

Expertise « Validierung des Modellansatzes zur Berechnung und Projektion der Einnahmen und Ausgaben der AHV »

**Mandat réalisé pour le compte de
l'Office fédéral des assurances sociales**

par

Laurent Donzé

Fribourg, le 20 août 2018

Table des matières

Table des matières	iii
Table des figures	v
Remerciements	1
Introduction	3
1 Modélisation des droits à une rente et des sommes des rentes	5
1.1 Les catégories d'intérêt et notation	5
1.2 Le calcul des Équivalents Pleine Rente Cumulés (EPRC)	5
1.3 Projection annuelle des EPRC	6
1.4 Le calcul des Somme des rentes (SR)	6
1.5 Commentaire	8
2 Implémentation en R	11
2.1 La gestion de projet RStudio	11
2.2 Le script <code>getting_started.R</code>	11
2.3 Les dépenses et les recettes	14
3 Évaluation finale	15
3.1 Modélisation	15
3.2 Programmation	15
3.3 Documentation	16
Acronymes	17
Notation	19
Bibliographie	21



Table des figures

1.1	Algorithme pour la projection annuelle des EPRC	7
2.1	Flux du script <code>getting_started.R</code>	13



Remerciements

Nous avons bénéficié tout au long de cette expertise de l'aide, des explications et des commentaires de Madame Loyal Pipoz, collaboratrice auprès de l'**Office fédéral des assurances sociales (OFAS)**, section « Mathématique ». Madame Pipoz nous a remis la documentation nécessaire, les scripts R et les données utilisées dans les calculs. Nous avons en outre rencontré pour une première présentation de ce rapport le responsable de la section, Monsieur le Dr Thomas K. Friedli, et son équipe. Leur collaboration a été très appréciée. Je tiens à remercier vivement toutes ces personnes.



Introduction

Dans le cadre de son projet « Phoenix », la section « Mathématique » de l'OFAS est chargée entre autres du développement et de la réimplémentation du calcul du budget de l'AVS à court et long terme. Celui-ci a été récemment implémenté avec le logiciel R. Ce nouvel outil permet de manière flexible la projection des sommes des rentes à longue échéance. Dans l'optique d'une nouvelle réforme du système de financement de l'AVS, il est donc primordial de disposer d'estimations fiables du budget de l'AVS.

Sur ces considérations, l'OFAS a décidé de faire valider sa modélisation des rentes par des experts externes. Nous avons été mandaté à cet effet. En particulier, il s'agit de valider le modèle de calcul à long terme des dépenses et des recettes selon :

1. la complétude du modèle ;
2. la paramétrisation et la plausibilité des hypothèses du modèle ;
3. l'implémentation en code informatique, notamment du point de vue de la cohérence entre la documentation et le code.

Pour répondre à notre mandat, dans un premier temps et pour mieux la comprendre, nous détaillons la modélisation adoptée. Nous décrivons ensuite l'implémentation qui en a été faite en code informatique. Enfin, nous donnons une évaluation globale.



Chapitre 1

Modélisation des droits à une rente et des sommes des rentes

Dans un souci de compréhension, sur la base de [Pipoz \(2018\)](#), nous reformulons en nos termes la modélisation du nombre de droits à une rente complète dès l'âge de la retraite, désigné également comme **Équivalents Pleine Rente Cumulés (EPRC)**, et la **Somme des rentes (SR)**. Ces quantités dépendent de catégories d'intérêt que nous décrirons dans un premier temps.

1.1 Les catégories d'intérêt et notation

Notons de manière générale par $c \in \mathcal{C}$ une catégorie particulière. Considérons par ailleurs les facteurs `sexe`, `nationalité`, `domicile`, `état-civil`, `genre_de_rente`. Les modalités de ces facteurs prennent leurs valeurs respectivement dans les ensembles \mathcal{S} , \mathcal{N} , \mathcal{D} , \mathcal{E} et \mathcal{G} . Par exemple, les modalités du facteur `sexe` sont données par $\mathcal{S} = \{m, f\}$, où m et f désignent respectivement un homme ou une femme. Ainsi, une catégorie d'analyse c peut-être un des éléments des ensembles ou un élément du produit cartésien de deux ou plusieurs de ces ensembles, e.g. $\mathcal{S} \times \mathcal{N} \times \mathcal{D} \times \mathcal{E} \times \mathcal{G}$.

