur différentes maladies etc... Enfin le modèle Logit + additif permet de faire une étude spécifique sur l'apparition et l'anticipation des cohortes aux mortalités particulières.





# Bibliographie

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. (2015), ‘TensorFlow : Large-scale machine learning on heterogeneous systems’. Software available from tensorflow.org.

**URL:** <https://www.tensorflow.org/>

Arlot, S. (2016), ‘Validation croisée’, *Journées d’Études en Statistique (JES)* .

**URL:** <https://arxiv.org/pdf/1703.03167.pdf>

Bongaarts, J. (2004), ‘Long-range trends in adults mortality : Models and projection methods’, *Population Council* .

Boyle, P. P. & Tan, K. S. (1997), ‘Quasi-monte carlo methods’, *International AFIR Colloquium Proceedings* **1**, 1–24.

Chai, C. M. H., Siu, K. & Zhou, X. (2013), ‘A double-exponential garch model for stochastic mortality’, *European Actuarial Journals* **3**(2), 385–406.

Chisari, C. (2016), ‘Sobol the sobol quasirandom sequence’.

**URL:** [http://people.sc.fsu.edu/~jburkardt/py\\_src/sobol/sobol.html](http://people.sc.fsu.edu/~jburkardt/py_src/sobol/sobol.html)

Cormen, T. H. (2009), ‘Introduction to algorithms’, *MIT press* .

Currie, I. (2018), ‘Demography’s dark matter : measuring cohort effects’.

**URL:** <https://www.longevity.co.uk/site/informationmatrix/demographysdarkmatter.measuringcohorteffects.htm>

Debonneuil, E. (2014), Modèle paramétrique de mortalité en fonction de l’âge, pour des applications à des portefeuilles de retraite, Master’s thesis, ISFA.

Dehuri, S. & Cho, S.-B. (2010), ‘A comprehensive survey on functional link neural networks and an adaptive pso–bp learning for cfnn’, *Neural Computing and Applications* **10**(2), 187–205.

Denney, J. T., McNown, R., Rogers, R. G. & Doubilet, S. (2013), ‘Stagnating life expectancies and future prospects in an age of uncertainty’, *US National Library of Medicine, National Institutes of Health* **94**(2), 445–461.

Fiallos, S. & Noel, V. (2018), ‘Modelling old age mortality rates, an extrapolation method based on the mode and the compression of mortality’, *Health Life Pensions* .

- Gao, Q. & Hu, C. (2009), ‘Dynamic mortality factor model with conditional heteroskedasticity’, *Insurance : Mathematics and Economics* **45**(3), 410–423.
- Giacommetti, R., Bertocchi, M., Rachev, S. & Fabozzi, F. J. (2012), ‘A comparison of the lee–carter model and ar-arch model for forecasting mortality rates’, *Insurance : Mathematics and Economics* **50**(1), 85–93.
- Heligman, L. & Pollard, J. H. (1980), ‘The age pattern of mortality’, *Journal of the Institute of Actuaries* **107**(1), 49–82.
- Horiuchi, S. (1983), ‘The long term impact of war on mortality : old-age mortality of first world war survivors in the federal republic of germany’, *Bulletin démographique des Nations Unies* (15), 80–92.
- Hunt, A. & M.Villegas, A. (2015), ‘Robustness and convergence in the lee–carter model with cohort effects’, *Insurance : Mathematics and Economics* **64**, 186–202.
- Joe, S. & Kuo, F. Y. (2008), ‘Notes on generating sobol sequences’.  
**URL:** <http://web.maths.unsw.edu.au/~fkuo/sobol/joe-kuo-notes.pdf>
- John Duchi, E. H. & Singer, Y. (2011), ‘Adaptive subgradient methods for online learning and stochastic optimization’, *J. Mach. Learn. Res.* **12**, 2121–2159.
- Kingma, D. & Ba, J. (2014), ‘Adam : A method for stochastic optimization’.
- Lee, R. D. & Carter, L. (1992), ‘Modeling and forecasting us mortality’, *Journal of the American statistical association* **87**(419), 659–671.
- Lee, R. & Miller, T. (2001), ‘Evaluating the performance of the lee-carter method for forecasting mortality’, *Demography* **38**(4), 537–549.
- Lütkepohl, H. (2005), ‘New introduction to multiple time series analysis’, *Springer Science and Business Media* .  
**URL:** <https://arxiv.org/pdf/1703.03167.pdf>
- Mavros, G., Cairns, A. J., Kleinow, T. & Streftaris, G. (2016), ‘Stochastic mortality modelling : Key drivers and dependent residuals’.
- Nesterov, Y. (1983), ‘A method for unconstrained convex minimization problem with the rate of convergence o  $(1/k^2)$ ’, *Doklady an SSSR* **269**(3), 543–547.
- Niederreiter, H. (1992), ‘Random number generation and quasi-monte carlo methods’, *SIAM* .
- Penrose, R. (1955), ‘A generalized inverse for matrices’, *Mathematical proceedings of the Cambridge philosophical society* **51**, 406–413.
- Planchet, F., Moudiki, T. & Cousin, A. (2018), ‘Multiple time series forecasting using quasi-randomized functional link neural networks’, *Risks* .
- Qian, N. (1999), ‘On the momentum term in gradient descent learning algorithms’, *Neural Networks* **12**(1), 145–151.
- Qiu, X., Suganthan, P. N. & Amaratunga, G. A. J. (2016), ‘Electricity load demand time series forecasting with empirical mode decomposition based random vector functional link network’, *Information Sciences* **367**.

- Rena, Y., P. N. Suganthana, N. S. & Amaratungac, G. (2016), ‘Random vector functional link network for short-term electricity load demand forecasting’, *Information Sciences* **367**, 1078–1093.
- Renshaw, A. & Haberman, S. (2006), ‘A cohort-based extension to the lee–carter model for mortality reduction factors’, *Insurance : Mathematics and Economics* **38**(3), 556–570.
- Shkolnikov, V., Barbieri, M. & Wilmoth, J. (2000), ‘The human mortality database’.  
**URL:** <http://www.mortality.org/>
- Tekian, M.-S. (2018), ‘Espérance de vie : peut-on gagner six heures par jour indéfiniment?’, *theconversation.com* .
- Thatcher, A. R., Kannisto, V. & Vaupel, J. W. (1998), ‘The force of mortality at ages 80 to 120’, *Odense, Denmark : Odense University Press* .
- Thomas, J. (2018), ‘Démographie et tables de mortalité’, *Cours, Université Paris Dauphine* .
- Thomas, J. & Planchet, F. (2013), ‘Multidimensional smoothing by adaptive local kernel-weighted log-likelihood : application to long-term care insurance’, *Insurance : Mathematics Economics* **52**(3), 573–589.
- Tibshirani, R. J. & Hastie, T. J. (1987), ‘Local likelihood estimation’, *Journal of the American Statistical Association* **82**(398), 559–567.
- Vallin, J. & Lopez, A. (1985), ‘La lutte contre la mort : influence des politiques sociales et des politiques de santé sur l’évolution de la mortalité’, *INED* **108**.
- Vallin, J. & Meslé, F. (1988), ‘Les causes de décès en france de 1925 à 1978’, *INED* **115**.
- Willets, R. (2004), ‘The cohort effect : insights and explanations’, *British Actuarial Journal* **10**(4), 833–877.
- Wilmoth, J., Vallin, J. & Caselli, G. (1989), ‘Quand certaines générations ont une mortalité différente de celle que l’on pourrait attendre’, *Population* **44**(2), 335–376.
- Zeiler, M. D. (2012), ‘Adadelata : An adaptive learning rate method’, *CoRR abs* **1212.5701**.
- Zhavoronkov, A., Debonneuil, E., Mirza, N. & Artyuhov, I. (2012), ‘Evaluating the impact of recent advances in biomedical sciences and the possible mortality decreases on the future of health care and social security in the united states’, *Pensions* **17**, 241–251.



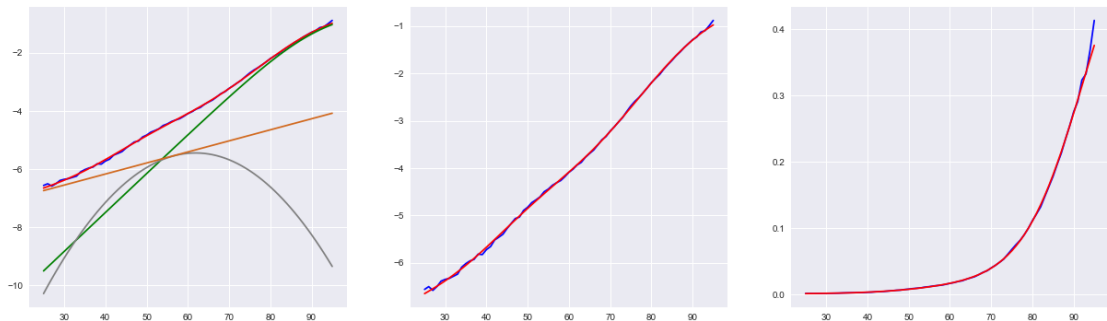
# Annexes



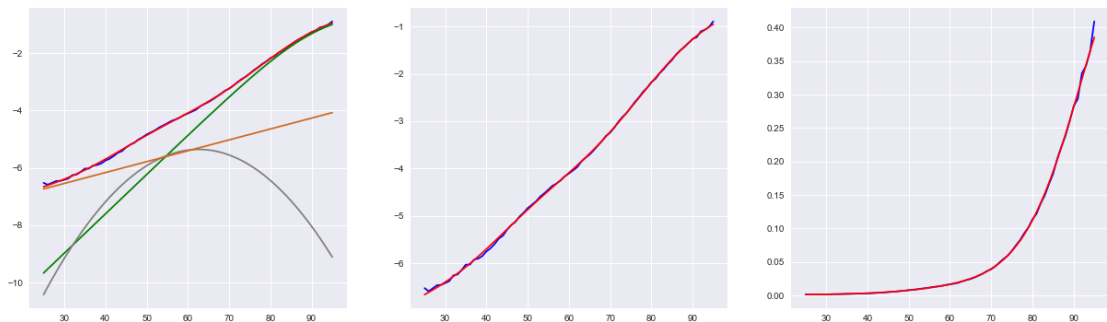
## Annexe A

# Ajustement de la mortalité française par modèle *Logit+ Additif*

Année 1955 :

