

Modèles dynamiques de population et tables de mortalité

Prix SCOR 2016 de la Thèse

INSTITUT DES
ACTUAIRES



Journée 100% Actuaires & 100% Datascience, 16 nov. 2017

Alexandre Boumezoued, Milliman R&D

1. Introduction

2. Les tables périodiques de la Human Mortality Database
3. Dynamique des populations et correction des tables de mortalité
4. Quelles conséquences pour l'assurance ?

- Différentes **sources d'incertitude** s'insèrent dans le processus de **modélisation des taux de mortalité futurs**
 - Les **données à disposition** constituent le point de départ d'un tel processus

Qualité des données

Les données de mortalité (internes ou externes) peuvent présenter des anomalies (erreurs, incomplétude,...) qui créent une **volatilité indue**

Erreur de spécification du modèle

Les caractéristiques de l'historique (ex : tendance) ne sont que **partiellement reproduites** par le modèle sélectionné

Erreur d'estimation des paramètres

Les valeurs des paramètres du modèle sont **incertaines** car estimées sur un nombre de points d'observation limité (âge x temps)

Stochasticité intrinsèque au modèle

Le modèle de mortalité contient une stochasticité intrinsèque, principalement incluse dans **l'évolution temporelle**

- L'évaluation du risque de longévité dépend de manière cruciale des **tables de mortalité nationales**
- Des travaux récents posent la question de la fiabilité de ces tables de référence et proposent des **améliorations méthodologiques** pour leur construction (voir notamment Boumezoued & Devineau, L'Actuariel Janvier 2017)
- L'objectif de cet atelier est d'illustrer **l'apport de la modélisation dynamique des populations pour la mesure des risques de mortalité et de longévité**
- Les anomalies étudiées dans cet atelier portent en particulier sur les **« effets cohorte » anormaux**, comme ceux des générations 1919 et 1920, qui présentent dans les données de la *Human Mortality Database* des taux de mortalité particulièrement faibles ou élevés

- La prise de conscience de possibles anomalies de données a émergé suite à la conjecture de **Richards (2008)**, qui suggère que certains 'effets cohorte' en Angleterre pourraient être dus à la **sensibilité de la méthode de calcul des taux de mortalité aux chocs de fertilité**
- Cette conjecture a été confirmée par **Cairns et al. (2016)** qui ont proposé une méthode de correction pour l'Angleterre sur la base de données de nombres de naissances à pas trimestriel
- Depuis, la correction a été étendue à l'ensemble des pays de la Human Mortality Database, en intégrant les données de la Human Fertility Database, voir **Boumezoued (2016)** et **Boumezoued & Devineau (2017)** – ce travail a mis en lumière le caractère universel des anomalies constatées
- Des travaux sont actuellement en cours (**Boumezoued, Hoffmann, Jeunesse, 2017**) afin d'établir un cadre mathématique robuste, et de proposer une méthode unifiée de construction de tables

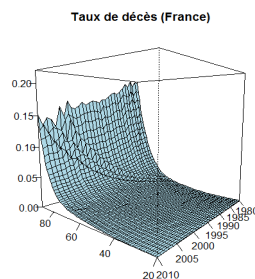
1. Introduction

2. Les tables périodiques de la Human Mortality Database

3. Dynamique des populations et correction des tables de mortalité

4. Quelles conséquences pour l'assurance ?

- La **Human Mortality Database** fournit des estimations de taux de mortalité
 - Son attractivité réside dans le fait que ces informations sont fournies de manière homogène pour différents pays
 - Ces informations nationales font référence en raison de leur richesse et de leur profondeur d'historique
- Des anomalies de données sont en réalité **déjà suggérées** dans la note technique associée à la Human Mortality Database (Wilmoth et al. 2007) :



HMDB Main Menu

- Registration
- New User
- Change Password
- User's Agreement
- Privacy
- FAQ
- Overview
- History
- Publicity
- Research Agreements
- Research Teams
- HMDB Publications
- Partners
- HMDB Summary
- HMDB Metadata
- HMDB Methods
- HMDB
- What's New
- Registration Status
- Data Availability
- Copyright Data Form
- Publication Guidelines
- Links
- Max Planck Institute
- UC Berkeley
- IC Reims Demography
- HMDB
- Human Life Table Database
- General
- Contact Us

The Human Mortality Database

Max Planck Institute for Demographic Research
University of California, Berkeley and INED, Paris

Magali Barbieri, Associate Director
United Nations and formerly University of California, Berkeley

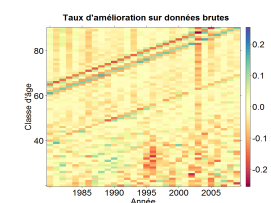
John Wilmoth, Founding Director

The Human Mortality Database (HMDB) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. The project began as an outgrowth of earlier projects in the Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany (see Dittus). It is the work of two teams of researchers in the USA and Germany (see research teams), with the help of financial backers and scientific collaborators from around the world (see acknowledgments). The French Institute for Demographic Studies (INED) has also supported the further development of the database in recent years.

We seek to provide open, international access to these data. At present the database contains detailed population and mortality data for the following 28 countries or areas:

Australia	Finland	Lithuania	Slovenia
Bulgaria	France	Luxembourg	Spain
Belarus	Germany	Netherlands	Sweden
Belgium	Hungary	New Zealand	Taiwan
Canada	Ireland	Norway	U.S.A.
Chile	Israel	Portugal	Ukraine
Czech Republic	Italy	Russia	
Denmark	Japan	Slovakia	
Estonia			

For more information, please begin by reading an [overview](#) of the database. If you have comments or questions, or trouble gaining access to the data, please write to us (Data@demography.berkeley.edu).