D'autre part, désignons par α la catégorie d'âge traitée. Pour un homme, $\alpha = 21, 22, \dots, 99$ et pour une femme, $\alpha = 20, 21, \dots, 99$. Nous désignerons par $\alpha_0^c \in \{62, \dots, 70\}$ l'âge observé de départ à la retraite. L'indice t va représenter l'année.

1.2 Le calcul des EPRC

Pour $c \in \mathcal{S} \times \mathcal{N} \times \mathcal{D}$, l'année t et la catégorie d'âge α , soient $K_{\alpha,t}^c$ le nombre d'EPRC, $V_{\alpha,t}^c$ le nombre de « nouveaux » **Équivalents Pleine Rente (EPR)** (arrivants) et $q_{\alpha,t}^c = 1 - p_{\alpha,t}^c$ le



taux de mortalité. Les **EPRC** $K_{\alpha,t}^c$ sont calculés par la formule (1.1)

$$K_{\alpha,t}^c = \begin{cases} V_{\alpha,t}^c, & \text{pour } t = 1971, \\ \alpha = 20 & \text{pour les femmes et} \\ \alpha = 21 & \text{pour les hommes;} \\ p_{\alpha-1,t-1}^c K_{\alpha-1,t-1}^c + V_{\alpha,t}^c & \text{pour } t > 1971. \end{cases} \quad (1.1)$$

Les quantités $K_{\alpha,t}^c$ dépendent via $V_{\alpha,t}^c$ de la statistique de la population (**Enquête sur la structure de la population (ESPOP)**, **Statistique de la population et des ménages (STATPOP)**). Celle-ci n'est disponible qu'à partir de 1971. L'équation (1.1), vu sa structure dynamique, donnera donc la possibilité de faire des projections.

Le **Registre des rentes (RR)** permet de calculer le nombre de rentiers. Une échelle allant de 1 à 44 indique le nombre d'années de cotisation, au maximum 44, d'un rentier ou d'une rentière. Légalement, un homme peut cotiser de 21 à 65 ans et une femme de 21 à 64 ans. Dans un souci de simplification et pour avoir la même échelle entre les hommes et les femmes, on considère qu'une femme peut cotiser à partir de 20 ans. Le volume des entrants $V_{\alpha,t}^c$ est déterminé à partir de la statistique de la population (**ESPOP**, **STATPOP**) selon la formule suivante :

$$V_{\alpha,t}^c = bw_{\alpha,t}^c \cdot n_{\alpha,t}^c, \quad t = 1971 + n; \quad n = 0, 1, 2, \dots \quad (1.2)$$

où $bw_{\alpha,t}^c$ est le facteur de pondération suivant :

$$bw_{\alpha,t}^c = \max \left(\frac{1}{44}, \frac{1}{44 - \alpha + t - 1927} \right), \quad \begin{array}{l} \alpha \in [20, \dots, 64] \text{ pour les femmes et} \\ \alpha \in [20, \dots, 65] \text{ pour les hommes,} \end{array} \quad (1.3)$$

et $n_{\alpha,t}^c$ est le nombre d'individus de la catégorie c , d'âge α en période t selon les statistiques de la population.

1.3 Projection annuelle des **EPRC**

Définissons tout d'abord le facteur annuel de croissance des **EPRC**, noté $\omega_{\alpha,t}^c$, comme :

$$\omega_{\alpha,t}^c = \frac{K_{\alpha,t}^c}{K_{\alpha,t-1}^c}, \quad (1.4)$$

où $K_{\alpha,t}^c$ est calculé selon l'équation (1.1). La projection annuelle des **EPRC** doit tenir compte de l'âge observé α_0^c de départ à la retraite. On considérera également l'année considérée t_{rr} du **RR** et notamment pour cette année les $\tilde{K}_{\alpha,t_{rr}}^c$ déduits du registre. Les taux de croissance $\omega_{\alpha,t}^c$ et de mortalité $q_{\alpha,t}^c$ vont nous permettre d'ajuster les **EPRC**. Cette projection va se faire pour les années 2016 à 2065 selon l'algorithme 1.1 décrit ci-dessous.

Algorithme 1: Projection annuelle des EPRC**Require:**

- α_0^c : âge observé de départ à la retraite ($\alpha_0^c \in \{62, \dots, 70\}$);
- t_{rr} : année considérée du RR;
- $\tilde{\mathbf{k}}^c$: vecteur des $\tilde{K}_{\alpha, t_{rr}}^c$, $\alpha = 62, \dots, 99$;
- ω^c , vecteur des $\omega_{\alpha_0^c, t}^c$, $t = t_{rr} + 1, \dots, 2065$;
- \mathbf{Q}^c : matrice des taux de mortalité $q_{\alpha, t}^c$.