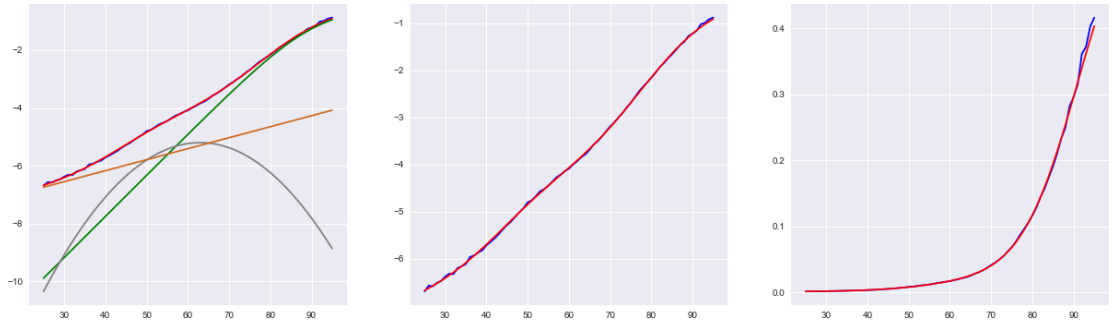


Année 1956 :

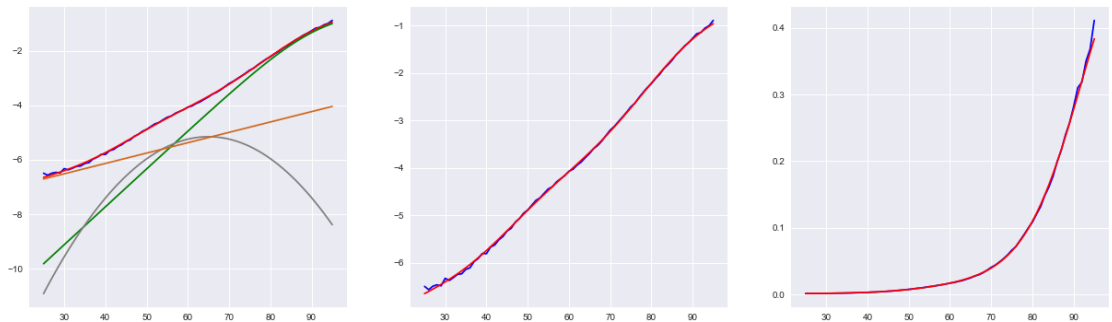


Année 1957 :

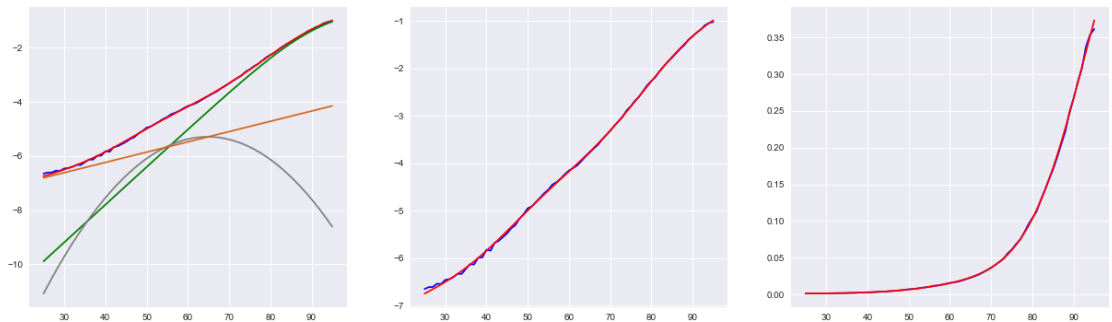
96 ANNEXE A. AJUSTEMENT DE LA MORTALITÉ FRANÇAISE PAR MODÈLE LOGIT+ ADDITIF



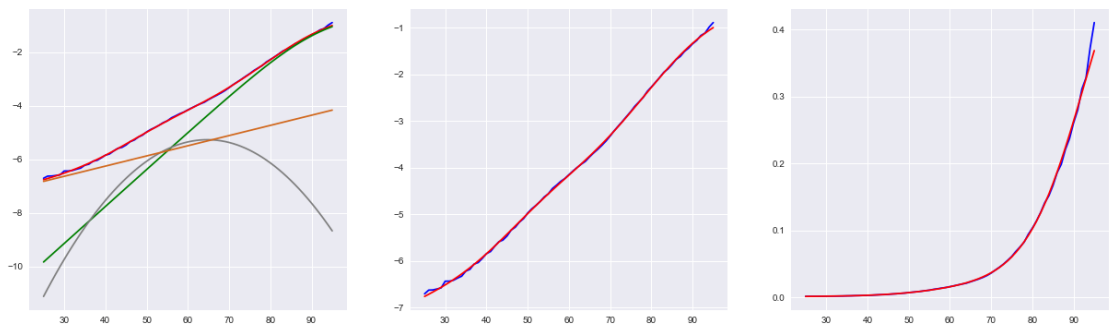
Année 1958 :



Année 1959 :

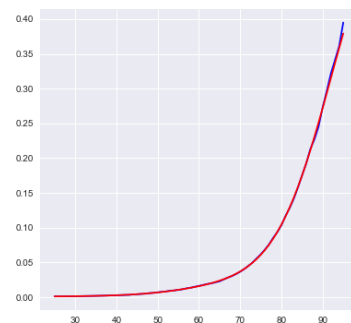
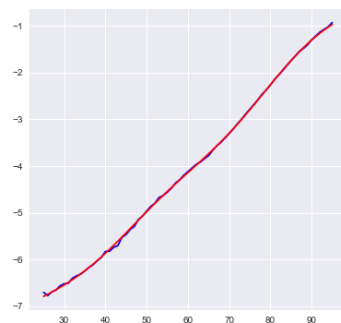
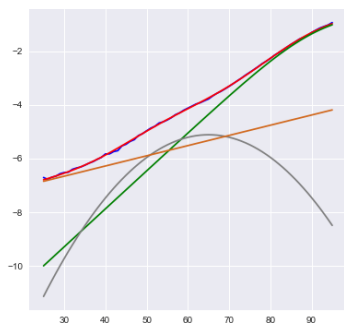


Année 1960 :

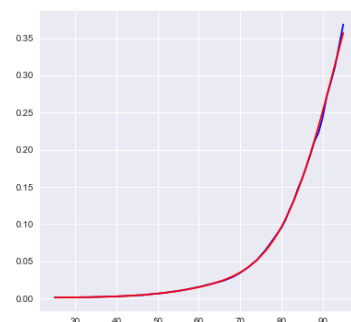
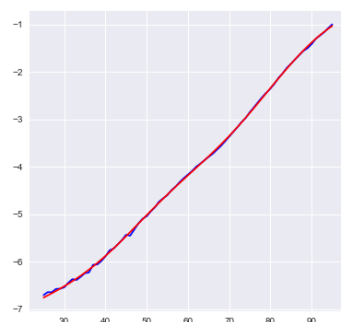
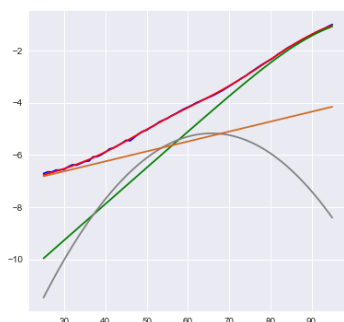




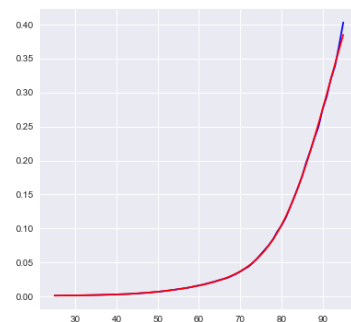
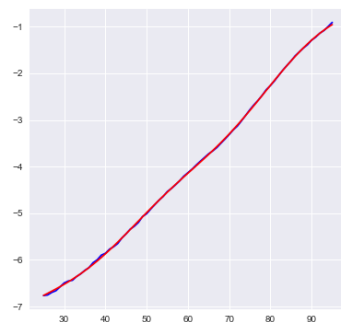
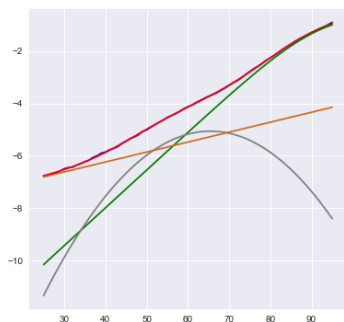
Année 1961 :



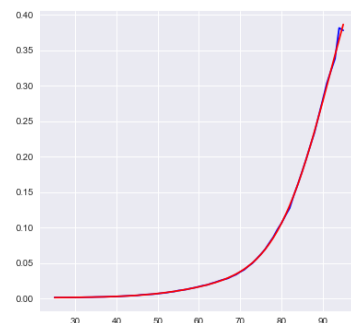
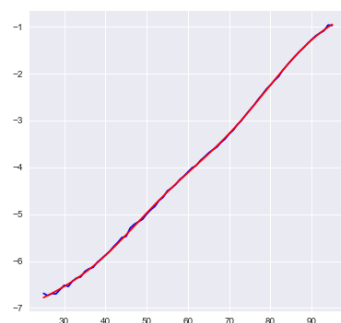
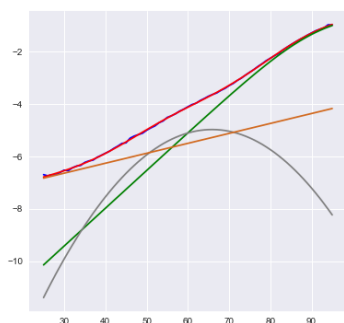
Année 1962 :



Année 1963 :

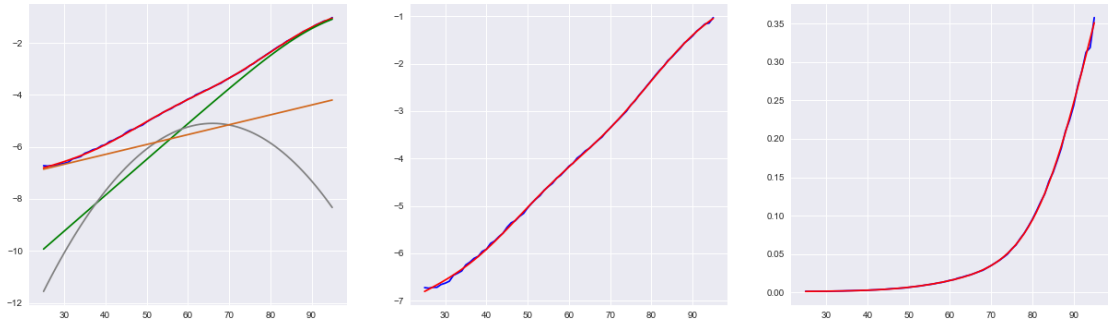


Année 1964 :

