A semi-Markov model to investigate the different transitions between states of dependency in elderly people

Vincent LEPEZ – Svetlana ROGANOVA – Antoine FLAHAULT*

AAI Colloquium – Lyon
June 25, 2013

Insuring Long Term Care for elderly people: a fast growing need of our society

Modeling the LTC risk

- ***Penalized actuarial tables and markov approach***
- ***semi-markov approach***

Estimating the parameters of a semi-markov model

- ***The data, the entry into dependency state and model dimension***
- ***Weibull - Linear model***
- ***Non-parametric approach and first results***
- ***Mixed Weibull – Linear model and final estimation results***

Application to a model-related insurance product design

- ***Mortality tables***
- ***Actuarial tables***

Perspectives

Insuring Long Term Care for elderly people: a fast growing need of our society (1 / 2)

The number of people aged 75 and more is about to triple until 2050.

At the same time, number of elderly dependant people will double.

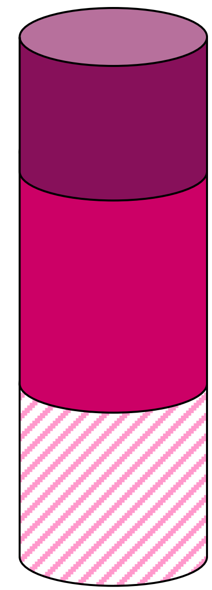
In France, public social insurance type of coverage is worth 20 G€ / year, or 1% GDP.

On average, the overall care costs for an elderly dependent person are 30 000 € / year.

The APA (social scheme) is worth a fifth of the need

An other 2 to 3 fifth comes from the revenues of the houshold

The rest is not covered



Overall, Long term Care is an emerging risk with foreseable extremely high costs…

…and only partially covered.

Insuring Long Term Care for elderly people: a fast growing need of our society (2 / 2)

The insurance industry must provide solutions, whether social economy, provident societies or private sector.

At the same time and due to financial stakes, one needs to develop a refined knowledge of the insurable risk

The objectives of this study are

- **providing a robust and detailed modeling of the survival process into dependency state.**

  we will pay particular attention to states transitions and their length;

- **building up annuities insurance products for LTC needs coverage.**

  the associated benefits will be adapted to the underlying LTC states.

The progressive french GIR grid is our reference for measuring LTC states and needs.

Modéling the risk : first approaches

The « old » penalized actuarial tables SCOR model

entry into total dependency state at age *a* > 60 yo:

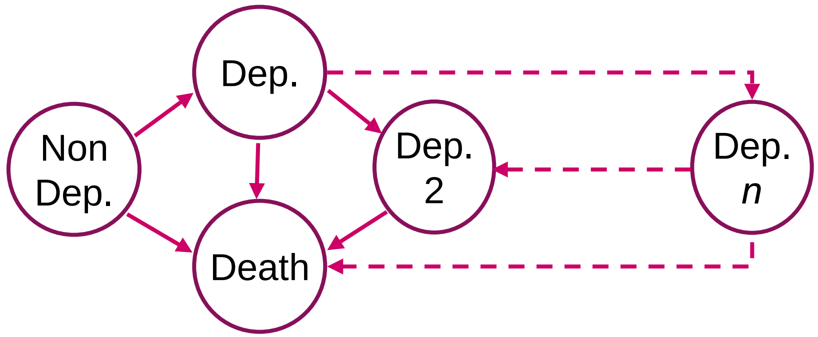
$$e_a = 0{,}0015 \times \exp(0{,}125 \times (a - 60))$$

mortality of non-dépendants:

$$q_a^{ND} = 0{,}8 \times q_a^{TD8890}$$

mortality of dépendants:

$$q_a^{ND} = 2 \times q_a^{TD8890} + 3{,}5\ \%$$



Markov model

Let *E* := { 0 ; 1 ; … ; *n* } be a set of states and *X* un process lying in *E*.

For *i* and *j* in *E* and 0 < *s* < *t*, the transition probability between *i* and *j* states, between *s* and *t* dates is

$$p_{i,j}(s,t) = P(X(t) = j | X(s) = i).$$

Let 0 < t_1 < … < t_n be the moments where the process jumps into the i_1 , …, i_n states, the transitions must only depend on the last previously visited state:

$$P(X(t_{n+1}) = i_{n+1} | X(t_1) = i_1, \ldots, X(t_n) = i_n) = P(X(t_{n+1}) = i_{n+1} | X(t_n) = i_n).$$

Problem : **the transition duration laws are then directly linked to the transition probabilities**

Modeling the risk : the semi-Markov approach (1 / 2)

The main goal of the semi-Markov approach is to have
transition probabilities and duration of transition probability laws made independant.