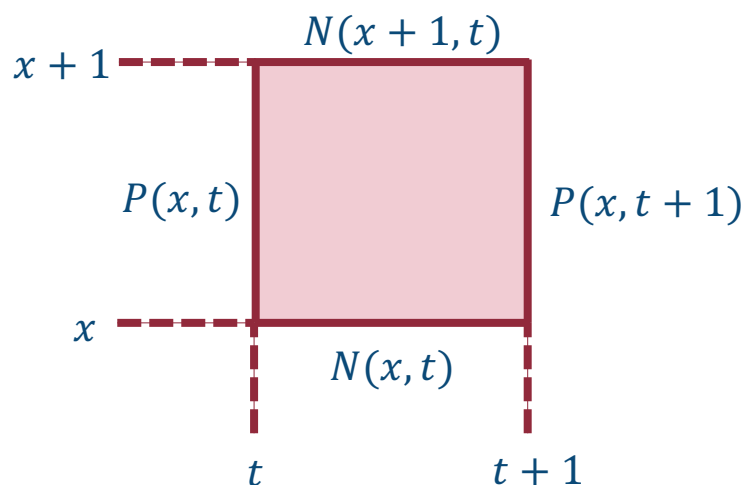
" The assumption [of uniform distribution of births] is violated most severely in situations where there are rapid changes in the size of successive cohorts, owing to fluctuations in the birth series many years before. The worst situation is when a sharp discontinuity in births occurs in the middle of one calendar year, creating a cohort that is "heavy" at one end and "light" at the other. We have not attempted to correct our mortality estimates for the error introduced by such occurrences, which may result in artificially elevated or depressed levels of mortality along a diagonal of the Lexis diagram that follows the cohort(s) in question. The user should be aware of this possibility and not misinterpret the data."

- Quantité de personnes **d'âge exact a au temps exact s** : $g(a, s)$
- Populations de référence sur le diagramme de Lexis :
 - $P(x, t)$: Nombre de personnes à **la date exacte t** d'âge x au dernier anniversaire

$$P(x, t) = \int_x^{x+1} g(a, t) da$$

- $N(x, t)$: Nombre de personnes qui atteignent **l'âge exact x** dans l'année $[t, t + 1)$

$$N(x, t) = \int_t^{t+1} g(x, s) ds$$



- **Référence en temps et âge continu :**

- Quantité (dans l'ensemble des réels) de personnes d'âge exact a au temps exact s : $g(a, s)$

- **Report des décès sur le diagramme de Lexis :**

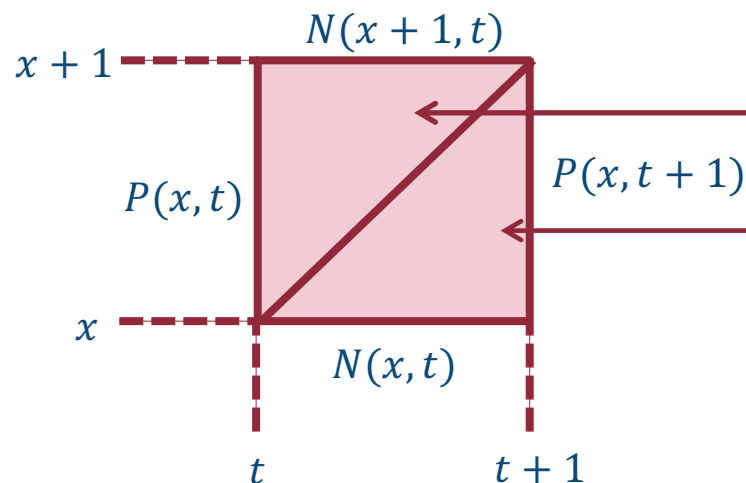
- Nombre de décès **dans l'année** $[t, t + 1)$ de personnes d'âge révolu x au moment du décès

- Nées dans l'année $[t - x - 1, t - x)$: $D_U(x, t)$

- Nées dans l'année $[t - x, t - x + 1)$: $D_L(x, t)$

- **Relations fondamentales :**

$$N(x + 1, t) = P(x, t) - D_U(x, t) \quad \text{et} \quad N(x, t) = P(x, t + 1) + D_L(x, t)$$



Nombre de décès dans les triangles supérieur (U) et inférieur (L)

Les triangles correspondent à deux cohortes différentes

- **Le nombre de décès dans le carré s'écrit :** $D^p(x, t) = D_U(x, t) + D_L(x, t)$

- **Le taux de décès est estimé comme** $\widehat{\mu}^p(x, t) = D^p(x, t)/E^p(x, t)$
 - Dans la vision période, la composante principale de l'exposition au risque **ne peut pas être mesurée directement** par une estimée de population à une date donnée
 - Une hypothèse de **répartition uniforme des naissances** chaque année est utilisée

Approximation sous hypothèse de répartition uniforme des naissances sur les deux années consécutives

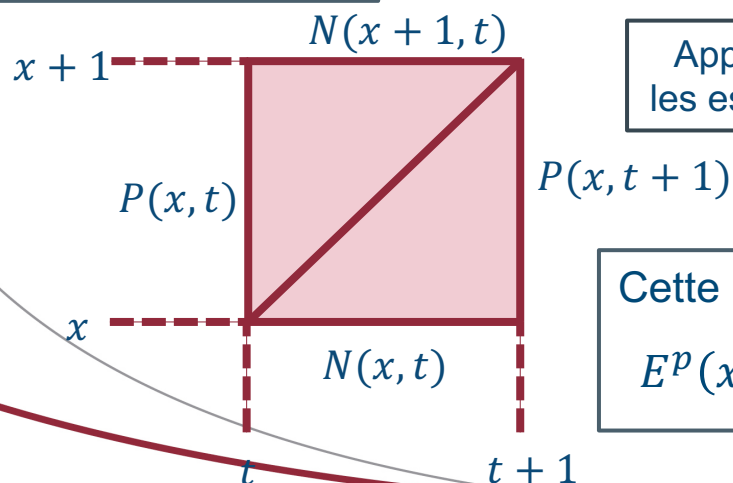
Contribution moyenne à l'exposition sous hypothèse de répartition uniforme des décès dans chaque triangle

Les triangles correspondent à deux cohortes différentes

$$E^p(x, t) = \frac{1}{2} [N(x, t) + N(x + 1, t)] - \frac{1}{3} [D_L(x, t) - D_U(x, t)]$$