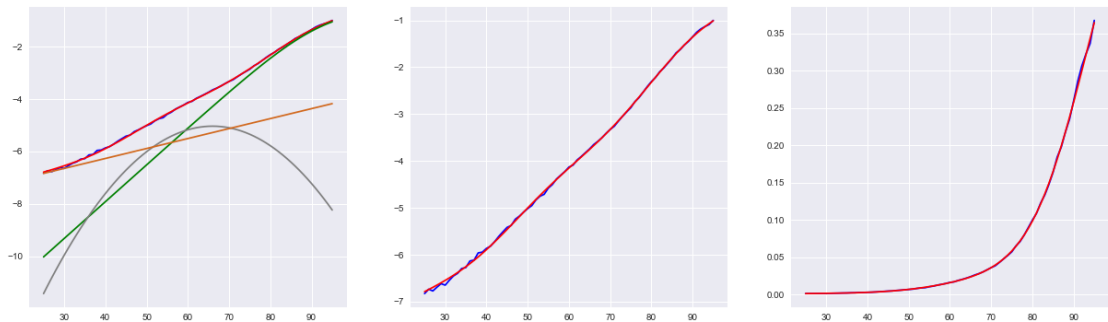


98ANNEXE A. AJUSTEMENT DE LA MORTALITÉ FRANÇAISE PAR MODÈLE LOGIT+ ADDITIF

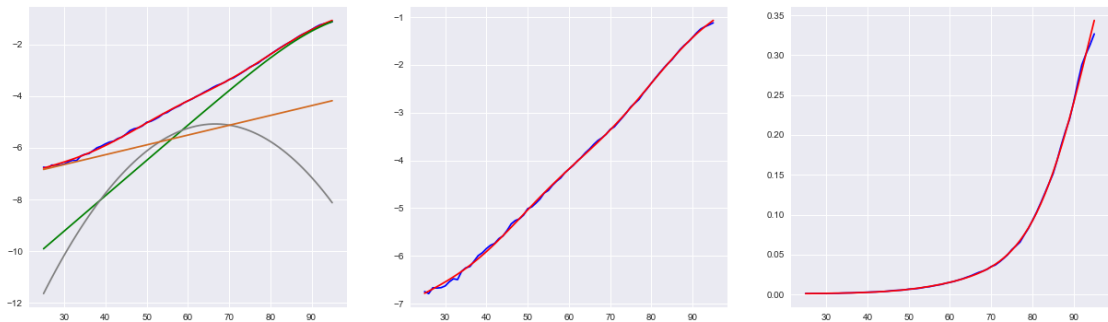
Année 1965 :



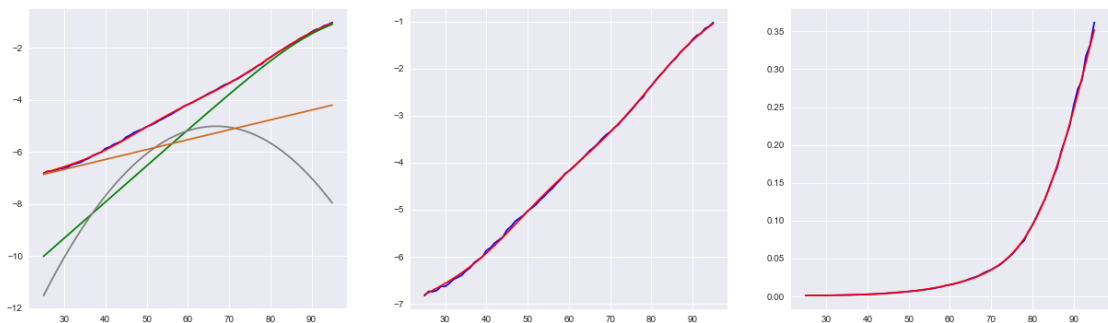
Année 1966 :



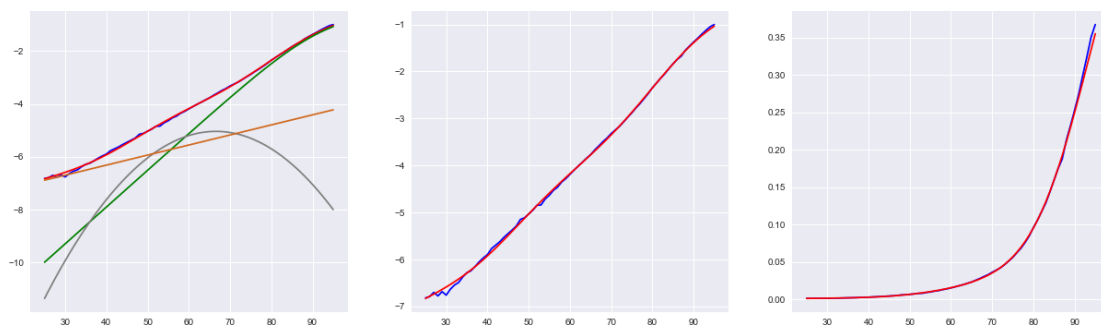
Année 1967 :



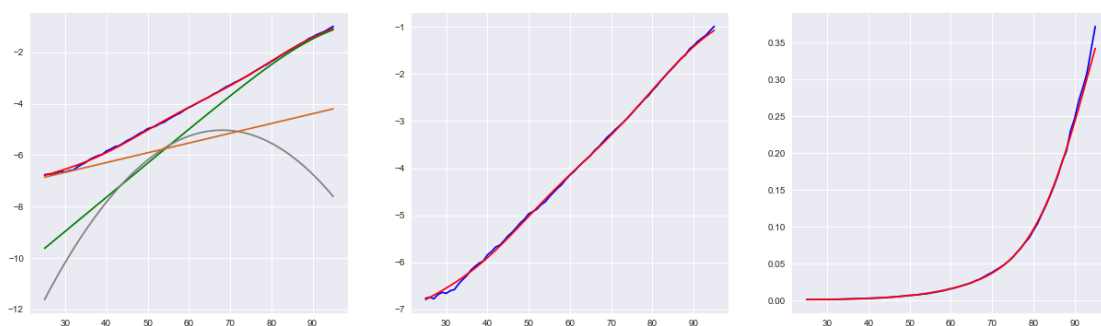
Année 1968 :



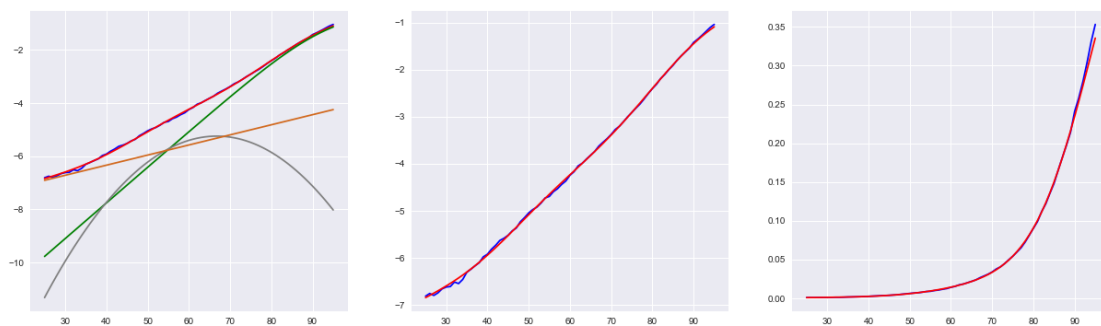
Année 1969 :



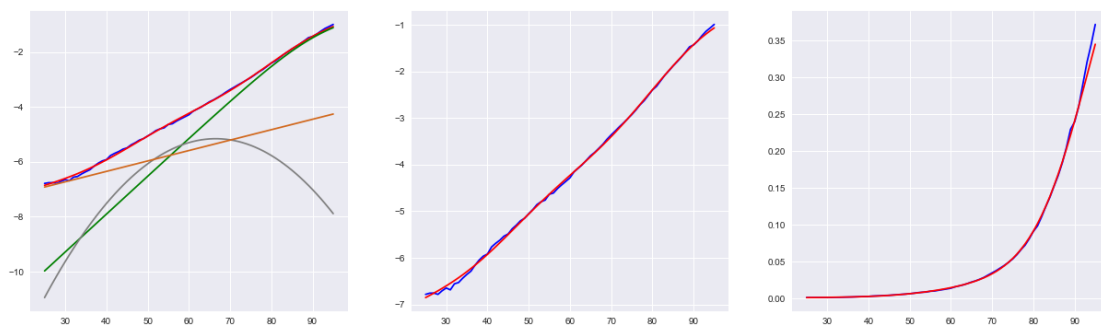
Année 1970 :



Année 1971 :

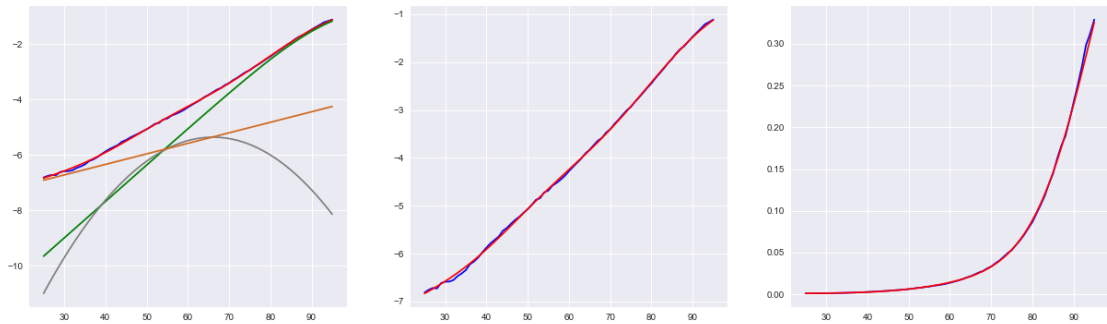


Année 1972 :

