Long-Term Care and Longevity

Christian Gouriéroux[1]   Yang Lu[2]

[1]Center for Research in Economics and Statistics (CREST), France and University of Toronto, Canada

[2]CREST

Shanghai, July 2014

Outline

INTRODUCTION

The background:

- ▶ The human longevity phenomenon is accompanied by an increase of the number of seniors who potentially need long-term care (LTC).
- ▶ How does the change of LTC features impact the longevity phenomenon and vice versa?
- ▶ We answer this question using lifetime data only.

- ▶ A person enters into LTC once he/she becomes unable to do some Activities of Daily Livings.
- ▶ The definition differs across countries, companies, products, etc.
- ▶ This transition is assumed irreversible.
- ▶ It is accompanied by a mortality jump.
- ▶ We consider only one LTC state.

- A person enters into LTC once he/she becomes unable to do some Activities of Daily Livings.
- The definition differs across countries, companies, products, etc.
- This transition is assumed irreversible.
- It is accompanied by a mortality jump.
- We consider only one LTC state.

Why LTC and mortality should be considered jointly even without observation of LTC?

The improvement of the aggregated mortality in cohort is a consequence of two processes:

1. The change of the proportion of people in LTC at each age.
2. The true mortality improvement, for both autonomous and disabled people.

Failing to correct the effect of (1) causes a bias in the estimation of the longevity phenomenon (2).

Why LTC and mortality should be considered jointly even without observation of LTC?

The improvement of the aggregated mortality in cohort is a consequence of two processes:

1. The change of the proportion of people in LTC at each age.
2. The true mortality improvement, for both autonomous and disabled people.

Failing to correct the effect of (1) causes a bias in the estimation of the longevity phenomenon (2).

Why using only lifetime data?

- ▶ LTC data are often unavailable or unreliable.
- ▶ Most LTC databases are too aggregated, and not longitudinal or cover a very short period, thus not suited for prediction purposes.
- ▶ For instance, for France, various databases show different trends (source: OECD).
- ▶ On the other hand, there exist good databases for the lifetime data.

We will show why LTC features can be identified with the lifetime data, for reasonable specification with longevity.

Why using only lifetime data?

- ▶ LTC data are often unavailable or unreliable.
- ▶ Most LTC databases are too aggregated, and not longitudinal or cover a very short period, thus not suited for prediction purposes.
- ▶ For instance, for France, various databases show different trends (source: OECD).
- ▶ On the other hand, there exist good databases for the lifetime data.

We will show why LTC features can be identified with the lifetime data, for reasonable specification with longevity.

Methodologically,

- ▶ We propose a joint model based on the intensity of LTC entry and mortality intensities.
- ▶ Longevity is taken into account via a factor common to LTC and mortality, either deterministic or stochastic.
- ▶ The model is applied to the French male population (HMD data).

THE MODEL

The general statistical setting

- A person can either enter first into a non terminal event (LTC) before death, or die directly.
- In the second case, the LTC does not happen.
- They are sometimes called **semicompeting** risks, or multi-state models.

Reduced form approach: Illness-Death Interpretation
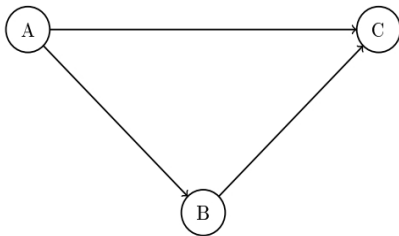
Figure: The potential transitions of an individual during its lifetime.

A: health (autonomous), B: illness (LTC), C: death

Structural form approach:
Define the following latent duration variables:

- X_1 potential time of entry into LTC,
- X_2 potential time of dying directly without LTC,
- X_3 residual lifetime up to the death upon enrollment of LTC.
- They are latent because X_2 and (X_1, X_3) cannot be simultaneously observed.
- The model is completed by specifying the joint distribution of (X_1, X_2, X_3).

Structural form approach:
Define the following latent duration variables:

- X_1 potential time of entry into LTC,
- X_2 potential time of dying directly without LTC,
- X_3 residual lifetime up to the death upon enrollment of LTC.
- They are latent because X_2 and (X_1, X_3) cannot be simultaneously observed.
- The model is completed by specifying the joint distribution of (X_1, X_2, X_3).

For instance, we assume that (X_1, X_3) and X_2 are independent, and denote by:

- $\lambda_1(x_1)$, $\lambda_2(x_2)$ and $\lambda_{2|1}(x_3|x_1)$, the hazard functions of the latent variables.
- $\Lambda_1(x_1)$, $\Lambda_2(x_2)$ and $\Lambda_{2|1}(x_3|x_1)$, the corresponding cumulative hazard functions.