1: **procedure** PROJECTION($\alpha_0^c, t_{rr}, \tilde{\mathbf{k}}^c, \omega^c, \mathbf{Q}^c$)

2: **Start**[Calcul des $\hat{K}_{\alpha, t}^c$]

3: $\hat{K}_{\alpha, t_{rr}}^c \leftarrow \tilde{K}_{\alpha, t_{rr}}^c, \forall \alpha$

4: **for** $t \leftarrow \{t_{rr} + 1, \dots, 2065\}$ **do**

5: **for** $\alpha \leftarrow \{62, \dots, 99\}$ **do**

6: **if** $\alpha = \alpha_0^c$ **then**

7: $\hat{K}_{\alpha, t}^c \leftarrow \omega_{\alpha_0^c, t}^c \hat{K}_{\alpha, t-1}^c$

8: **else**

9: $\hat{K}_{\alpha, t}^c \leftarrow (1 - q_{\alpha-1, t-1}^c) \hat{K}_{\alpha-1, t-1}^c$

10: **end if**

11: **end for**

12: **end for**

13: **End**

14: **end procedure**

FIGURE 1.1 – Algorithme pour la projection annuelle des EPRC

1.4 Le calcul des SR

Pour le calcul de la **Somme des rentes (SR)**, on utilisera comme point de départ les informations du **RR** en vigueur. Notons par $\widetilde{SR}_{\alpha, t_{rr}}^c$, cette somme des rentes pour la catégorie d'âge α . Désignons par $a_{\alpha, t}^c$ le coefficient d'ajournement ou d'anticipation de départ à la retraite. Ce coefficient peut s'écrire comme :

$$a_{\alpha, t}^c = \begin{cases} \left(\frac{1+s_{\alpha, t}}{1+s_{\alpha, t_{rr}}} \right), & \alpha \in \{62, 63, \dots, 70\}, \\ 1 & \text{sinon,} \end{cases} \quad (1.5)$$

où les taux $s_{\alpha, t}$ sont fixés dans la loi.¹ D'autre part, la quantité $\hat{g}_{\alpha, t}^c$ estime la croissance des **EPRC** en l'année t par rapport à l'année t_{rr} et est formulée selon 1.6 :

$$\hat{g}_{\alpha, t}^c = \frac{\widehat{K}_{\alpha, t}^c}{\widetilde{K}_{\alpha, t_{rr}}^c}, \quad t \geq t_{rr}, \alpha \geq \alpha_0^c, \quad (1.6)$$

où les $\widehat{K}_{\alpha, t}^c$ ont été trouvés par l'algorithme 1.1 et où $\widetilde{K}_{\alpha, t_{rr}}^c$ sont les **EPRC** observés dans le registre ; $\alpha \geq 62$. Le facteur $\hat{g}_{\alpha, t}^c$ permettra d'ajuster $\widetilde{SR}_{\alpha, t_{rr}}^c$. Enfin, on notera par $\kappa_{\alpha, t}^c$, un facteur dit « empirique » d'ajustement.²

La somme des rentes estimées $\widehat{SR}_{\alpha, t}^c$ pour la catégorie d'âge $\alpha \geq 62$, en période t , est donnée par :

$$\widehat{SR}_{\alpha, t}^c = \hat{g}_{\alpha, t}^c \kappa_{\alpha, t}^c \widetilde{SR}_{\alpha, t_{rr}}^c, \quad t > t_{rr}, \alpha \geq \alpha_0^c. \quad (1.7)$$

Si l'on considère toutes les catégories d'âge, on obtiendra la **SR** en période t pour la catégorie c par :

$$\widehat{SR}_t^c = \sum_{\alpha=62}^{99} a_{\alpha, t}^c \widehat{SR}_{\alpha, t}^c, \quad (1.8)$$

et la **SR** totale à la période t comme :

$$\widehat{SR}_t = \sum_c \widehat{SR}_t^c. \quad (1.9)$$

1.5 Commentaire

Le but final de la modélisation est la projection de la **SR**. Comme nous l'avons constaté dans les sections précédentes, ce calcul passe successivement par une projection des **EPRC**. Cela permet de définir un facteur de croissance de la somme des rentes. Cette somme des rentes est encore ajustée par différents facteurs pour tenir compte notamment des ajournements ou des anticipations des départs à la retraite. Un facteur particulier lié à un rapport

1. Actuellement, $a_{\alpha, t}^c = 1$. Aucun facteur de correction n'est appliqué.

2. Le facteur $\kappa_{\alpha, t}^c$ est exprimé par la valeur de la fonction de densité $f_X(x_t)$ d'une loi chi-carrée à 3 degrés de liberté où x_t est une quantité qui dépend d'un rapport d'indices de production industrielle.

d'indices de production industrielle est intégré au modèle. Ce dernier terme est tout à fait heuristique et mériterait une meilleure justification théorique.