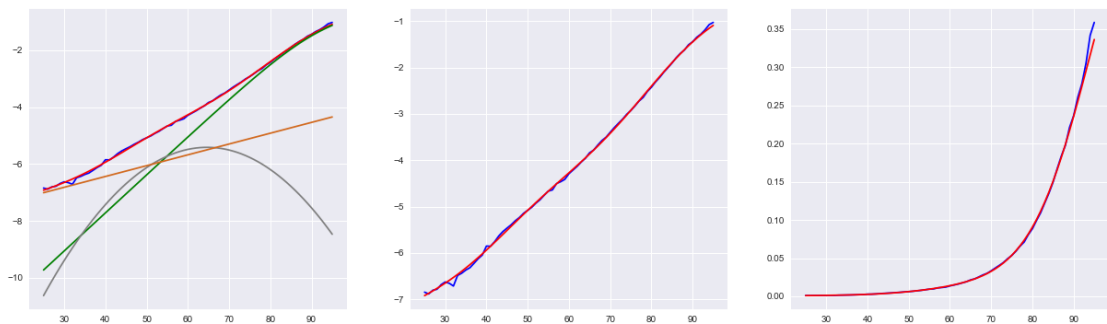


100ANNEXE A. AJUSTEMENT DE LA MORTALITÉ FRANÇAISE PAR MODÈLE LOGIT+ ADDITIF

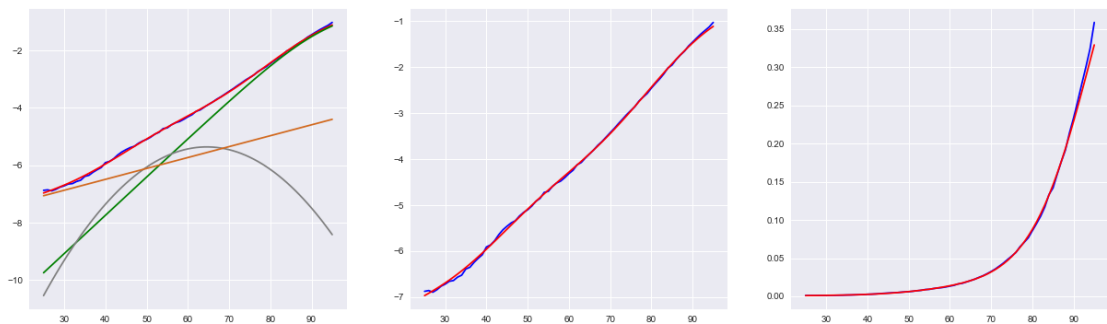
Année 1973 :



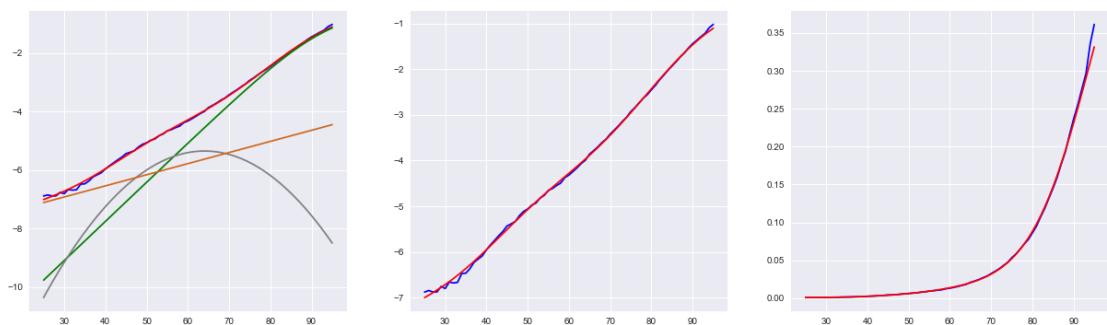
Année 1974 :



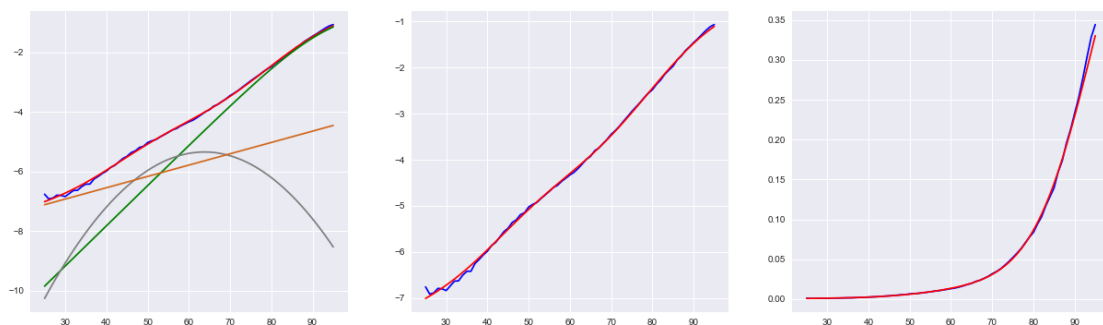
Année 1975 :



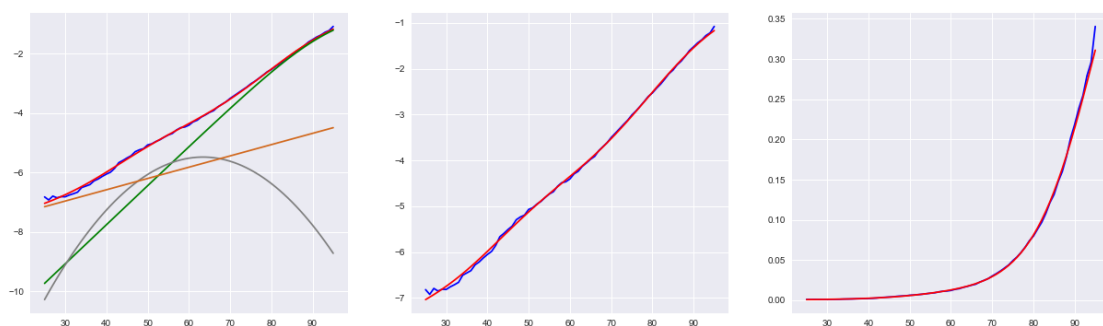
Année 1976 :



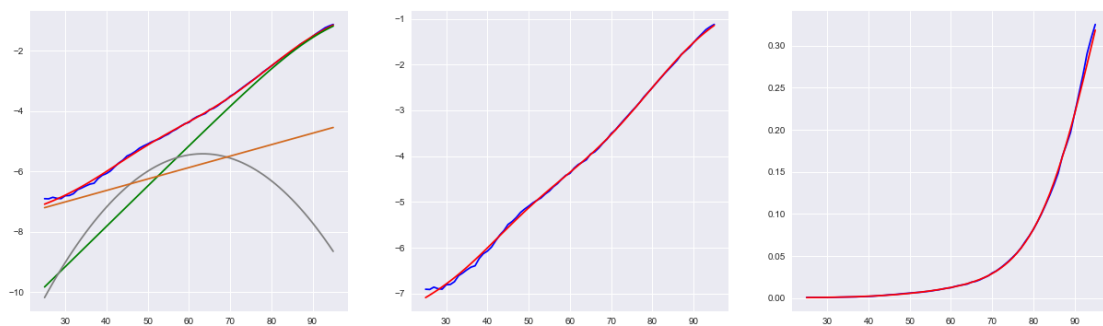
Année 1977 :



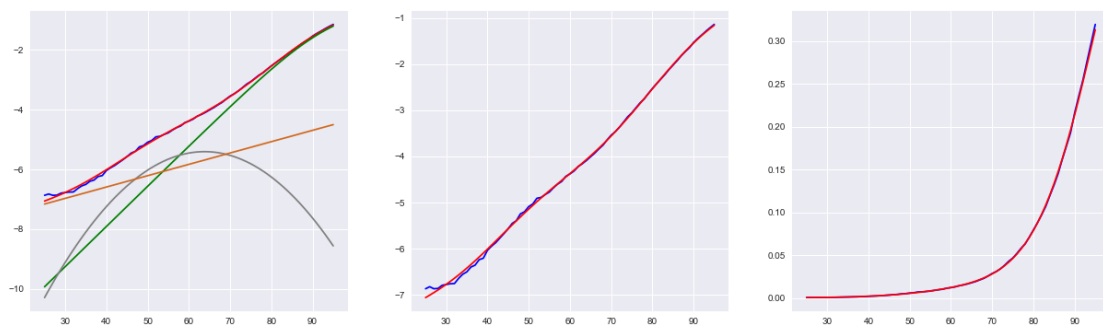
Année 1978 :



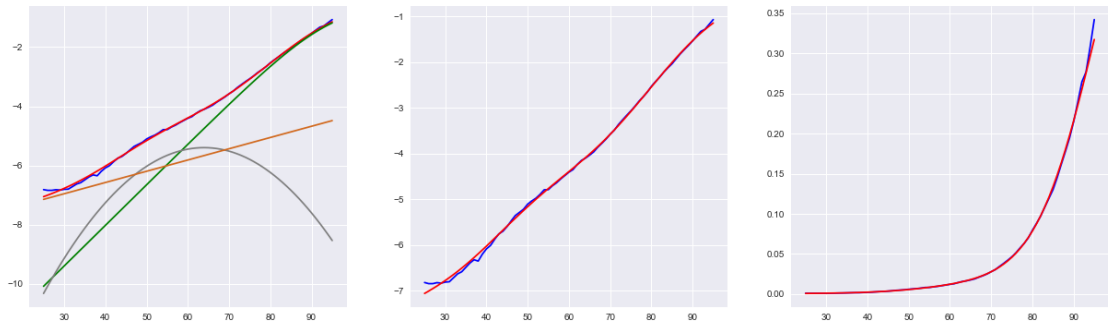
Année 1979 :



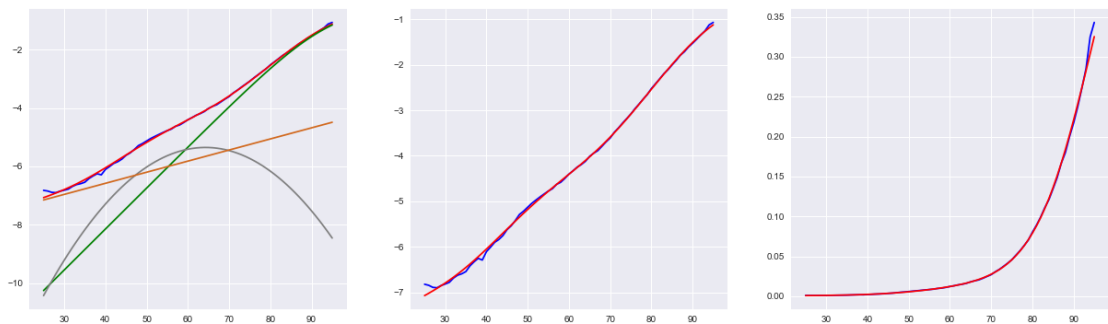
Année 1980 :



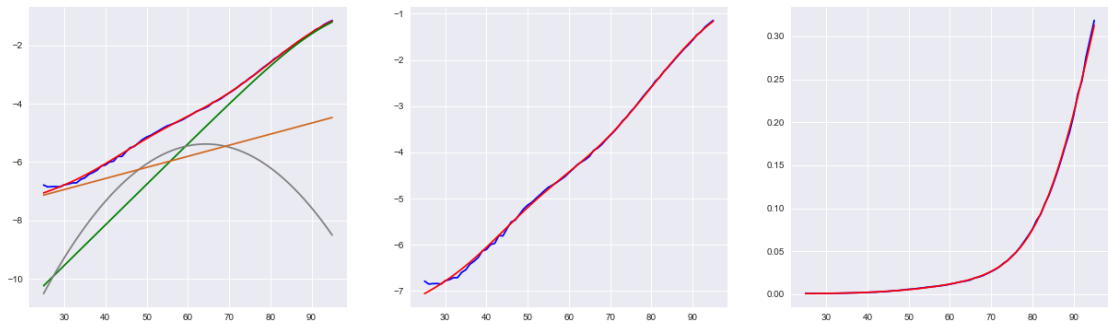
Année 1981 :