Defining potentially observable variables Y_1, Y_2 **in terms of latent variables** X_1, X_2, X_3:

1. $Y_1 = \mathbb{1}_{X_1 < X_2} X_1$
2. $Y_2 = \mathbb{1}_{X_1 < X_2}(X_1 + X_3) + X_2 \mathbb{1}_{X_1 > X_2}$.

The indicator $\mathbb{1}_{X_1 < X_2}$ denotes the regime:
if $X_1 > X_2$ then $Y_1 = 0$, $Y_2 = X_2$, death without LTC.
if $X_1 < X_2$ then $Y_1 = X_1$, $Y_2 = X_1 + X_3$, death with LTC.
As such (Y_1, Y_2) is the maximum observable information.

The marginal density of the lifetime Y_2 is:

$$f_2(y_2) = \int_0^{y_2} \lambda_1(t)\lambda_{2|1}(y_2 - t|t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt$$
$$+ \lambda_2(y_2)e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}.$$

and its survivor function is:

$$S_2(y_2) = \mathbb{P}(Y_2 > y_2)$$
$$= \int_0^{y_2} \lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)},$$

Illustration: constant intensities

Assume that λ_1, λ_2 and $\lambda_{2|1}$ are constant, then the survivor function of lifetime Y_2 becomes:

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big[\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 - \lambda_{2|1}}e^{-\lambda_{2|1}y_2} - \frac{\lambda_{2|1}}{\lambda_1 + \lambda_2 - \lambda_{2|1}}e^{-(\lambda_1 + \lambda_2)y_2}\Big]$$
$$+ \frac{\lambda_2}{\lambda_1 + \lambda_2}e^{-(\lambda_1 + \lambda_2)y_2}, \text{ if } \lambda_1 + \lambda_2 \neq \lambda_{2|1},$$

and:

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big[1 + (\lambda_1 + \lambda_2)y_2\Big]e^{-(\lambda_1 + \lambda_2)y_2} + \frac{\lambda_2}{\lambda_1 + \lambda_2}e^{-(\lambda_1 + \lambda_2)y_2}$$
$$\text{if } \lambda_1 + \lambda_2 = \lambda_{2|1}.$$

Illustration: constant intensities

Assume that λ_1, λ_2 and $\lambda_{2|1}$ are constant, then the survivor function of lifetime Y_2 becomes:

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big[\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 - \lambda_{2|1}}e^{-\lambda_{2|1}y_2} - \frac{\lambda_{2|1}}{\lambda_1 + \lambda_2 - \lambda_{2|1}}e^{-(\lambda_1+\lambda_2)y_2}\Big]$$
$$+ \frac{\lambda_2}{\lambda_1 + \lambda_2}e^{-(\lambda_1+\lambda_2)y_2}, \text{ if } \lambda_1 + \lambda_2 \neq \lambda_{2|1},$$

and:

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big[1 + (\lambda_1 + \lambda_2)y_2\Big]e^{-(\lambda_1+\lambda_2)y_2} + \frac{\lambda_2}{\lambda_1 + \lambda_2}e^{-(\lambda_1+\lambda_2)y_2}$$
$$\text{if } \lambda_1 + \lambda_2 = \lambda_{2|1}.$$

In both cases it is written as a mixture of two survivor functions.

Theorem

i) If $\lambda_1 + \lambda_2 - \lambda_{2|1} \neq 0$ and $\lambda_2 \neq \lambda_{2|1}$,
   the three parameters $\lambda_1, \lambda_2, \lambda_{2|1}$ can be identified.

ii) If $\lambda_2 = \lambda_{2|1}$,
   the LTC has no effect on the mortality intensity. We get
   $S_2(y_2) = e^{-\lambda_{2|1}y}$. *The parameter $\lambda_2 = \lambda_{2|1}$ is identifiable,*
   but not the parameter λ_1.

iii) If $\lambda_1 + \lambda_2 - \lambda_{2|1} = 0$,
   all three parameters can be identified.

In other words, the possibility of identifying the parameters is
due to the mortality jump upon entry into LTC.

In both cases it is written as a mixture of two survivor functions.

Theorem

i) If $\lambda_1 + \lambda_2 - \lambda_{2|1} \neq 0$ and $\lambda_2 \neq \lambda_{2|1}$,
   the three parameters $\lambda_1, \lambda_2, \lambda_{2|1}$ can be identified.

ii) If $\lambda_2 = \lambda_{2|1}$,
   the LTC has no effect on the mortality intensity. We get
   $S_2(y_2) = e^{-\lambda_{2|1}y}$. *The parameter $\lambda_2 = \lambda_{2|1}$ is identifiable,*
   but not the parameter λ_1.

iii) If $\lambda_1 + \lambda_2 - \lambda_{2|1} = 0$,
   all three parameters can be identified.