La modélisation est en soi cohérente. Bien que simple dans sa conception, elle permet des projections plausibles à plus ou moins long terme. Elle utilise à bon escient les informations dont on dispose. Cette formulation permet en outre d'intégrer très rapidement différents scénarios et de les tester. Elle constitue donc un outil bien approprié.

D'un point de vue statistique, la limite d'une telle approche est sa non-stochasticité. La modélisation ne permet pas de fournir des estimateurs avec leur variance. Des intervalles de confiance pour les quantités projetées ne sont donc pas disponibles. Cependant, une modélisation statistique des **SR** n'est pas évidente et est un projet à part entière. Différentes pistes pourraient être possibles,³ notamment dans la modélisation des taux de croissance ou d'ajustement.

3. Cf. par exemple [Leitner \(2018\)](#).

Chapitre 2

Implémentation en R

Le calcul des **EPRC** et des **SR** a été implémenté dans un environnement R (**R Core Team (2018)**). L'**OFAS** nous a remis deux scripts R (`getting_started.R` et `prepare_data.R`), des fonctions ad hoc, des données, le tout permettant de reproduire ces quantités. Les lignes de code nous ont été mises à disposition dans un projet de type « **RStudio** ».

2.1 La gestion de projet RStudio

L'**OFAS** a décidé de travailler dans un éco-système RStudio. Le programme RStudio permet la gestion de projets et le développement de logiciels dans un environnement propre, notamment en utilisant la plate-forme « github ». L'organisation des données en tables `tidy` fait également partie du concept.¹ Il s'agit d'une façon moderne de procéder qui convient certainement à une équipe de développeurs et est bien adaptée en l'occurrence.

Le choix d'évoluer en mode `tidy` permet une structuration des données qui serait très profitable en termes de clarté et d'efficacité. Cette façon de faire est donc une bonne option.

2.2 Le script `getting_started.R`

Il s'avère que le but des scripts R `getting_started.R` et `prepare_data.R` est le même. Nous ferons donc référence uniquement à `getting_started.R`. Celui-ci démarre par lire un objet appelé `PARAM_GLOBAL` dans lequel sont stockés tous les paramètres nécessaires au calcul et à la simulation. Cela permet une grande flexibilité puisque si l'on veut tester un scénario particulier il suffit de ne changer que l'une ou l'autre valeur de la liste des paramètres.

Ensuite, le programme « source » un autre script R, appelé `mod_input_ahv_data.R` qui active toutes les fonctions ad hoc construites pour lire les données à traiter. Celles-ci sont sous formes d'objets `tidy` et de « `tidy lists` ». Cette première phase étant faite, on peut passer à la préparation des données pour les calculs.

On construit d'abord la population totale à partir des données relatives à la population résidente, à l'émigration, aux assurés facultatifs et aux frontaliers. Ensuite, à partir des « `Eckwerte` » — ensemble d'indices (salaires, prix) observés et projetés —, on génère les tables

1. On consultera par exemple [Müller et Wickham \(2018\)](#), [Wickham \(2014\)](#), [Wickham et Henry \(2018\)](#).



de l'évolution du revenu, du développement des rentes et des rendements nominaux. Enfin, on calcul le scénario pour le paramètre d'ajustement des premières rentes. L'enchaînement de ces opérations est décrit dans le diagramme des flux 2.1.