Transition durations are as fundamental for the insurer as the transition probabilities themselves because
- dependency level is driven by transition probabilities, hence annuitiy level;
- transition durations drive the length of payment of the annuity on a given level.

Hence, the present value of annuities, which is a critical input for pricing, must be constructed from a model able to precisely determine those two components.

Definitions:

- J_n : variable aléatoire discrète à valeurs dans l'espace d'états E égale au $n^{\text{ème}}$ état visité par le processus ;
- S_n : variable aléatoire continue positive égale à la date du $n^{\text{ème}}$ saut du processus ;
- X_n : variable aléatoire continue positive égale à la durée de station dans l'état J_n ;
- $N(s)$: variable aléatoire discrète égale aux nombres de transitions intervenues avant la date s.

For *i* and *j* in *E* and 0 < *s* < *x*, one defines the semi-markov kernel *Q* between *i* and *j* and between *s* and *x* by

$$Q_{i,j}(s,x) = P(J_{N(s)+1} = j, X_{N(s)+1} \leq x | J_{N(s)} = i, S_{N(s)} = s)$$

Modeling the risk : the semi-Markov approach (2 / 2)

Characteristics of the semi-Markov model:

It is a generalization of the standard Markov model :

$$p_{i,j}(s) = \lim_{x\to\infty} Q_{i,j}(s,x) = P(J_{N(s)+1} = j | J_{N(s)} = i, S_{N(s)} = s).$$

The law of transition length from state *i*, entered in in *s* and state *j*, for any *x* > 0 *is*

$$F_{i,j}(s,x) = P(X_{N(s)+1} \leq x | J_{N(s)+1} = j, J_{N(s)} = i, S_{N(s)} = s).$$

Fundamental assumption and property:

Under the asumption that the time spent in a given state and the entry date into it are independant

$$Q_{i,j}(s,x) = F_{i,j}(x) \times p_{i,j}(s).$$

Hence, to build-up of the *Q* kernel, one only needs

- the transition durations laws between states ;

- the functional forms of the proability transitions between states.

Practically speaking, the underlying models for the two can be parametric or not.

Estimation: the data

A fundamental source : the APA data

The DREES granted us access to compulsory APA data for 4 french « départements »

- 04 : Alpes de Haute Provence ;

- 71 : Saône et Loire ;

- 76 : Seine Maritime ;

- 95 : Val d'Oise.

They consist in the complete trajectories of dependant 52 000 dependent people between 1st Jan 2002 and 31st Dec 2005.

More than 27 000 GIR transitions are observed.

A secondary source : the CPRPSNCF data

The CPRPSNCF built up a dependency social coverage since 1st Oct 2005.

Data regarding 6 500 elderly dependent people are observed over 15 months.

But too few GIR transitions are actually observed.