Approximation basée sur les estimées de populations

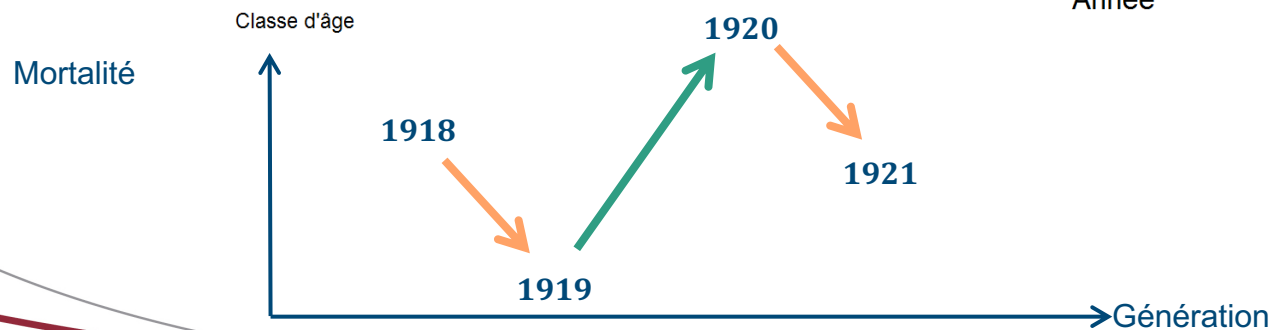
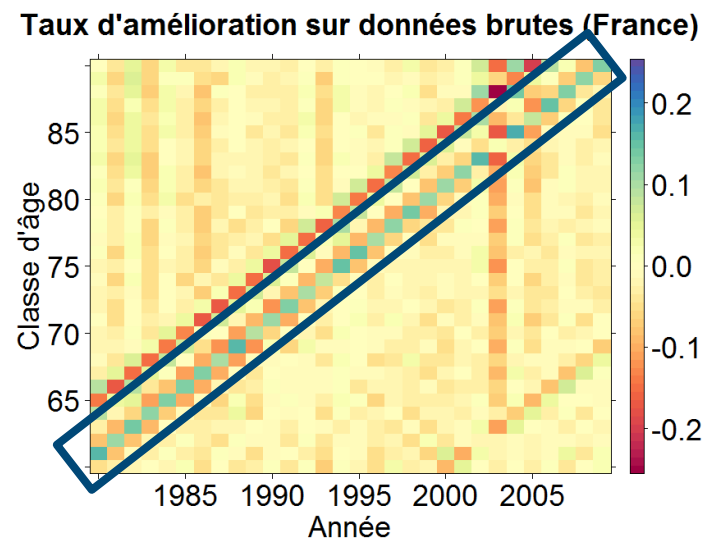
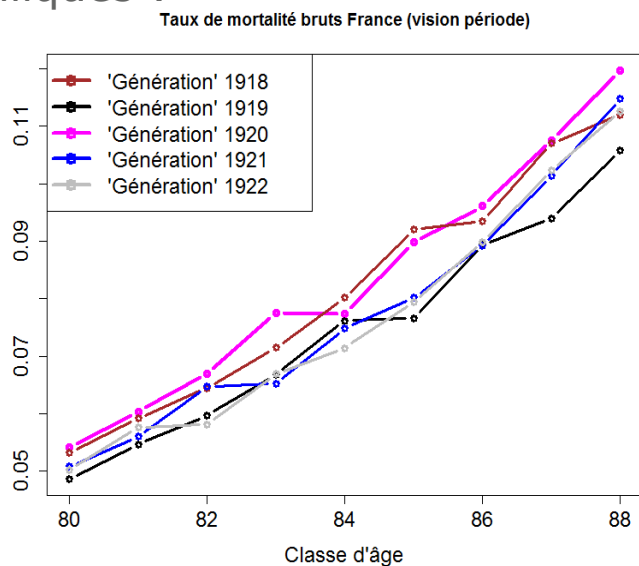
Correction (faible) de la réalisation des décès dans chaque triangle



Cette approximation s'écrit finalement

$$E^p(x, t) = \frac{1}{2} [P(x, t) + P(x, t + 1)] + \frac{1}{6} [D_L(x, t) - D_U(x, t)]$$

- Relations d'ordre inversées pour certaines générations et 'effets cohorte' anormaux :