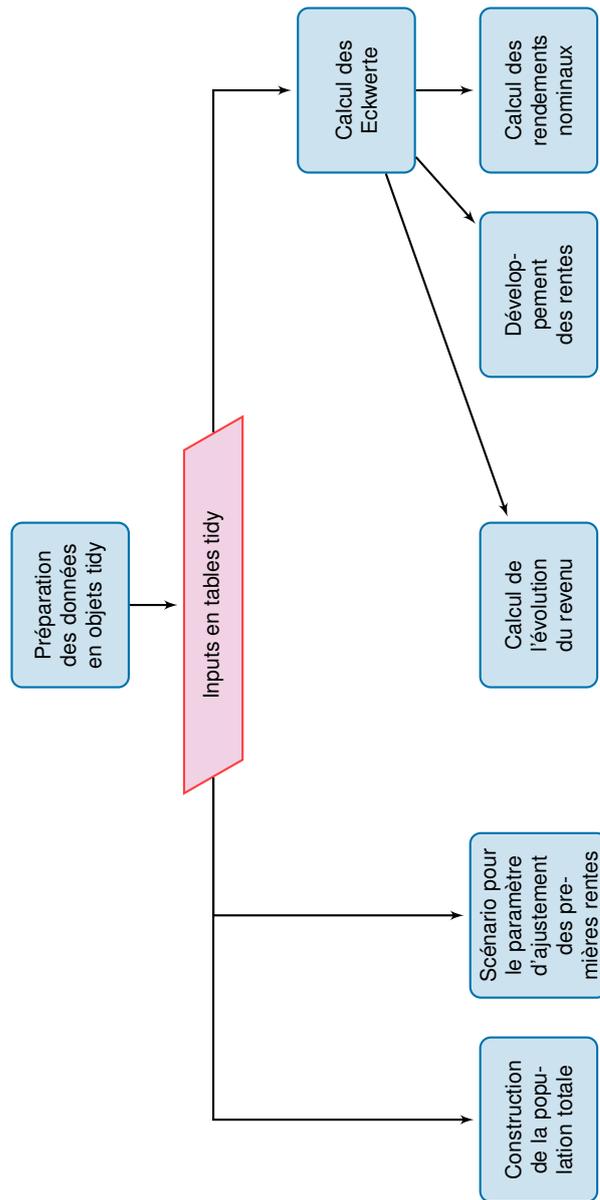


FIGURE 2.1 – Flux du script getting_started.R

2.3 Les dépenses et les recettes

Une librairie de fonctions R fait également partie de l'environnement RStudio mis à notre disposition. Ces fonctions vont permettre le calcul des dépenses et des produits liés à l'AVS. Elles dépendront en input des objets produits par les scripts précédemment évoqués.

Les dépenses sont données par la **SR**, les allocations pour impotents (« Hilfslosenschaedigungen ») et d'autres dépenses. La fonction `mod_ahv_ausgaben` génère le total des dépenses. Quant aux recettes, celles-ci proviennent des cotisations (registre des rentes) ainsi que d'autres recettes. La fonction `mod_ahv_einnahmen` nous permet ce calcul.

L'estimation des dépenses et des recettes va finalement permettre de prédire le résultat à long terme du système de financement de l'AVS. Celui-ci est généré par la fonction `mod_ahv_bilanz`. Une dernière étape dite de « postprocessing » permet finalement d'actualiser les résultats.

Chapitre 3

Évaluation finale

Notre expertise s'est concentrée sur deux points essentiels de la modélisation et du calcul des rentes AVS. Le premier est évidemment de juger de la cohérence et de la justesse de la modélisation. Le second est d'apprécier la transcription en code informatique des modèles développés.

3.1 Modélisation

Le chapitre 1 nous a permis, sur la base de la documentation que l'on nous a remise, de formuler dans nos termes la modélisation mathématique des EPRC et des SR. Cet exercice était nécessaire vu, d'une part, la complexité du problème et, d'autre part, la difficulté de le présenter de manière simple et cohérente.

L'approche choisie par l'OFAS est tout à fait naturelle et pertinente pour ce type d'analyse. En se basant sur l'évolution de la population et sur le registre des rentes, en se laissant la possibilité d'introduire différents paramètres selon les scénarios de développement et de politiques futures, l'OFAS s'est doté d'un modèle efficace, robuste, et durable.

3.2 Programmation

Il est impossible de vérifier tout le code produit. Nos contrôles ont porté essentiellement sur la logique de programmation, sur les soins apportés au développement et à l'écriture du code, sur l'efficacité de la programmation et bien évidemment sur la transcription sous forme de code informatique des développements mathématiques présentés au chapitre 1.

Nous avons déjà eu la possibilité de souligner que l'utilisation du logiciel R, notamment dans un environnement RStudio, est une option moderne qui permettra sans problème une évolution, des mises à jour et des développements ultérieurs. Ce choix est donc judicieux.

Nous avons analysé le code des fonctions principales décrites succinctement au chapitre précédent. Elles répondent aux standards de programmation en la matière. Nos contrôles ne nous ont pas permis de déceler de fautes ou d'erreurs évidentes. Les fonctions sont bien commentées de manière à en faciliter la lecture. De toute évidence, ces fonctions transcrivent en code les modèles mathématiques développés.