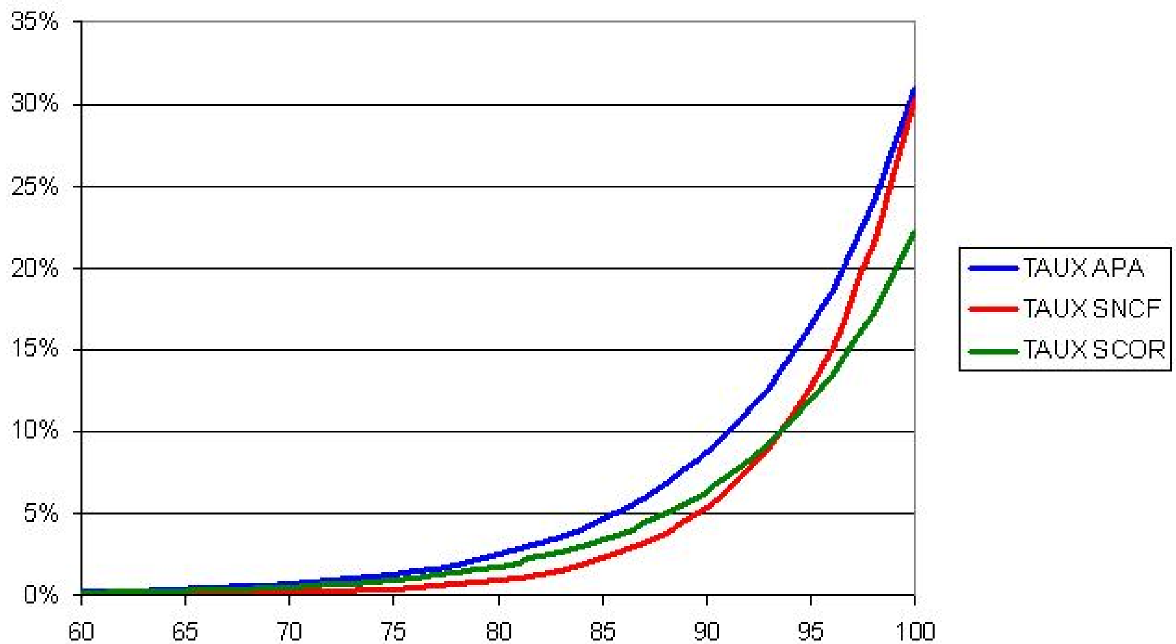
Estimation : entry into dependancy

Entry into dependency by GIR state is estimated by the measured empirical frequency

An exponential smoothing of these frequencies, all GIR levels compounded, allows us to compare measured frequencies to the genuine SCOR model.

For a given age *a* > 60 ans, one gets a entry into dependency rate e_a :

- **SCOR :** $e_a = 0{,}0015 \times \exp(0{,}125 \times (a - 60))$;
- **APA :** $e_a = 0{,}0020 \times \exp(0{,}126 \times (a - 60))$;
- **SNCF :** $e_a = 0{,}0003 \times \exp(0{,}173 \times (a - 60))$;

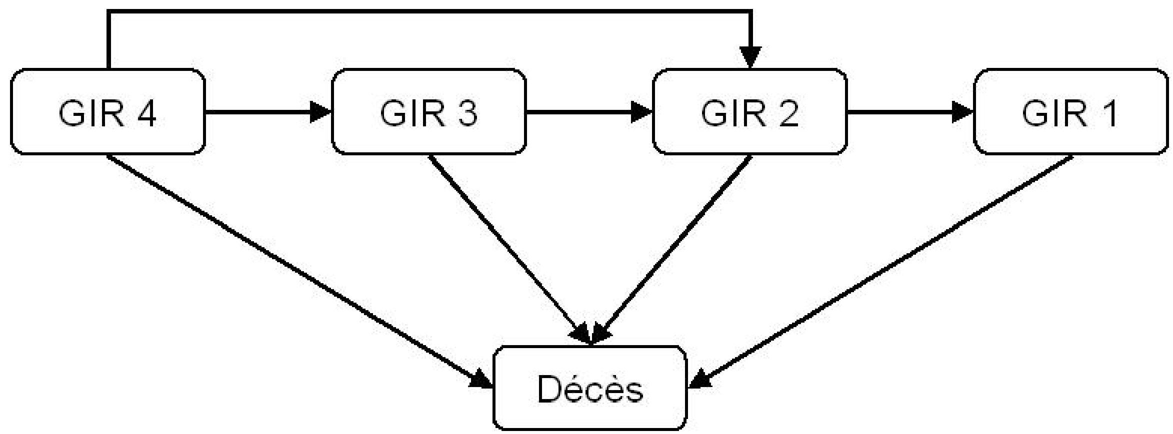


Estimation : reduction of model dimension

Observed frequencies of state transitions in the APA data:

	GIR 4	GIR 3	GIR 2	GIR 1	GIR 0	Censures
GIR 4		8,5%	9,3%	1,1%	6,9%	74,2%
GIR 3	2,3%		16,4%	2,8%	10,8%	67,7%
GIR 2	1,8%	2,6%		4,2%	18,3%	73,1%
GIR 1	0,5%	0,5%	1,8%		25,9%	71,3%

Final transition model:



Estimation : model likelihood function and censorship

Observed data show a heavy right hand side censorship.

Likelyhood function must take this censorship into account.