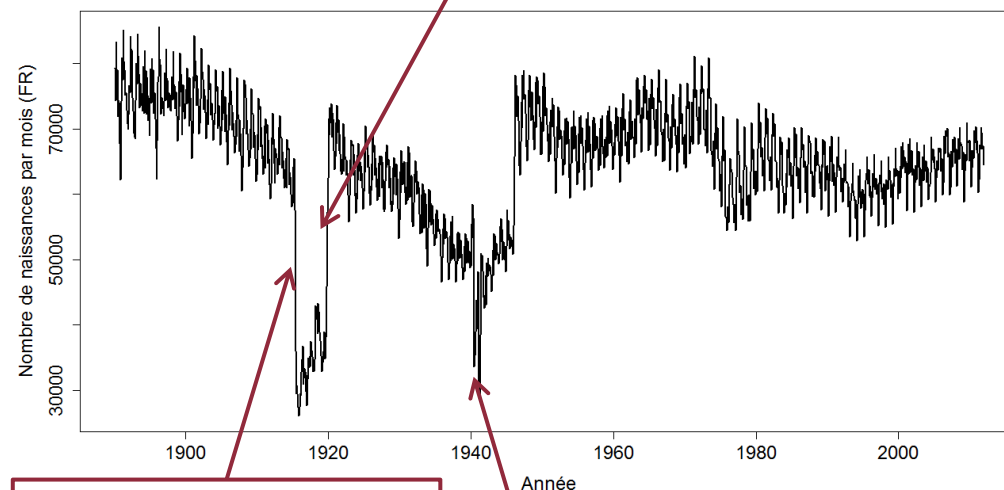
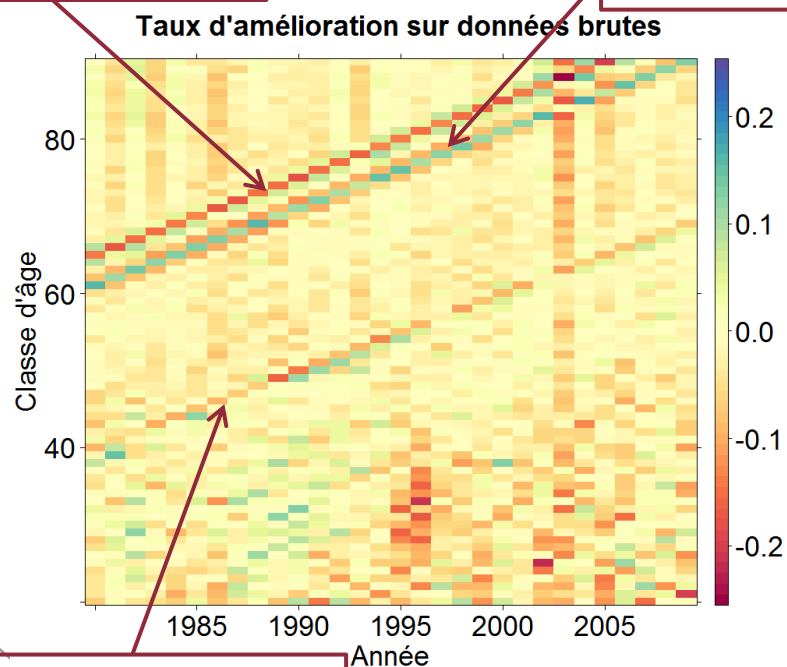


- Parallèle entre chocs sur les naissances et effets cohorte isolés...

Génération ~ 1915

Génération ~ 1920

Choc ~ 1920



Choc ~ 1915

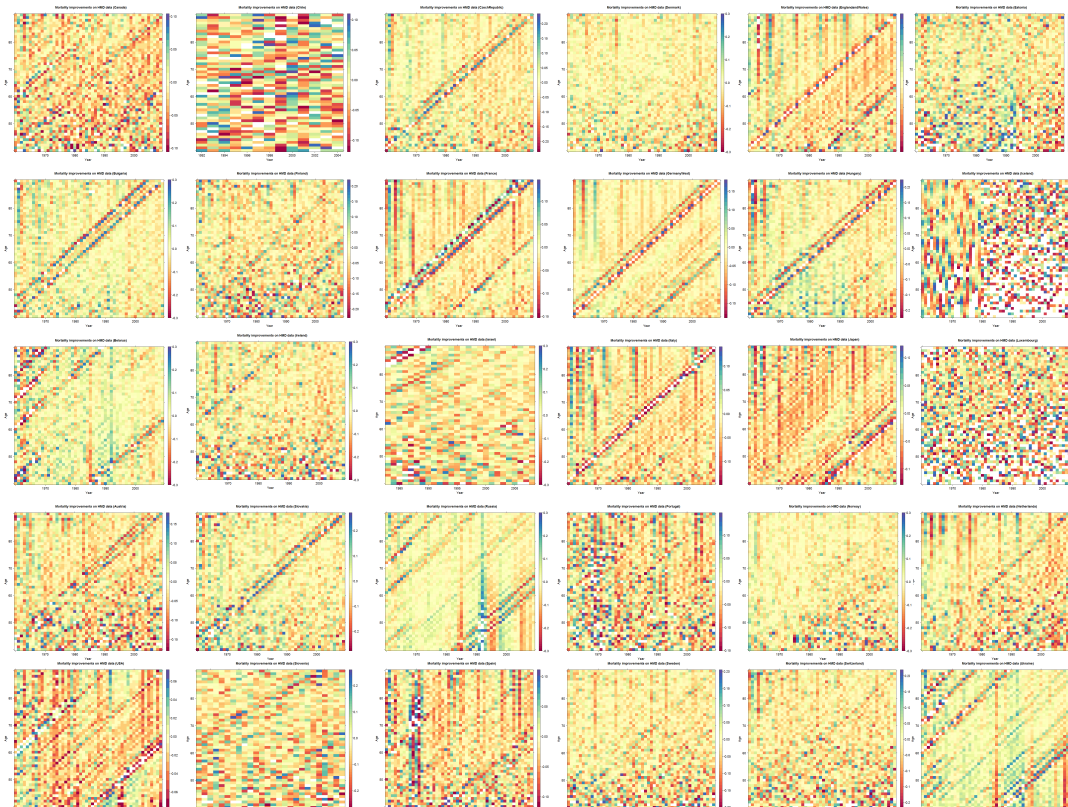
Choc ~ 1940

Effets cohorte ou anomalies ?

Génération ~ 1940

- Les effets cohorte isolés constituent des anomalies universelles

Taux d'amélioration de la mortalité pour 30 régions du monde, issus de la Human Mortality Database

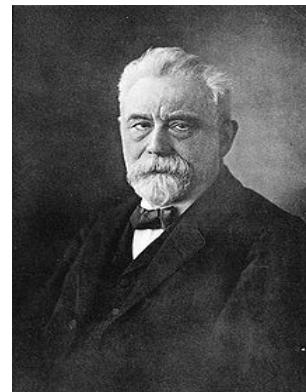
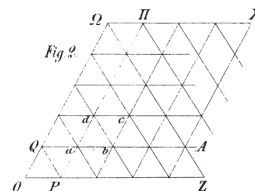
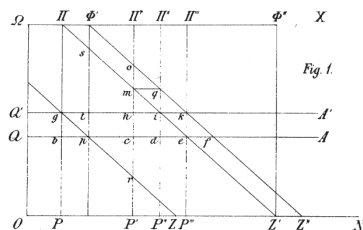