3.3 Documentation

La documentation est en générale écrite en « markup language » et produite sous différents formats (html, pdf). Celle-ci est relativement difficile à lire et n'est pas toujours cohérente. Un effort de présentation dans un document unique, structuré serait le bienvenu.

Acronymes

EPR	Équivalents Pleine Rente 5, 20
EPRC	Équivalents Pleine Rente Cumulés iii, v, 5–8, 11, 15, 20
ESPOP	Enquête sur la structure de la population 6
OFAS	Office fédéral des assurances sociales i, 3, 11, 15
RR	Registre des rentes 6–8, 19, 20
SR	Somme des rentes iii, 5, 8, 11, 14, 15
STATPOP	Statistique de la population et des ménages 6

Notation

t_{rr}	Année de référence du RR considéré.
α	Catégorie d'âge traitée.
α_0^c	Âge observé de départ à la retraite ($\alpha_0^c \in \{62, \dots, 70\}$).
\mathcal{C}	Ensemble des catégories c
\mathcal{D}	Ensemble des modalités du facteur domicile. $\mathcal{N} = \{\text{au}, \text{ch}\}$, où au et ch désignent respectivement « étranger » ou « suisse ».
\mathcal{E}	Ensemble des modalités du facteur état-civil. $\mathcal{E} = \{\text{geschieden}, \text{ledig}, \text{verheiratet}, \text{verwitwet}\}$.
\mathcal{G}	Ensemble des modalités du facteur genre_de_rente. $\mathcal{G} = \{\text{rcompl_femme}, \text{renfant_mere_simple}, \text{renfant_pere_simple}, \text{rorphelin_double}, \text{rorphelin_mere_simple}, \text{rorphelin_pere_simple}, \text{rveuve}, \text{rvieillesse_simple}\}$.
\mathcal{N}	Ensemble des modalités du facteur nationalité. $\mathcal{N} = \{\text{au}, \text{ch}\}$, où au et ch désignent respectivement « étranger » ou « suisse ».
\mathcal{S}	Ensemble des modalités du facteur <code>sexe</code> . $\mathcal{S} = \{\text{m}, \text{f}\}$, où m et f désignent respectivement un homme ou une femme.

$a_{\alpha,t}^c$	Facteur d'ajournement ou d'anticipation de départ à la retraite.
$\omega_{\alpha,t}^c$	Facteur annuel de croissance des EPRC.
$\hat{g}_{\alpha,t}^c$	Facteur annuel de croissance estimé des EPRC.
$bw_{\alpha,t}^c$	Facteur de pondération.
$n_{\alpha,t}^c$	Nombre d'individus de la catégorie c , d'âge α , en période t .
$V_{\alpha,t}^c$	Nombre de nouveaux EPR pour la catégorie d'âge α au temps t .
$K_{\alpha,t}^c$	Nombre EPRC pour la catégorie d'âge α au temps t . On désignera par $\tilde{K}_{\alpha,t}^c$ les unités déduites du RR et par $\hat{K}_{\alpha,t}^c$ celles qui ont été estimées.
$\widehat{SR}_{\alpha,t}^c$	Somme des rentes estimées.
$q_{\alpha,t}^c$	Taux de mortalité pour la catégorie d'âge α au temps t .
$p_{\alpha,t}^c$	Taux de survie pour la catégorie d'âge α au temps t .

Bibliographie

- [1] Leitner, R. 2018, *Stochastische Modellierung der Schweizerischen AHV*, mémoire de maîtrise, ETH Zürich.
- [2] Müller, K. et H. Wickham. 2018, *tibble : Simple Data Frames*. URL <https://CRAN.R-project.org/package=tibble>, r package version 1.4.2.
- [3] Pipoz, L. C. 2018, «Modèle des sommes de rentes», techreport, Office fédéral des assurances sociales. Division Mathématiques, analyses et statistiques, Effingerstrasse 20, CH-3003 Bern, Switzerland.
- [4] R Core Team. 2018, *R : A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [5] Wickham, H. 2014, «Tidy data», *The Journal of Statistical Software*, vol. 59. URL <http://www.jstatsoft.org/v59/i10/>.
- [6] Wickham, H. et L. Henry. 2018, *tidyr : Easily Tidy Data with 'spread()' and 'gather()' Functions*. URL <https://CRAN.R-project.org/package=tidyr>, r package version 0.8.1.