Likelyhood contribution of an observed transition from state *i* (entry in date *s*) to state *j* with duration *x*:

$$c_{i,j}(s,x) = p_{i,j}(s) \times f_{i,j}(x),$$

likelyhood contribution of a censored trajectory since state *i* (entry in date *s*) after a waiting time *x*:

$$c_i(s,x) = S_i(s,x) = \sum_{j \neq i} p_{i,j}(s) \times S_{i,j}(x) \quad \text{avec} \quad S_{i,j}(x) = 1 - F_{i,j}(x).$$

For *K* individuals following, for *k* = 1, …, *K* a series of

- **n_k visited stated;**
- **at dates $s_{n0}, \ldots, s_{nk-1}$;**
- **with durations $x_{n1}, \ldots, x_{nk}$,**

The partial likelyhood function to be maximazed with respect to model parameters is:

$$L(s,x) = \prod_{k=1}^{K} S_{j^k_{n_k}}(s^k_{n_k-1}, x^k_{n_k}) \prod_{n=1}^{n_k-1} c_{j^k_{n-1}, j^k_n}(s^k_{n-1}, x^k_n)$$

Estimation : Weibull – linear model (1 / 2)

First parametrisation choice :

- **linear transition probabilities** $p_{i,j}(s) = a_{i,j} \times s + b_{i,j}$;
- **Weibull-type transition duration laws** $f(x) = \nu\sigma^\nu x^{\nu-1} \exp\left(-(\sigma x)^\nu\right)$
- **Hence: 24 paramètres ;**
- **a number of individuals *K* = 52 000 ;**
- **a total number of visited states n_k = 1 to 4 per individual.**

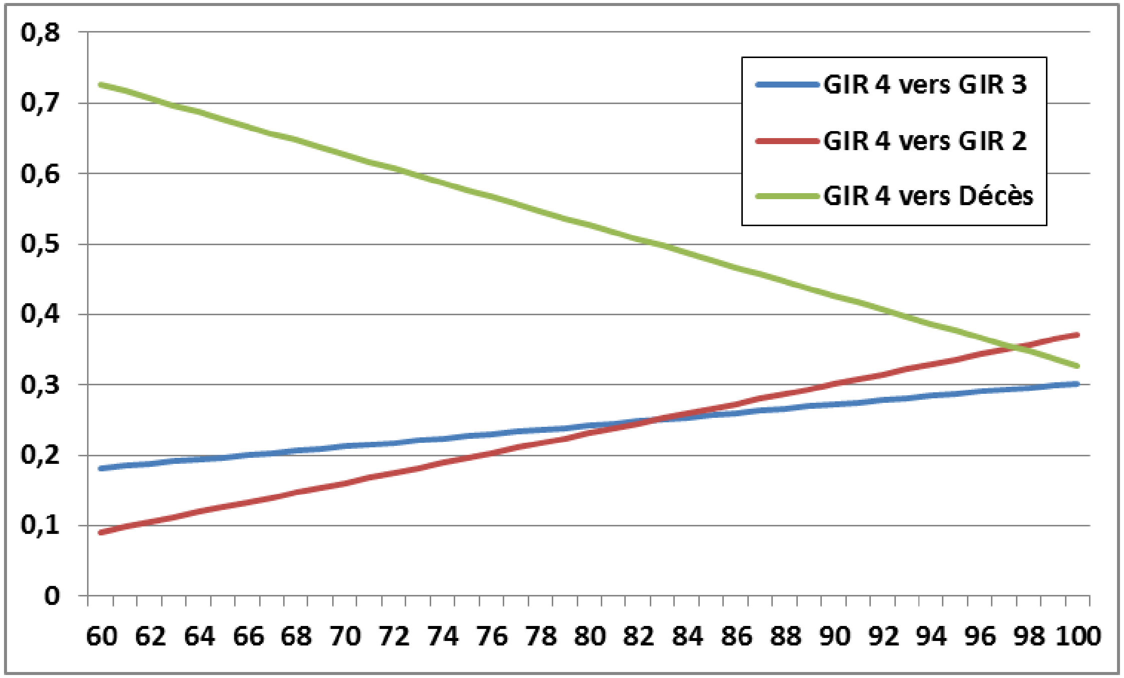
Parameters estimation results :

Transition vers GIR	Transition vers le deces
$f_{4,3} = W(1,66;0,36)$	$f_{4,0} = W(1,24;0,13)$
$f_{4,2} = W(1,81;0,39)$	$f_{3,0} = W(1,18;0,20)$
$f_{3,2} = W(1,58;0,40)$	$f_{2,0} = W(1,35;0,25)$
$f_{2,1} = W(1,46;0,27)$	$f_{1,0} = W(1,17;0,22)$