1. Introduction

2. Les tables périodiques de la Human Mortality Database

3. Dynamique des populations et correction des tables de mortalité

4. Quelles conséquences pour l'assurance ?



- Les représentations graphiques innovantes de **Lexis (1875)** et de ses contemporains ont permis d'appréhender deux caractéristiques majeures de la construction de tables à l'échelle d'un pays :
 - d'abord, la structuration de la mortalité autour de deux dimensions : l'âge et le temps,
 - ensuite, le caractère dynamique de l'évolution de la population, notamment la décroissance par les décès d'une part et le renouvellement par les naissances d'autre part
- Ces deux caractéristiques posent des problèmes **théoriques et pratiques** encore **irrésolus à ce jour**

Dynamique des populations et correction des tables

Le cadre continu déterministe de référence

- Quantité (dans l'ensemble des réels) de personnes **d'âge exact a au temps exact s** : $g(a, s)$
- Le taux de décès $\mu(a, s)$ donne l'évolution de chaque cohorte au cours du temps

- Cette évolution s'écrit à partir du nombre de naissances $g(0, s)$ au temps s :

$$g(a, s + a) = g(0, s) \underbrace{\exp \left\{ - \int_0^a \mu(u, s + u) du \right\}}_{\text{Survie pour les individus nés en } s}$$

- Si l'on dérive en âge et temps, on retrouve la variation infinitésimale de la population :

$$(\partial_a + \partial_s)g(a, s) = -\mu(a, s)g(a, s)$$

Nombre de décès $d(a, s)$
d'âge exact a au temps exact s

- Le problème de l'estimation statistique du taux de décès
 - Les méthodes d'estimation non-paramétriques classiques ne peuvent pas s'appliquer en présence de ces **dimensions âge et temps croisés**
 - En pratique, on suppose que le taux de décès est constant pour les plages d'âges et temps
 - Exemple (**vision période**) : $\mu(x, t) = \mu(a, s) \forall (a, s) \in [x, x + 1) \times [t, t + 1)$

- Exposition au risque annuelle (vision période)**

- Sous l'hypothèse précédente, le nombre de décès dans l'année t de personnes d'âge révolu x est

$$D(x, t) = \int_x^{x+1} \int_t^{t+1} d(a, s) ds da = \mu(x, t) \int_x^{x+1} \int_t^{t+1} g(a, s) ds da$$

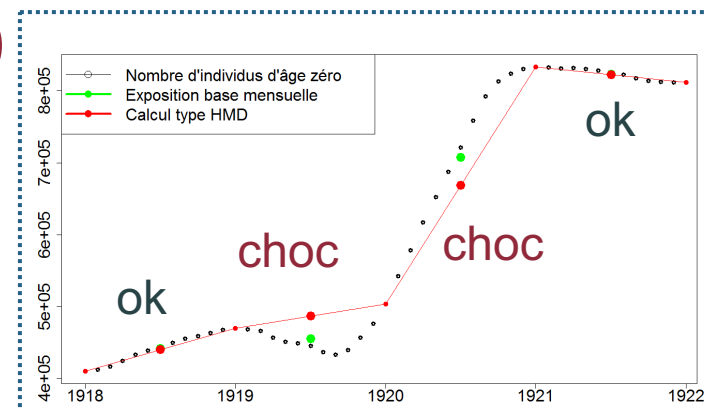
Exposition au risque annuelle =
durée totale vécue dans l'année t
par les individus d'âge révolu x

- **Construction des tables de mortalité** : le taux de mortalité $\mu(x, t)$ peut être estimé par $\hat{\mu}(x, t) = D(x, t)/E(x, t)$
- $D(x, t)$ est le nombre de décès d'individus d'âge x l'année t
- L'exposition au risque $E(x, t)$ quantifie la durée totale vécue par les individus d'âge x l'année t
 - En pratique, **cette quantité est estimée** comme moyenne des populations de début et de fin d'année, car la population nationale n'est pas observée continûment
 - Mathématiquement, cela revient à mettre en œuvre **l'approximation intégrale** :

$$E(x, t) = \int_t^{t+1} P(x, s) ds \approx \frac{1}{2} (P(x, t) + P(x, t + 1))$$

➔ **Problème** : cette approximation n'est pas valable en présence de chocs dans la population

- La méthode de correction basée sur les séries de fertilité prend en compte des **naissances à pas mensuel** permettant d'affiner le calcul de l'exposition au risque



- **Objectif** : mise en place d'un indicateur de qualité pour une table périodique
- Construction de l'indicateur de qualité en trois étapes :
 1. Calcul de tailles de populations $P^*(0, s)$ pour s mensuel.

âge révolu 0 au temps exact s

$$P^*(0, s) = \int_0^1 g^*(a, s) da = \int_0^1 g(0, s - a) da$$

2. Calcul d'expositions au risque sur la base des estimées mensuelles

$$E^*(0, t) = \int_t^{t+1} \int_0^1 g^*(a, s) da ds = \int_t^{t+1} P^*(0, s) ds$$

3. Calcul de l'indicateur de qualité pour la cohorte née l'année t

Un ratio $I(t)$ proche de 1 indique une bonne approximation pour la cohorte née l'année t

$$I(t) = \frac{E^*(0, t)}{\frac{1}{2} [P^*(0, t) + P^*(0, t + 1)]}$$

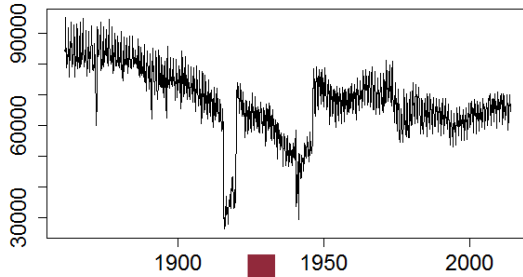
Cette première approche ne tient pas compte des décès et des migrations