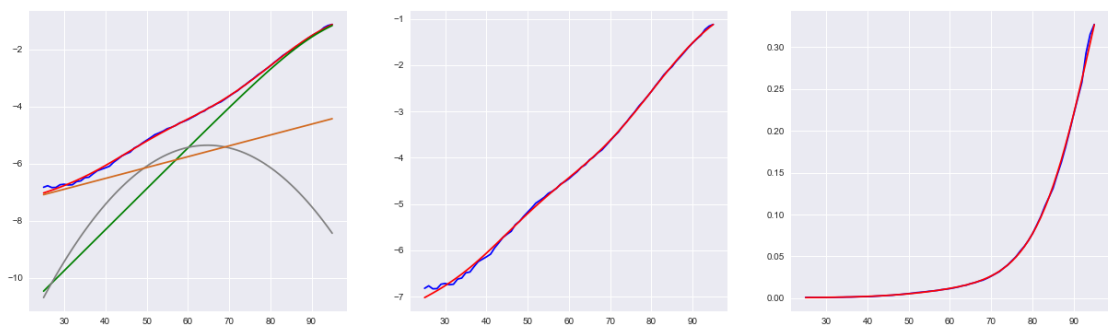
Année 1982 :



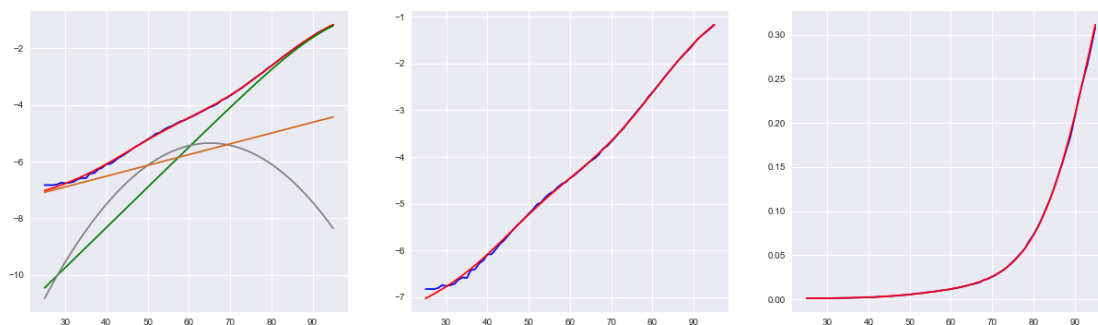
Année 1983 :



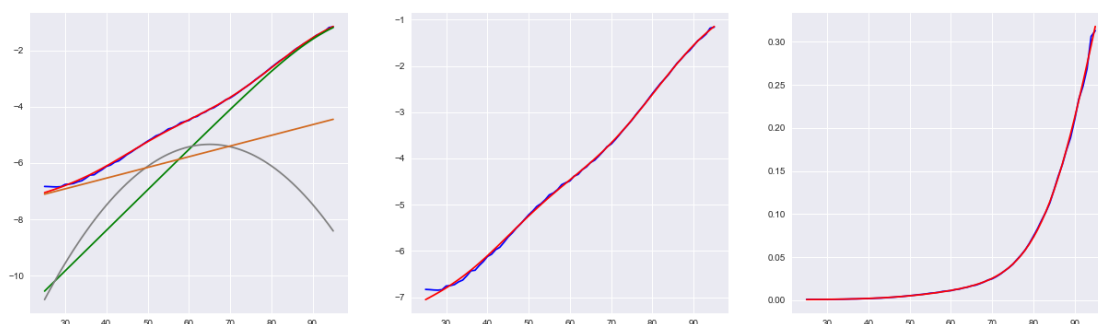
Année 1984 :



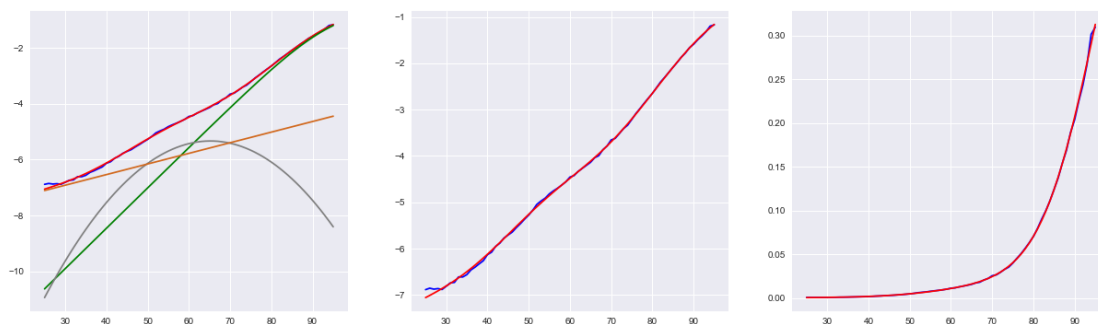
Année 1985 :



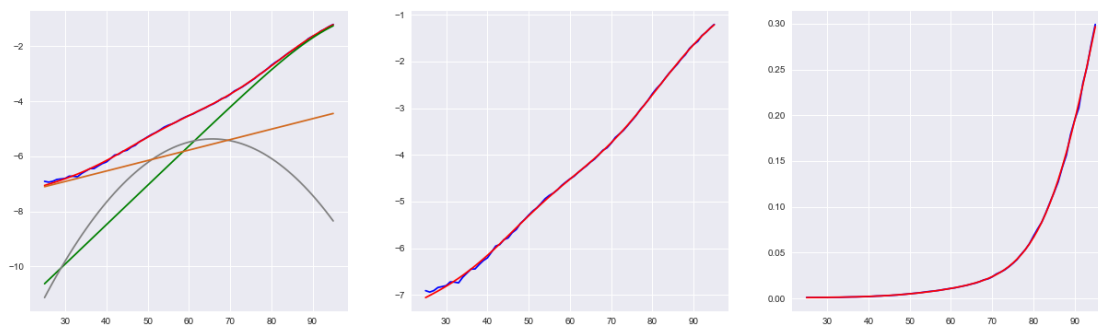
Année 1986 :



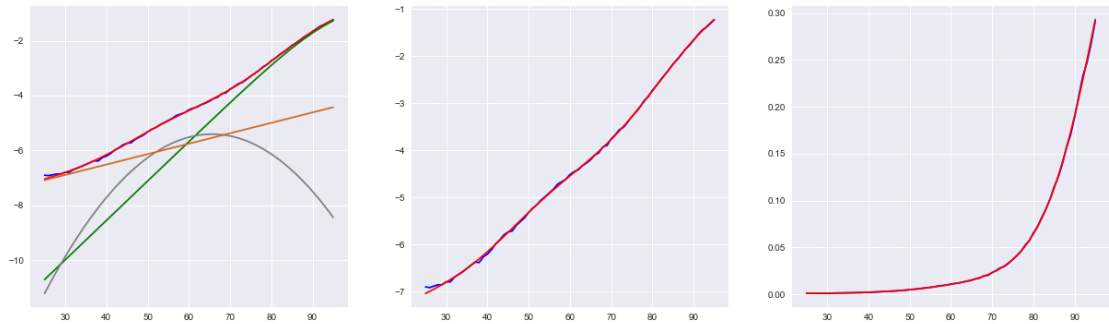
Année 1987 :



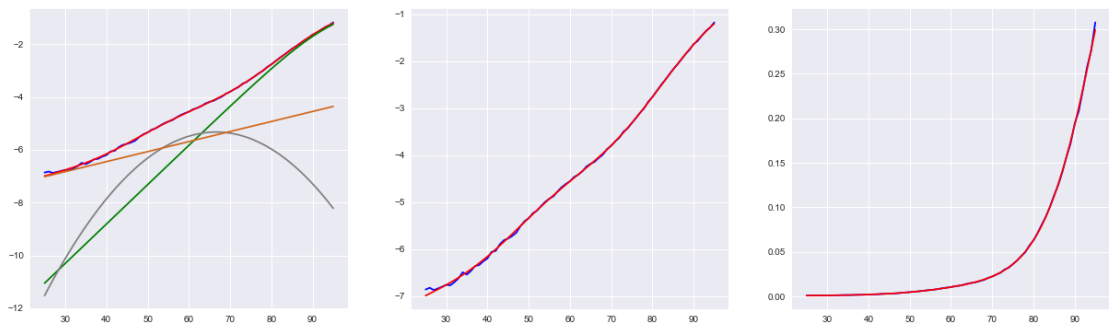
Année 1988 :



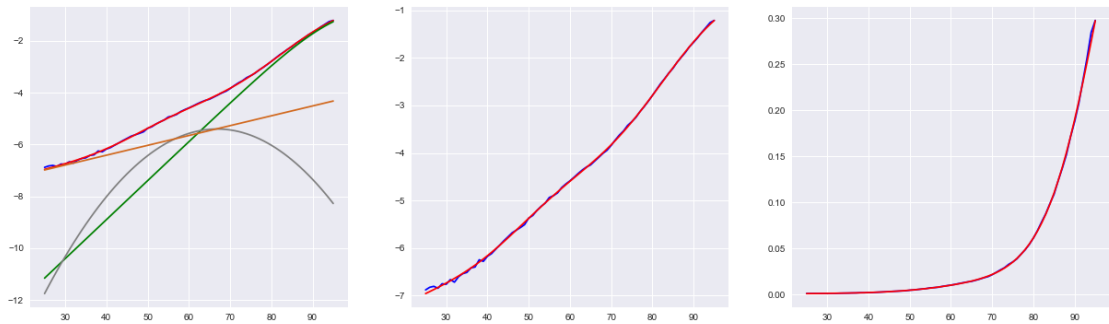
Année 1989 :



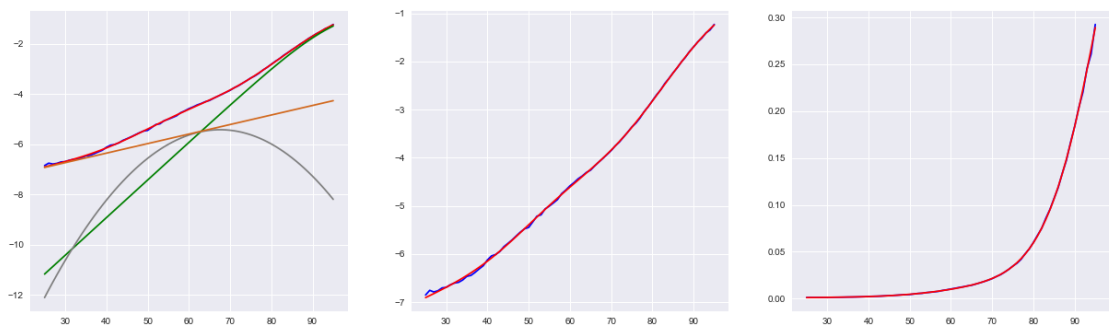
Année 1990 :