	GIR 3	GIR 2	GIR 1	GIR 0
GIR 4	$0,003\,s + 0,182$	$0,007\,s + 0,091$		$-0,010\,s + 0,727$
GIR 3		$0,002\,s + 0,426$		$-0,002\,s + 0,574$
GIR 2			$-0,006\,s + 0,331$	$0,006\,s + 0,669$
GIR 1				$0,000\,s + 1,000$

Estimation : Weibull – linear model (2 / 2)

Example of transition probabilities behavior (from GIR 4 to other states):



Average transition lengths between states:

	GIR 3	GIR 2	GIR 1	GIR 0
GIR 4	2, 4	2, 2		7, 1
GIR 3		2, 2		4, 7
GIR 2			3, 4	3, 7
GIR 1				4, 2

Model to be fine tuned…

Estimation : non-parametric model

Our goal: get hints about the shapes of the laws to be estimated.

To do so, we estimate a model where:

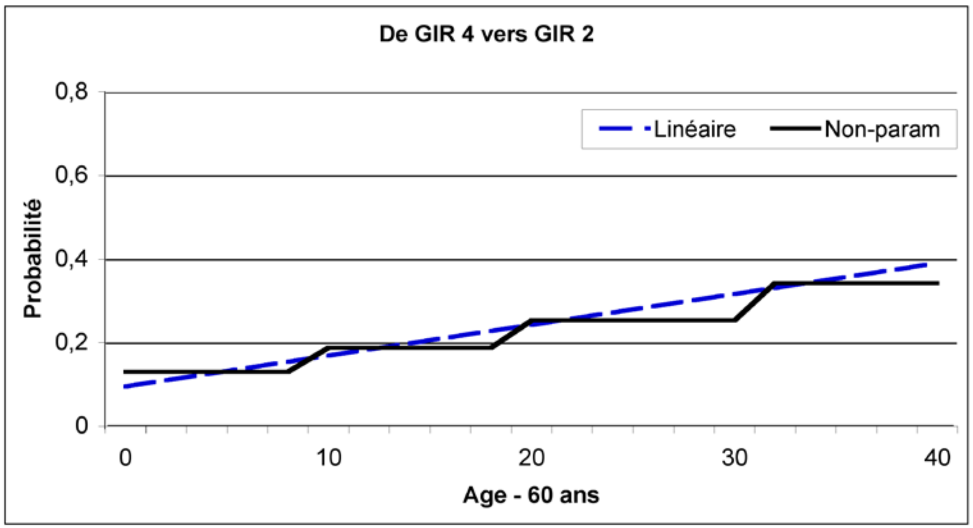
- transition probabilities are assumed to be piecewise constant;

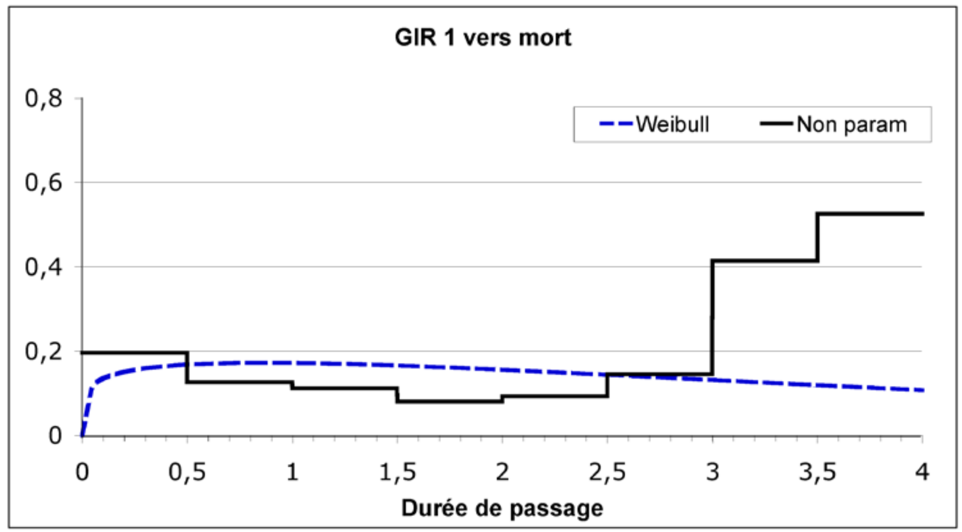
- transition durations laws are also assumed to piecewise constant.