L'indicateur est ensuite utilisé pour ajuster les taux de mortalité bruts

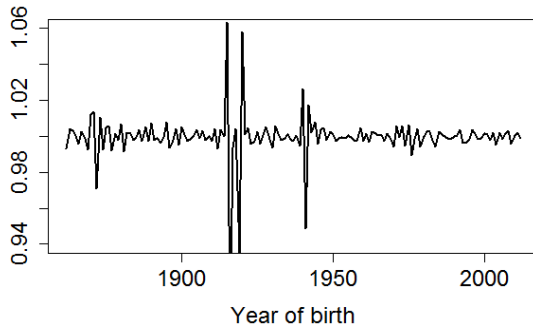
Dynamique des populations et correction des tables

Correction basée sur les données de fertilité : ex. France et Italie

Number of births by month (France)



Correction indicator (France)



Les données de fertilité à pas mensuel pour la France et l'Italie sont extraites à partir de la **Human Fertility Database**

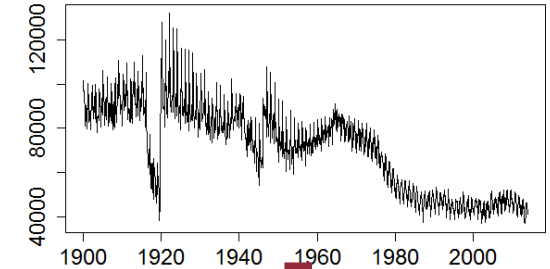


Le modèle de correction produit un **indicateur de qualité** à partir de données de fertilité à pas mensuel → l'indicateur quantifie, pour chaque génération, l'**ordre de grandeur des anomalies** causées par des chocs dans les séries de fertilité

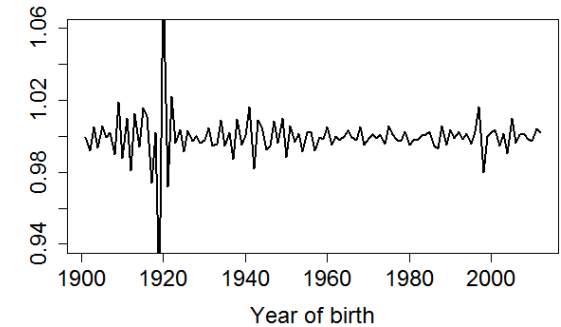


L'indicateur est ensuite utilisé pour produire les **tables corrigées**

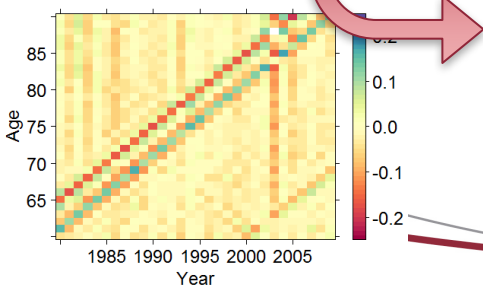
Number of births by month (Italy)



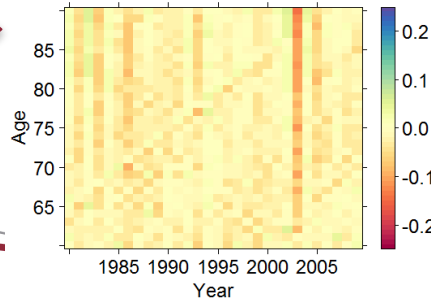
Correction indicator (Italy)



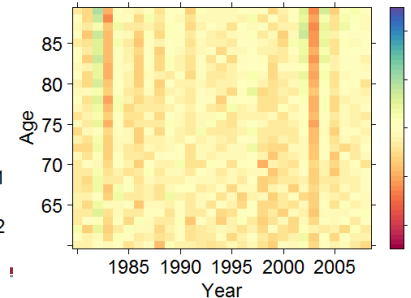
Mortality improvements (France)



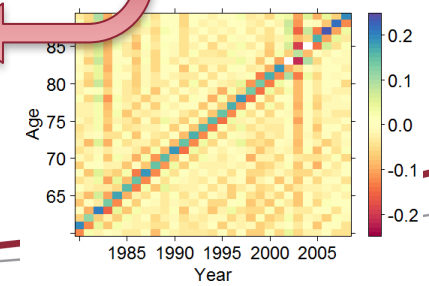
Mortality improvements (France)



Mortality improvements (Italy)



Mortality improvements (Italy)



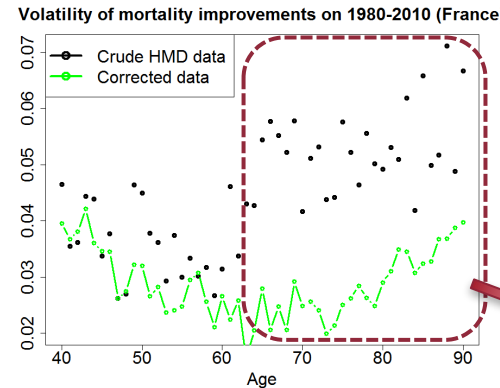
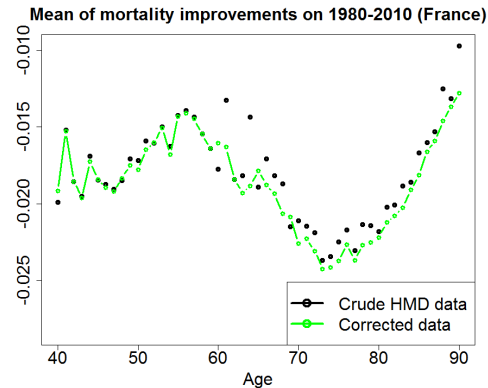
1. Introduction
2. Les tables périodiques de la Human Mortality Database
3. Dynamique des populations et correction des tables de mortalité
- 4. Quelles conséquences pour l'assurance ?**

Quelles conséquences pour l'assurance ?