Année 1991 :

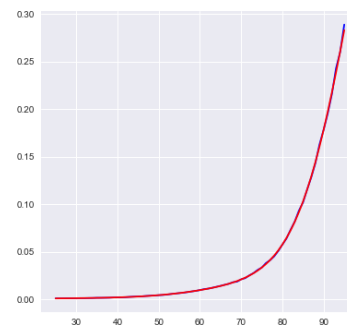
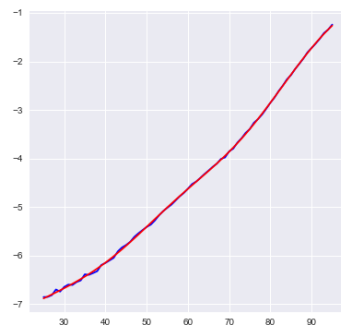
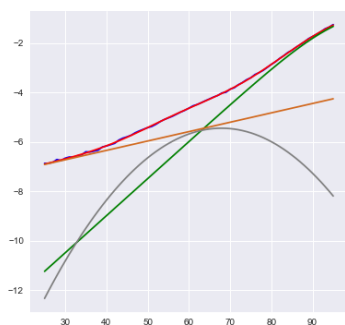


Année 1992 :

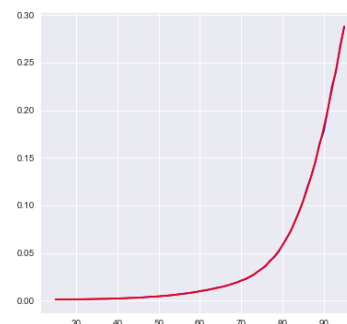
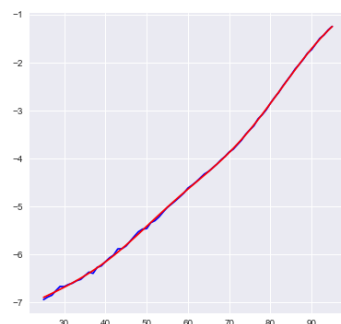
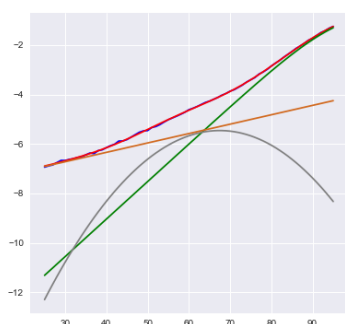




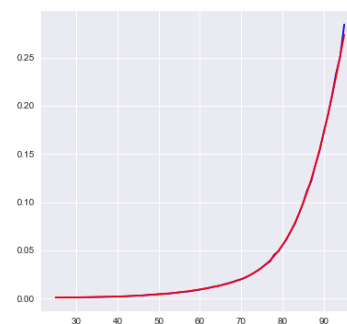
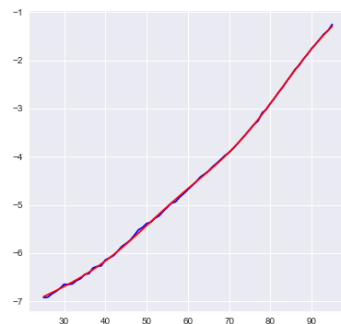
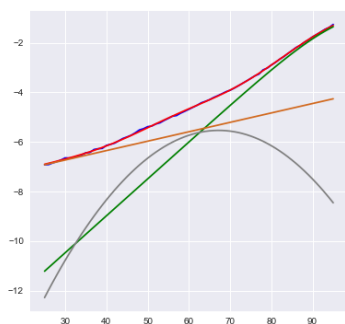
Année 1993 :



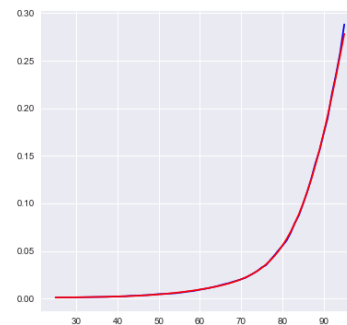
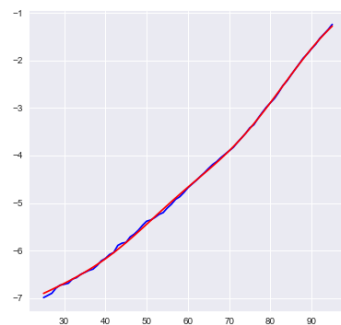
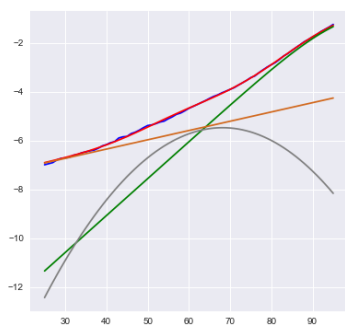
Année 1994 :



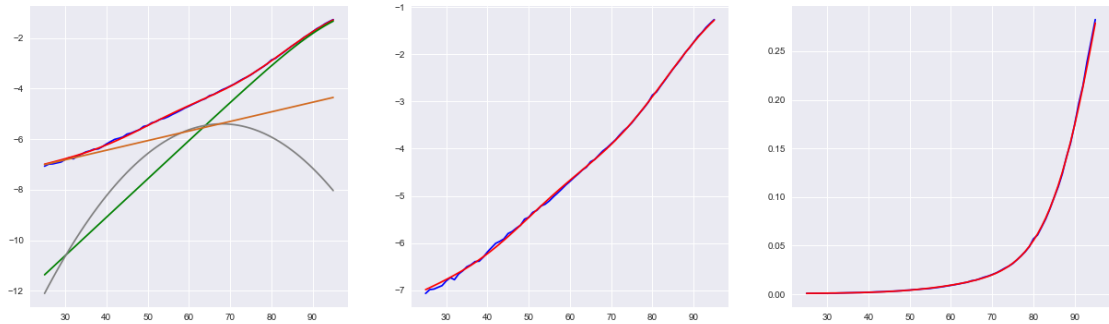
Année 1995 :



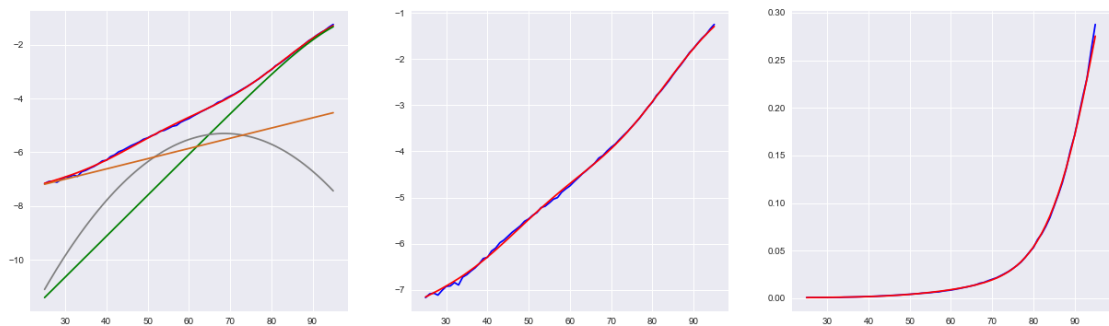
Année 1996 :



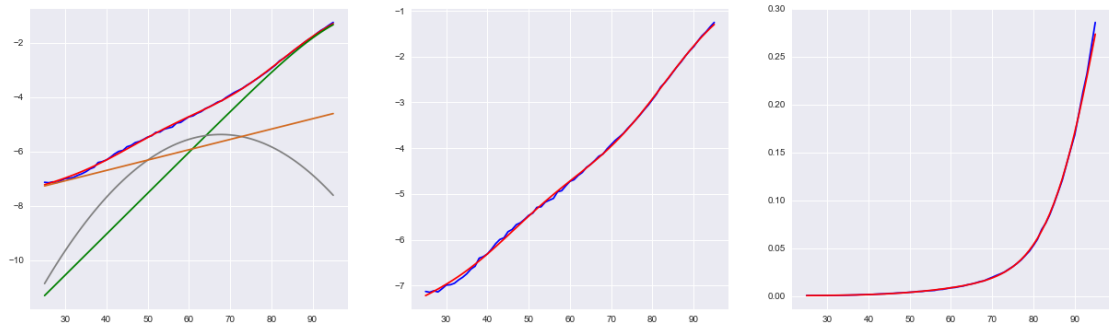
Année 1997 :



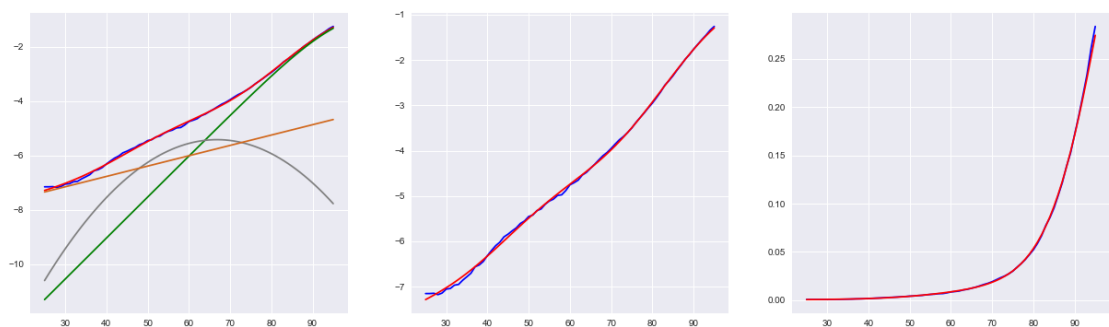
Année 1998 :