In other words, the possibility of identifying the parameters is due to the mortality jump upon entry into LTC.

A more general affine semi-parametric model:

$$\lambda_1(x_1|\underline{F}, t_0) = \lambda_1(x_1, F_{t_0+x_1}) = a_1(x_1) + b_1(x_1)F_{t_0+x_1},$$
$$\lambda_1(x_2|\underline{F}, t_0) = \lambda_2(x_2, F_{t_0+x_2}) = a_2(x_2) + b_2(x_2)F_{t_0+x_2},$$
$$\lambda_{2|1}(x_3|\underline{F}, x_1, t_0) = \lambda_{2|1}(x_3|x_1, F_{t_0+x_1+x_3})$$
$$= a_3(x_3|x_1) + b_3(x_3|x_1)F_{t_0+x_1+x_3}.$$

Some comments on this specification.

- We expect the factor process (F_t) goes to 0 when t goes to infinity.
- a_1, a_2 and a_3 are respectively the limits of the three intensity functions, when t_0 goes to infinity (the far future).
- Such a model is not identifiable yet if we only observe (Y_2, t_0). We need further constraints.

Assumptions for non parametric identification:

- Observation of a large number of cohorts t_0.
- $a_3(x_3|x_1)$, $b_3(x_3|x_1)$ can be written in a more constrained form, such as:

$$a_3(x_3|x_1) = a_3(x_3 + x_1), \qquad b_3(x_3|x_1) = b_3(x_3 + x_1),$$

that is the Markov model, or

$$a_3(x_3|x_1) = a_4(x_3) + a_5(x_1), \qquad b_3(x_3|x_1) = b_4(x_3) + b_5(x_1),$$

that is a semi-Markov model.

Model with deterministic exponential factor.

$$F_t = \exp(-mt),$$

Model with a stochastic common factor process.
The aim is to add uncertainty to the previous deterministic time
factor. This is essential for risk management purpose.
We choose an unobserved Cox, Ingersoll, Ross (CIR) process:

$$dF_t = -mF_t dt + \sigma\sqrt{F_t}dW_t,$$

where $\sigma > 0$, and W is a standard Brownian motion.
It nests the deterministic case (when $\sigma = 0$).

Model with deterministic exponential factor.

$$F_t = \exp(-mt),$$

Model with a stochastic common factor process.
The aim is to add uncertainty to the previous deterministic time
factor. This is essential for risk management purpose.
We choose an unobserved Cox, Ingersoll, Ross (CIR) process:

$$dF_t = -mF_t dt + \sigma \sqrt{F_t} dW_t,$$

where $\sigma > 0$, and W is a standard Brownian motion.
It nests the deterministic case (when $\sigma = 0$).

APPLICATION

We estimate the following three models:

1. Model 1: Markov mortality intensity for disabled people, and deterministic exponential factor *F*
2. Model 2: semi-Markov intensity + deterministic factor *F*
3. Model 3: semi-Markov intensity + dynamic unobserved factor *F*

I will focus here on Model 3.

The model is completed by the specification of the functions $a_1, a_2, a_3, b_1, b_2, b_3$.

We choose linear spline. This is a flexible non-parametric method (although for parsimony we fix the number of knots=3). The age range we cover is $[50, 110]$, and the three knots are $60, 70, 90$. In other words, we only consider the population who survive up to age 50.

The model is estimated by the maximum likelihood estimator.

The model is completed by the specification of the functions $a_1, a_2, a_3, b_1, b_2, b_3$.

We choose linear spline. This is a flexible non-parametric method (although for parsimony we fix the number of knots=3). The age range we cover is $[50, 110]$, and the three knots are $60, 70, 90$. In other words, we only consider the population who survive up to age 50.

The model is estimated by the maximum likelihood estimator.

Goodness of fit

Once the model estimated, we can compute the value of the intensity of Y_2:

$$\lambda(y_2, t_0, \theta) = f_2(y_2, t_0, \theta)/S_2(y, t_0, \theta),$$

and compare them to the historical values.

Model 3
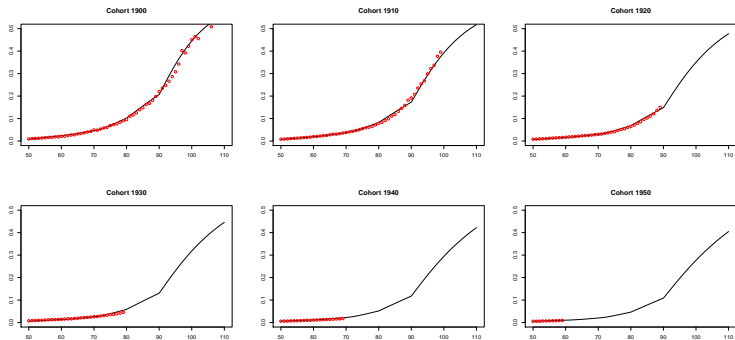
Figure: Hazard function of the lifetime variable. In points: historical data. In full line: the model (for both the past and the future years).