Une volatilité historique réduite

- Suite à la correction des tables de mortalité, la **volatilité historique** des taux d'amélioration sur les trente dernières années présente une **réduction significative** :

La **moyenne** historique des taux d'amélioration est **lissée en âge**

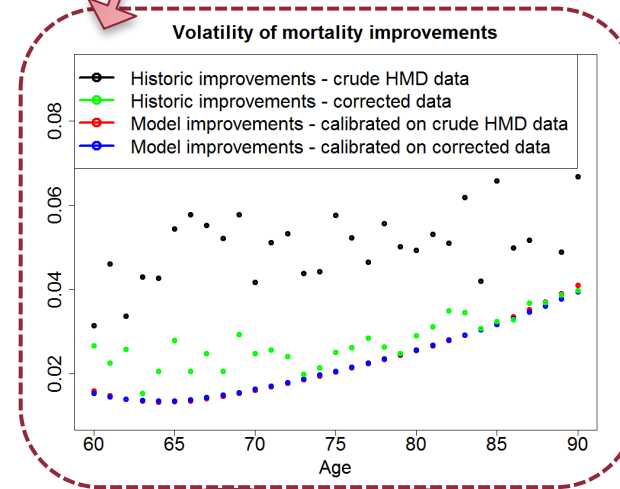


La **volatilité historique** des taux d'amélioration est **significativement réduite**

Les approches rétrospective et prospective **coïncident sur données corrigées**

- Une étude détaillée a montré la **réconciliation des visions prospective et rétrospective**, par ailleurs analysées de façon différente dans les études règlementaires
 - Voir** : CEIOPS' Advice for Level 2 Implementing Measures on Solvency II // Annex B Longevity risk calibration analysis

3.44. The results of this analysis indicated that, on average (across the nine countries for which data was analysed), historic improvements in mortality rates over 15 years from 1992 to 2006 were higher than 25%. Although the results of the stochastic model of future mortality improvements may imply a lower stress, CEIOPS has attached more weight to the analysis of historic improvements because of the significant uncertainty inherent in modelling mortality.



- Illustration de la problématique sur la base du modèle suivant (proche M7) :

$$\ln(\mu(x, t)) = \kappa_1(t) + \kappa_2(t)(x - \bar{x}) + \kappa_3(t) \left((x - \bar{x})^2 - \widehat{\sigma_x^2} \right) + \epsilon(x, t)$$

Moyenne sur la plage d'âges
considérée

Moyenne de la partie quadratique
 $(x - \bar{x})^2$ sur la plage d'âges considérée

Résidus centrés

- Ce choix de modèle illustratif est motivé par **différentes propriétés** :
 - Ce modèle tire parti de la log-linéarité du taux de mortalité aux âges élevés avec ajustement de la forme (légèrement) quadratique => le **nombre de paramètres en jeu est raisonnable**
 - La partie quadratique permet de répliquer les **moyennes historiques** des taux d'amélioration
 - Il n'inclut **pas de paramètre de cohorte**, ce qui va permettre une analyse des résidus
- Le calibrage est réalisé sur la période 1980-2010 et la plage d'âges 60-90

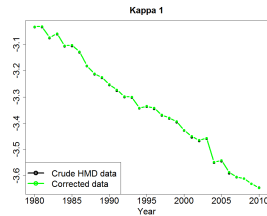
Quelles conséquences pour l'assurance ?

Choix et calibrage des modèles de mortalité (2/2)

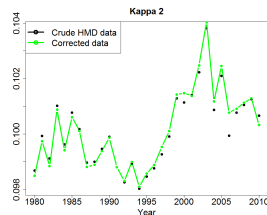
Conclusion 1

Les **paramètres estimés** sont proches

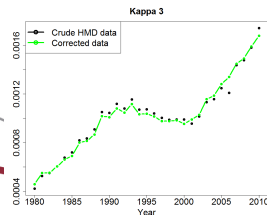
$\kappa_1(t)$



$\kappa_2(t)$



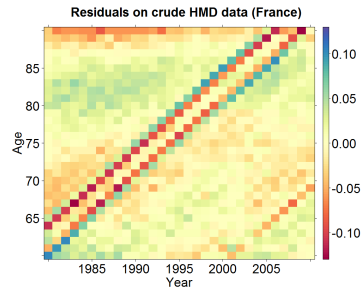
$\kappa_3(t)$



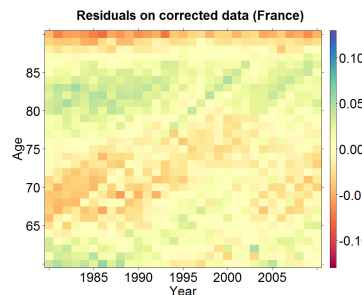
Conclusion 2

Les **résidus** ne présentent plus de tendance cohorte marquée

– Sur données brutes :



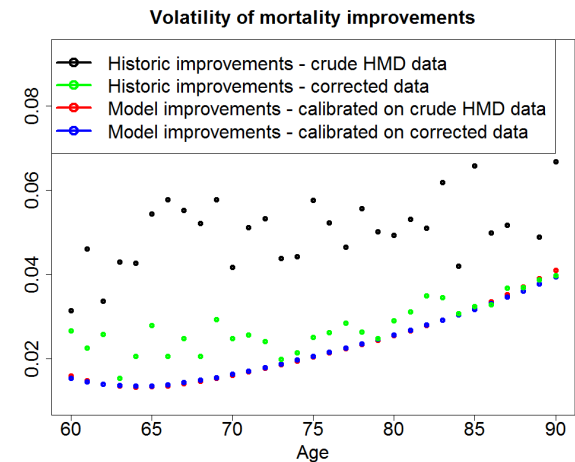
– Sur données corrigées :



Conclusion 3

La **volatilité historique** est davantage répliquée

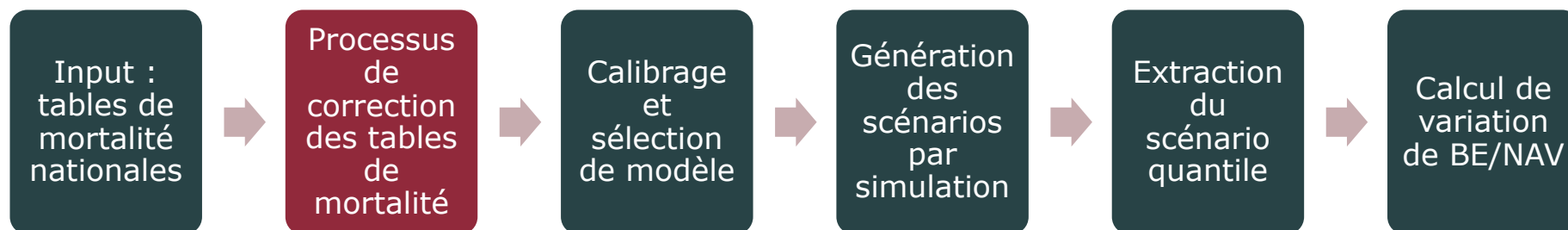
– La volatilité historique des taux d'amélioration est mieux répliquée sur la base des données corrigées :



Quelles conséquences pour l'assurance ?