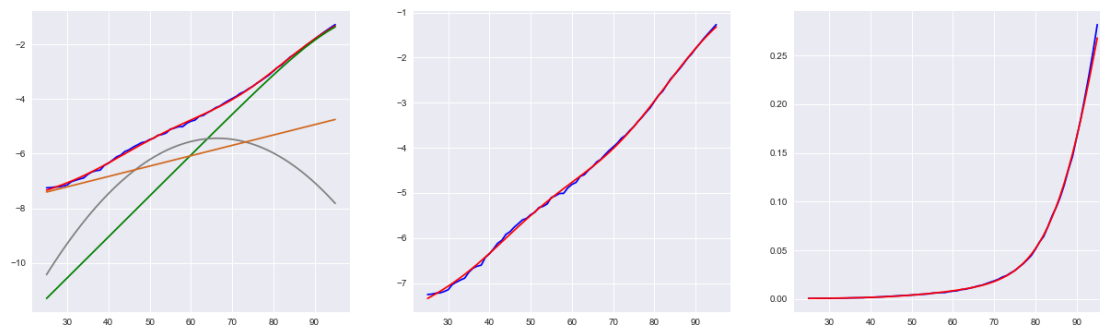
Année 1999 :



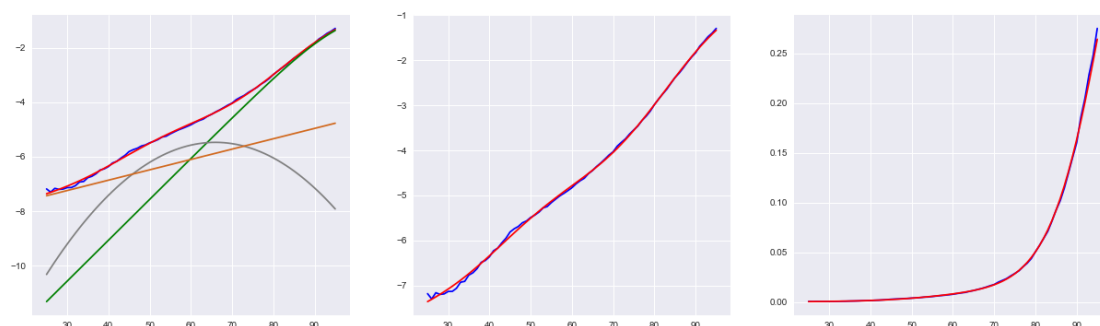
Année 2000 :



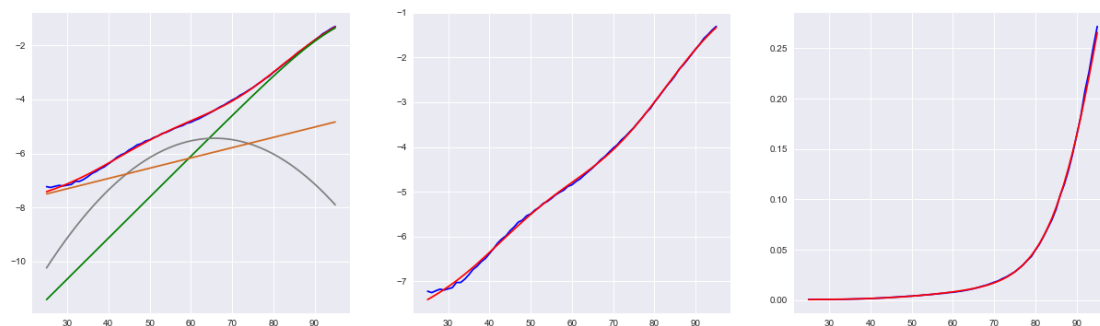
Année 2001 :



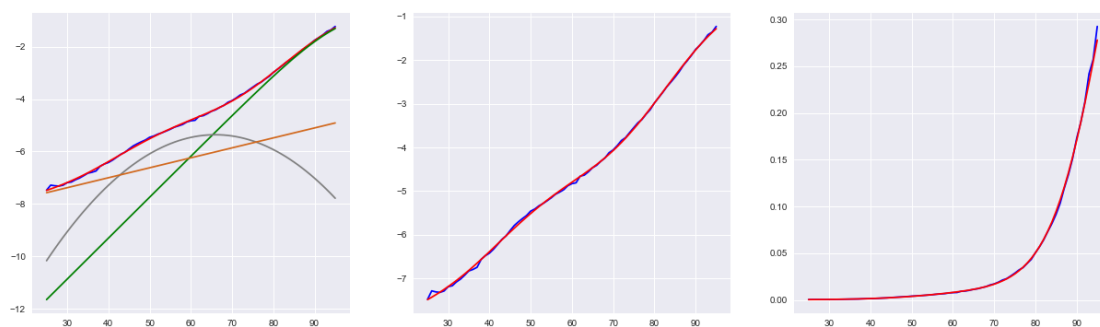
Année 2002 :



Année 2003 :

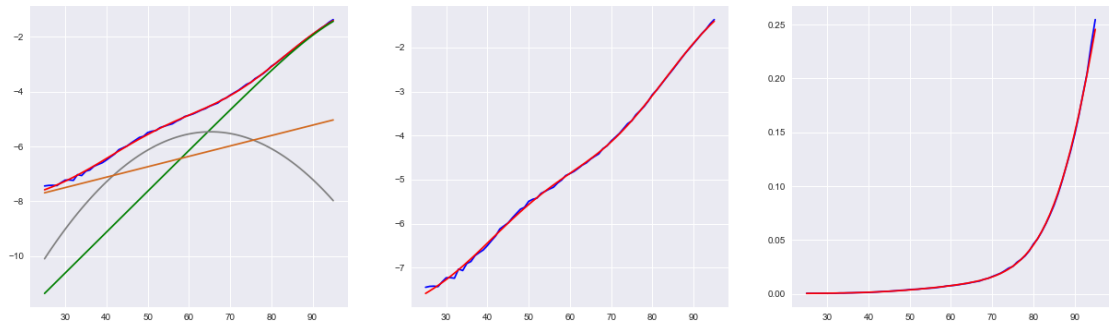


Année 2004 :

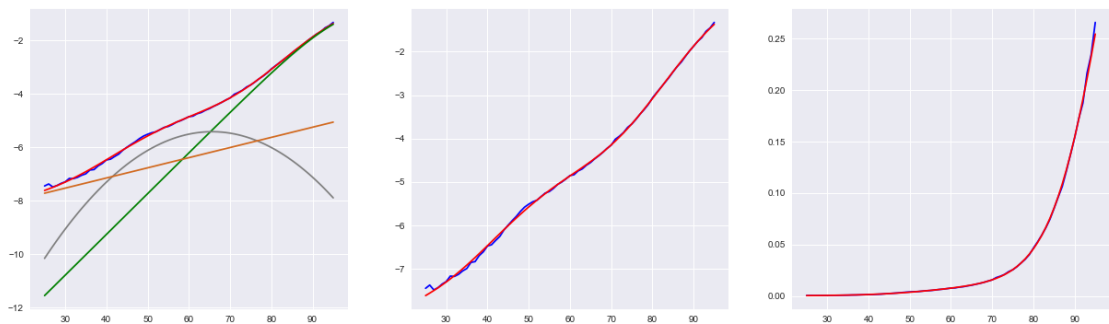


108 ANNEXE A. AJUSTEMENT DE LA MORTALITÉ FRANÇAISE PAR MODÈLE LOGIT+ ADDITIF

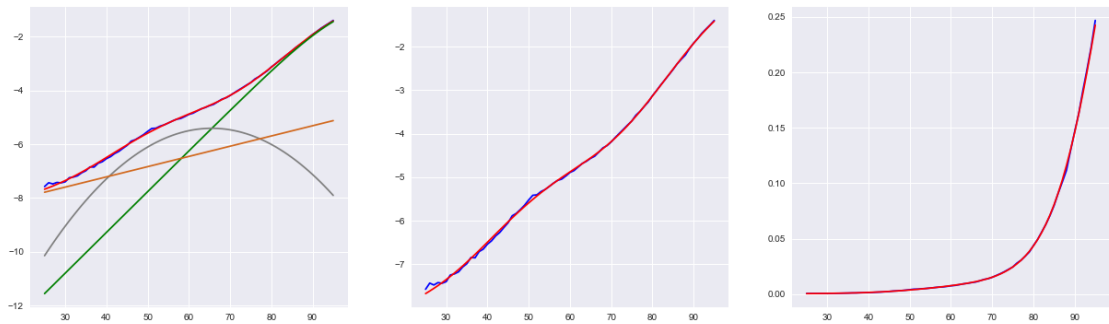
Année 2005 :



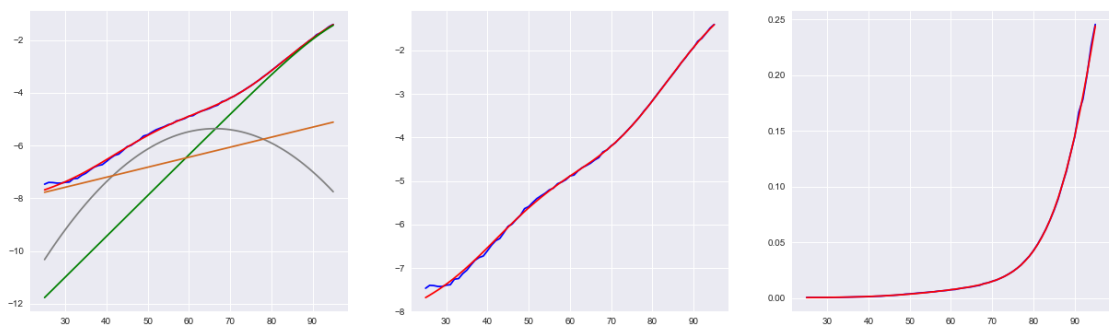
Année 2006 :



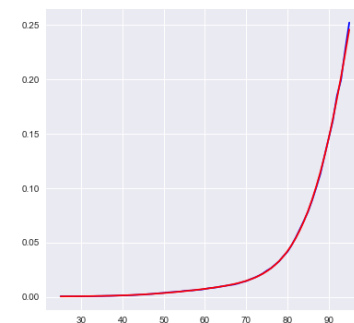
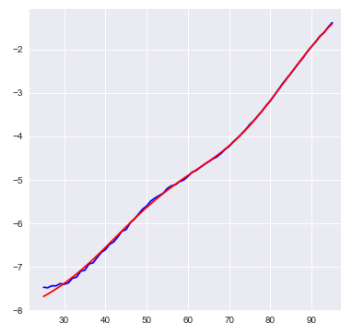
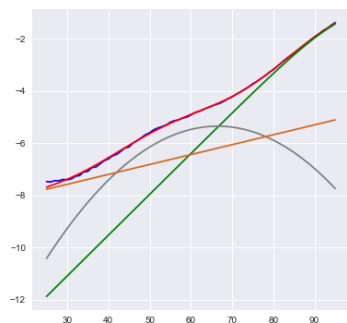
Année 2007 :