Results invalidate a standard markov approach

Estimation Results:

transition probabilities linearity validated **and** **inadequacy of Weibull for transitions GIR ➔ death**





Estimation : Mixed Weibull model (1 / 3)

Final choice of modeling:

- **linear transition probabilities** $p_{i,j}(s) = a_{i,j} \times s + b_{i,j}$;
- **Weibull type laws for GIR ➔ GIR transitions** $f(x) = \nu\sigma^\nu x^{\nu-1} \exp\left(-(\sigma x)^\nu\right)$
- **Convex combination of two Weibull laws GIR ➔ Death transitions, with unknown weight**

$$f \;=\; \alpha \times w(\nu_1, \sigma_1) + (1-\alpha) \times w(\nu_2, \sigma_2)$$

- **hence 36 parameters to estimate, with still K = 27 000 individuals and n_k = 1 to 4.**

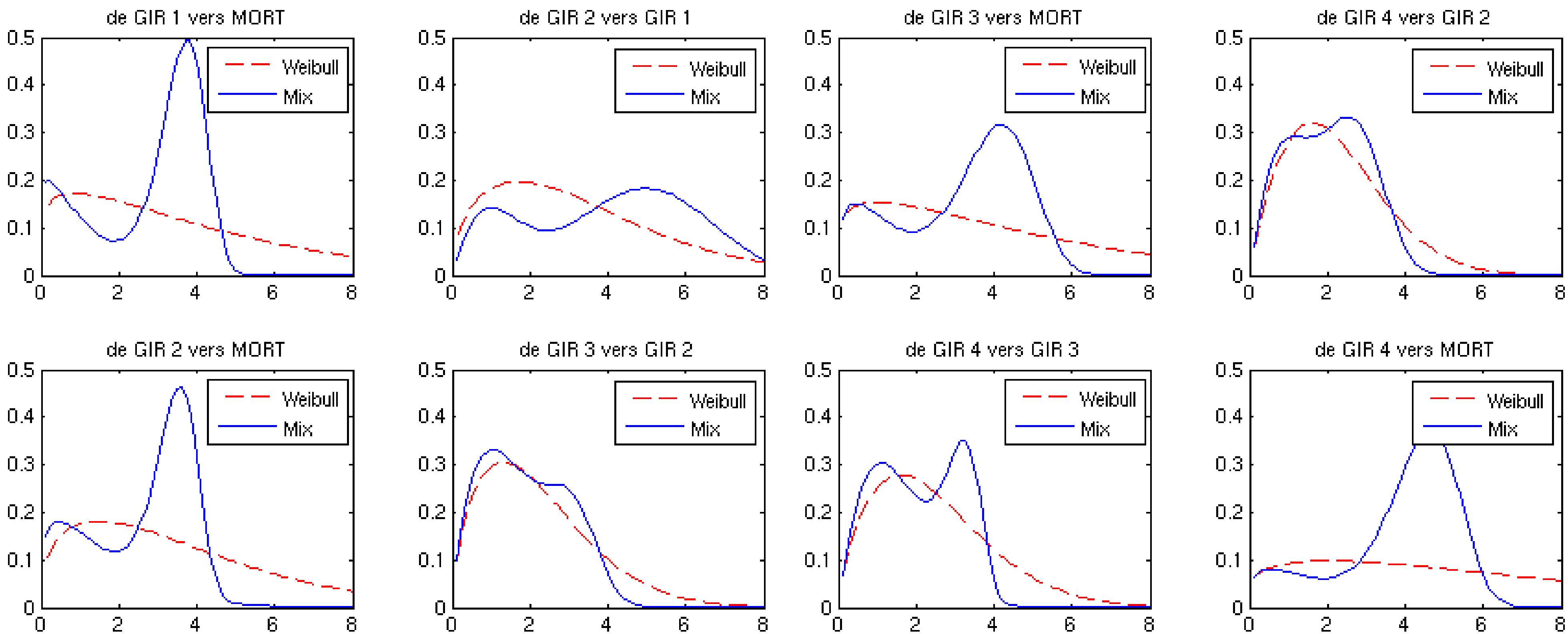
Parameters estimation:

Transition vers GIR	Transition vers le deces
$f_{4,3} = W(1,40; 0,22)$	$f_{4,0} = 0,42 \times W(1,35; 0,69) + 0,59 \times W(5,08; 0,28)$
$f_{4,2} = W(1,69; 0,40)$	$f_{3,0} = 0,73 \times W(1,08; 0,31) + 0,27 \times W(5,90; 0,27)$
$f_{3,2} = W(1,47; 0,30)$	$f_{2,0} = 0,51 \times W(1,17; 0,51) + 0,49 \times W(5,98; 0,28)$
$f_{2,1} = W(1,47; 0,20)$	$f_{1,0} = 0,26 \times W(1,16; 0,95) + 0,74 \times W(4,14; 0,24)$

	GIR 3	GIR 2	GIR 1	GIR 0
GIR 4	$-0,008\,s + 0,708$	$0,006\,s + 0,139$		$0,002\,s + 0,153$
GIR 3		$-0,001\,s + 0,638$		$0,001\,s + 0,362$
GIR 2			$-0,011\,s + 0,652$	$0,011\,s + 0,348$
GIR 1				$0,000\,s + 1,000$

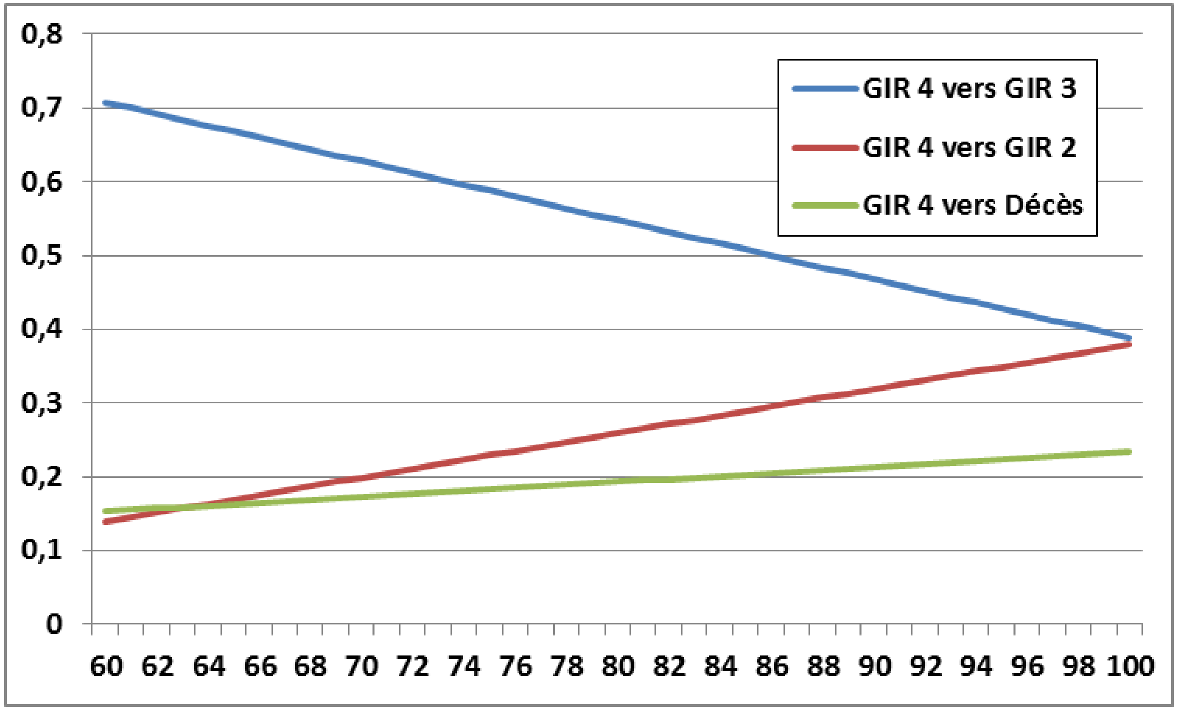
Estimation : Mixed Weibull model (2 / 3)

Mixed Weibull *vs* Weibull…



Estimation : Mixed Weibull model (3 / 3)

Example of transition probabilities behavior (from GIR 4 to other states):



Average transition lengths between states:

	GIR 3	GIR 2	GIR 1	GIR 0
GIR 4	4, 1	2, 2		2, 4
GIR 3		2, 3		3, 2
GIR 2			4, 6	2, 6
GIR 1				3, 1

Application : mortality tables

The model allows heavy simulation of trajectories into dependency states for an entry at given age *a*, according to the following protocol:

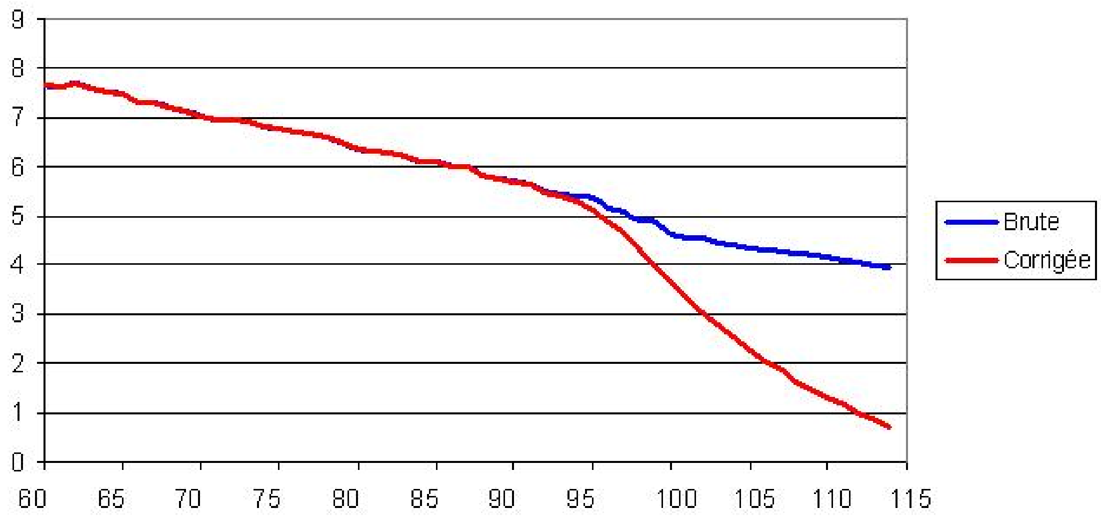
(1) generate initial state i_1 according to entry tables;

(2) générate next state i_2 according to estimated probabilities $p_{i1,i2}$ estimated for age *a*;

(3) générate transition length between states i_1 and i_2 according to estimated Weibulls if i_2 is not death state and estimated Mixed-Weibull otherwise;

(4) if last visited state is not Death then repeat stages (2) and (3).

The table is built out of 100 000 simulated trajectories for each starting age, corrected by TPG93 table is death rates prove to be too low. Here are the derived life expectancies at entry:



Application : costs tables

Following the same idea as for mortality table, Tables of Costs per GIR measure the Probable Present Value of a 100€ monthly annuity paid by the insurer to an individual that became dependent at age *a*, according to current GIR state.

For a portfolio of insured, the entry table per GIR and the Costs tables per GIR allow to Derive the Probable Present Value of an annuity product with benefits depending on the various GIR states.

GIR 4	1	2	3	4	5	6	7	8	9	10	11	12
60	720	591	452	315	209	146	102	69	46	31	20	13
61	721	592	451	313	208	144	100	68	46	31	20	13
62	727	597	454	315	209	145	101	69	47	31	20	13
63	730	595	451	314	207	143	99	68	46	30	19	12
64	725	592	448	309	202	139	97	67	45	29	19	12
65	720	586	443	305	200	137	96	66	45	29	19	12
66	676	552	417	286	188	130	90	63	42	28	18	12
67	677	551	415	282	184	126	88	60	41	27	18	11
68	675	550	414	282	183	127	88	60	40	26	17	11
69	676	549	412	280	180	124	86	59	39	26	17	11
70	664	538	403	273	176	121	84	57	38	25	16	10
71	662	536	400	270	174	119	83	57	38	25	16	11
72	666	537	401	269	172	116	80	55	38	25	16	10
73	665	537	399	268	171	117	80	55	36	24	15	10
74	663	535	398	266	169	114	79	53	35	23	15	10
75	666	536	397	265	168	114	79	55	36	24	16	10
76	663	533	394	262	164	110	76	52	34	22	14	9

Perspectives

Still to be addressed…

On the insurance side:

- evaluate all kind of provisions thanks to the model;
- account for waiting periods and other product features;
- get access to longer time series of data.

About modeling:

- Evaluate robustness of central modeling assumption;
- Assess volatility of parameters estimation;
- Take into account covariables or explanatory variables (diseases) through Cox-type modeling;
- Challenge Markov hypothesis ?

On the socio-political debate :

- Eventually create a fifth branch of Social Security scheme in France !

Thanks a lot for your attention