Intégration des données corrigées dans les modèles internes

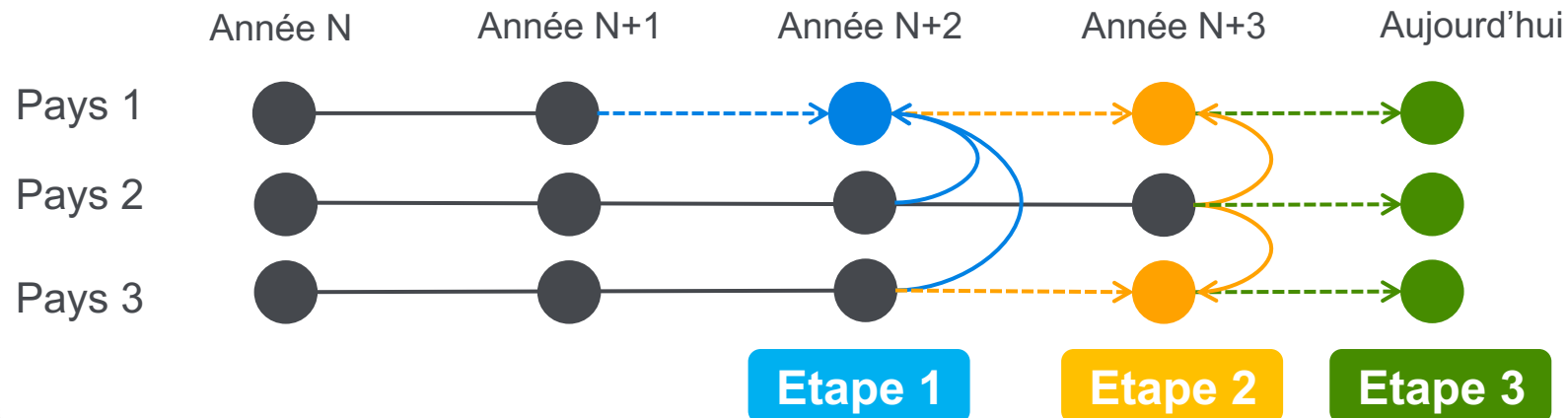
- L'intégration de nouvelles données s'accompagne de différents enjeux opérationnels :
 - Implémentation au sein des outils existants
 - Tests d'impacts sur le périmètre complet
 - Processus de validation du changement de modèle, d'audit (interne / externe)
- Ces corrections s'insèrent **en amont** du processus de calcul du SCR :



- *Issu de: La qualité des données de mortalité dans le cadre d'un modèle interne Solvabilité 2 : diagnostic, retraitements et mesures d'impacts, A. Boumezoued, L. Devineau (Milliman R&D), M. Habart, F. Balland et T. Popa (AXA GRM), présenté au 'Congrès des Actuaires', 23 juin 2017*

Un nouveau défi: le délai de publication des tables

- **Objectif:** proposer une estimation robuste de données de mortalité non encore disponibles
- **Idée:** tirer parti de l'information disponible pour d'autres pays
 - Ceci fait appel aux modèles stochastiques multi-population
 - Le graphique ci-dessous illustre le fonctionnement de l'algorithme de complétion



- **Conséquences opérationnelles:**
 - Permet de disposer **d'hypothèses biométriques à jour** pour un pays d'intérêt
 - Fournit un **cadre cohérent** pour des acteurs dont l'exposition porte sur plusieurs pays

1. Arnold, S., A. B., H. Labit Hardy, N. El Karoui. 2015. **Cause-of-death mortality: What can be learned from population dynamics?** To Appear in *Insurance: Mathematics and Economics*. HAL preprint: <https://halshs.archives-ouvertes.fr/hal-01157900/>
2. H. Bensusan, A. B., N. El Karoui, and S. Loisel. 2015. **Bridging the gap from microsimulation practice to population models: a survey.** Preliminary version as Chapter 2 of A. B. PhD thesis available at <https://tel.archives-ouvertes.fr/tel-01307921/document>.
3. A. B., N. El Karoui, S. Loisel. 2015. **Measuring mortality heterogeneity dynamics with interval-censored data.** To Appear in *Insurance: Mathematics and Economics*. HAL preprint: <https://hal.archives-ouvertes.fr/hal-01215350/>
4. A. B. 2016. **Improving HMD mortality estimates with HFD fertility data.** To Appear in the *North American Actuarial Journal*.. HAL preprint: <https://hal.archives-ouvertes.fr/hal-01270565v1>
5. A. B., L. Devineau. 2017. **Enjeux de fiabilité dans la construction des tables de mortalité nationales.** L'Actuariel, janvier 2017.
6. A.B., with L. Devineau, M. Habart, F. Balland and T. Popa. **The impact of mortality data anomalies on an internal model,** presented at 'Congrès des Actuaires', 23 juin 2017
7. A. B., M. Hoffmann, P. Jeunesse. 2017. **A new inference strategy for general population mortality tables.** *Working paper*
8. A. B., M. Hoffmann, P. Jeunesse. 2017. **Non-parametric inference for in-homogeneous and age-dependent population processes.** *Working paper*
9. Mémoire d'Actuariat de N. El Bekri, 2017. **Mise à jour des tables de mortalité par approche multi-population. Application au calcul de choc et au transfert du risque de longévité.**