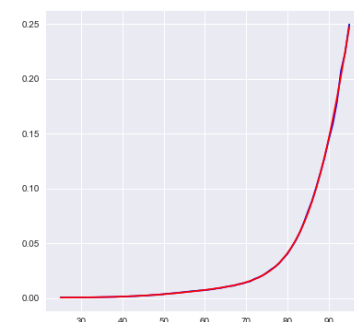
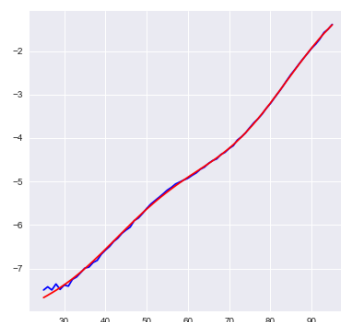
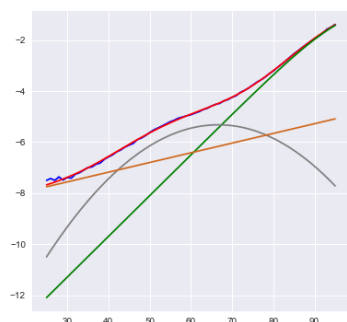
Année 2008 :



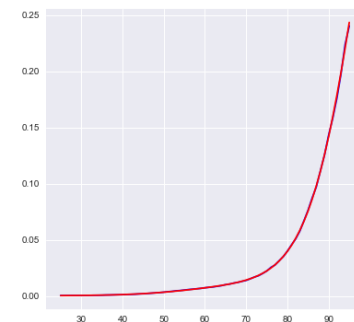
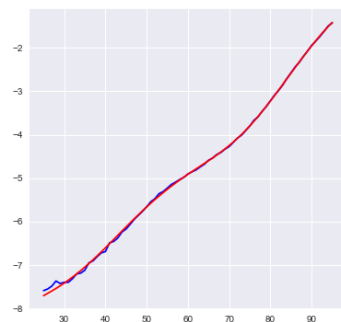
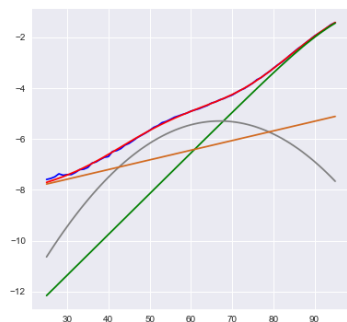
Année 2009 :



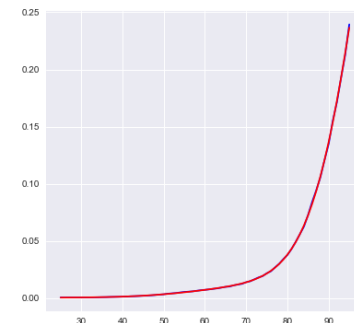
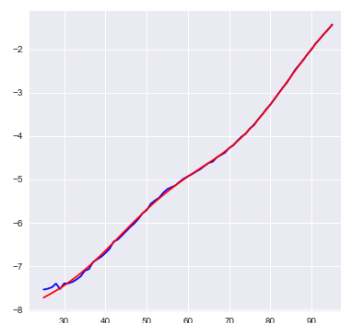
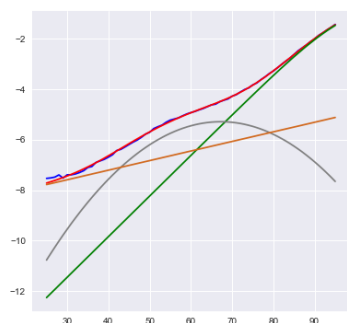
Année 2010 :



Année 2011 :

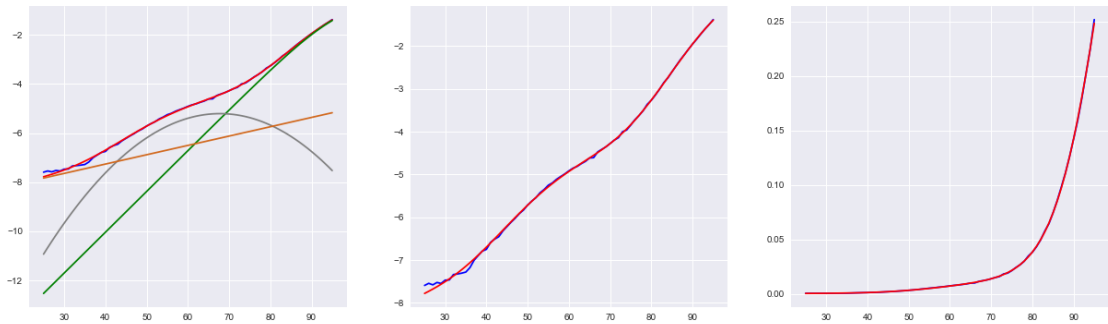


Année 2012 :

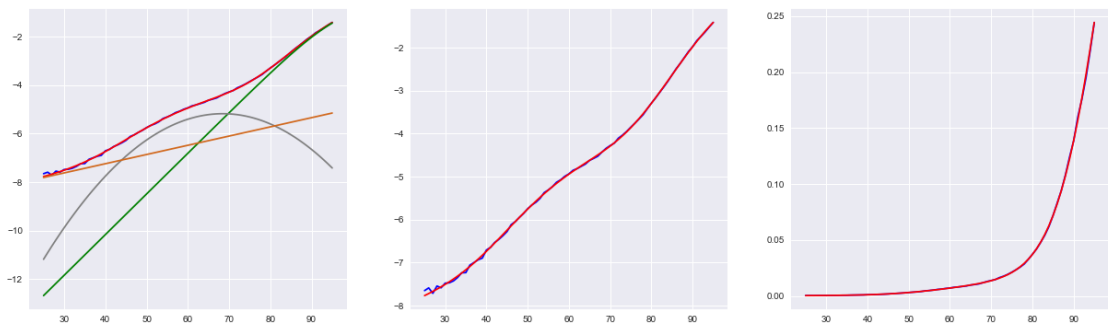


110ANNEXE A. AJUSTEMENT DE LA MORTALITÉ FRANÇAISE PAR MODÈLE LOGIT+ ADDITIF

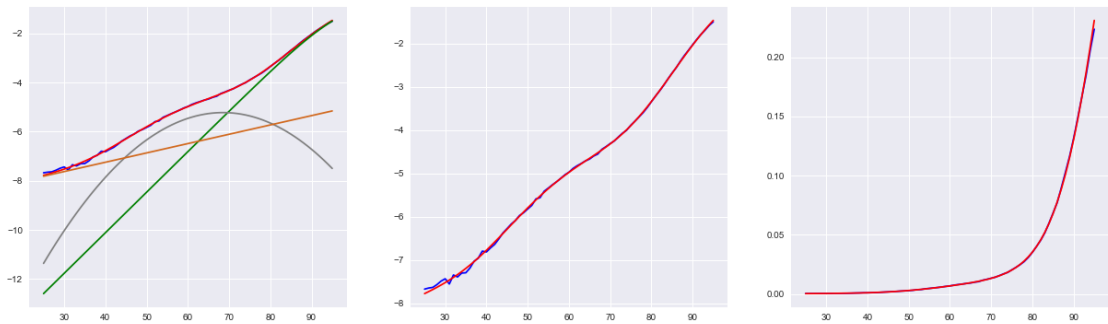
Année 2013 :



Année 2014 :



Année 2015 :



# Annexe B

## Détailles de la projection par VAR(1)

Cette annexe présente les résultats de la projection par vecteur autoregressif d'ordre 1.

Summary of Regression Results				
Model:	VAR			
Method:	OLS			
Date:	Thu, 23, Aug, 2018			
Time:	16:26:16			
No. of Equations:	6.00000	BIC:	-38.0580	
Nobs:	55.0000	HQIC:	-38.9981	
Log likelihood:	662.499	FFI:	6.44936e-18	
AIC:	-39.5909	Det(Omega_mle):	3.14298e-18	

Results for equation a				
	coefficient	std. error	t-stat	prob
const	0.266538	0.091401	2.916	0.004
L1.a	-1.658893	0.756558	-2.193	0.028
L1.a_young	-0.015596	0.008190	1.904	0.057
L1.b	-0.025827	0.008222	-3.141	0.002
L1.beta	2.089069	1.466764	1.424	0.154
L1.m	-0.001898	0.000794	-2.391	0.017
L1.sigma	0.000010	0.000041	0.254	0.799

Results for equation b				
	coefficient	std. error	t-stat	prob
const	-16.155435	6.233513	-2.592	0.010
L1.a	161.302184	51.597091	3.126	0.002
L1.a_young	-0.828095	0.558531	-1.483	0.138
L1.b	2.557580	0.560705	4.561	0.000
L1.beta	-104.063723	100.033000	-1.040	0.298
L1.m	0.120224	0.054134	2.221	0.026
L1.sigma	0.000494	0.002798	0.176	0.860

Results for equation beta				
	coefficient	std. error	t-stat	prob
const	0.006437	0.013227	0.487	0.627
L1.a	-0.009031	0.109489	-0.082	0.934
L1.a_young	-0.000356	0.001185	-0.300	0.764
L1.b	-0.000055	0.001190	-0.046	0.963
L1.beta	0.750933	0.212269	3.538	0.000
L1.m	-0.000073	0.000115	-0.640	0.522
L1.sigma	-0.000009	0.000006	-1.536	0.125

Results for equation m				
	coefficient	std. error	t-stat	prob
const	36.319725	15.428802	2.354	0.019
L1.a	-210.540653	127.709899	-1.649	0.099
L1.a_young	0.847487	1.382440	0.613	0.540
L1.b	-2.732848	1.387822	-1.969	0.049
L1.beta	500.560530	247.595440	2.022	0.043
L1.m	0.475935	0.133988	3.552	0.000
L1.sigma	-0.015159	0.006925	-2.189	0.029

Results for equation sigma				
	coefficient	std. error	t-stat	prob
const	-43.509551	245.101784	-0.178	0.859
L1.a	-666.669405	2028.798132	-0.329	0.742
L1.a_young	9.706397	21.961431	0.442	0.659
L1.b	-6.193935	22.046927	-0.281	0.779
L1.beta	548.371045	3933.298607	0.139	0.889
L1.m	1.723387	2.128539	0.810	0.418
L1.sigma	1.061654	0.110003	9.651	0.000

Correlation matrix of residuals						
	a	a_young	b	beta	m	sigma
a	1.000000	0.161140	-0.981718	0.687360	0.068657	-0.071512
a_young	0.161140	1.000000	-0.164780	0.047457	0.427726	-0.553060
b	-0.981718	-0.164780	1.000000	-0.594337	-0.070833	0.089261
beta	0.687360	0.047457	-0.594337	1.000000	0.389703	0.235346
m	0.068657	0.427726	-0.070833	0.389703	1.000000	0.202853
sigma	-0.071512	-0.553060	0.089261	0.235346	0.202853	1.000000

Nous pouvons remarquer que peu de coefficients sont significatifs.