Filtering of the longevity factor
Once the parameter estimated, we can infer the path of the
unobserved factor process *F* (by MCMC).
Below we plot the simulated factor mean conditional on the
observation, $\mathbb{E}[F_t|\theta, Y_2]$, as well as its unconditional mean
$\mathbb{E}[F_t|\theta]$.

Filtering of the longevity factor

Once the parameter estimated, we can infer the path of the unobserved factor process F (by MCMC).

Below we plot the simulated factor mean conditional on the observation, $\mathbb{E}[F_t|\theta, Y_2]$, as well as its unconditional mean $\mathbb{E}[F_t|\theta]$.
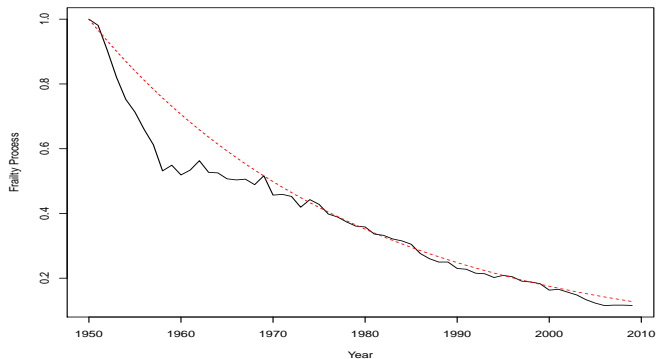
Figure: Simulated mean values of the unobserved frailty process conditional on the observation $\mathbb{E}[F_t|\theta, Y_2]$.

PREDICTION

The probability of entering into LTC during his or her lifetime, or (total) cumulative incidence, given survival until age 50 (and therefore not enrolled in LTC):

$$\mathbb{P}(X_1 < X_2 | X_1 > 50, X_2 > 50, t_0)$$

We plot the evolution of this probability as a function of the cohort t_0.
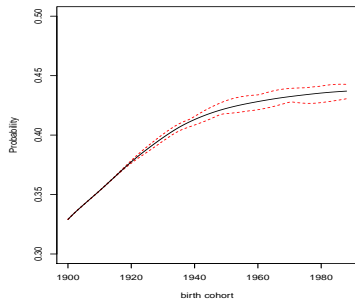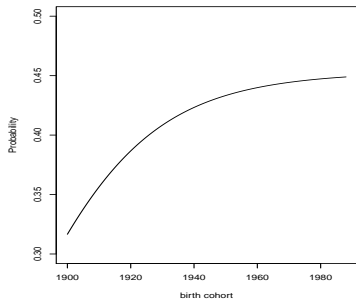
Figure: Evolution of the probability of entering into LTC during its lifetime as a function of the cohort. Left: Markov model without frailty, right: semi-Markov model with frailty

Then we define the residual lifetimes, with and without LTC:

$e_1(y) = \mathbb{E}[Y_2 - 50 | X_1 > 50, X_2 > 50, t_0]$ (with LTC)

$e_2(y) = \mathbb{E}[\min(X_1, X_2) - 50 | X_1 > 50, X_2 > 50, t_0]$ (without LTC)
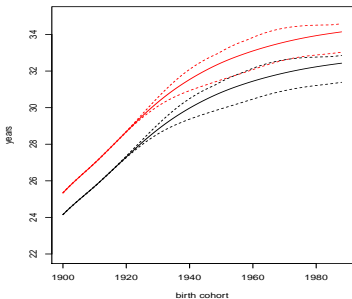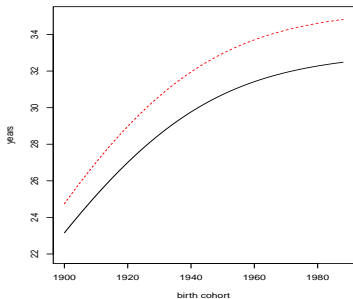
Figure: The residual life expectancy, with (dashed line) and without (full line) LTC, at age 50 for each cohort. Left: Markov model without frailty, right: semi-Markov model with frailty

Summary

- ▶ All three models provide satisfactory fit.
- ▶ The model with dynamic frailty factor is preferred since it measures the uncertainty in terms of prediction.
- ▶ It also allows to study the correlation between the two risks.
- ▶ It is compatible with the few observable data in LTC.

CONCLUSION

We proposed a model to identify LTC from lifetime data only.

- ▶ In some sense we get the model-implied LTC state.
- ▶ It would be interesting to compare this implied state to other existing definitions.

Theoretical contribution: identification results (with or without cohort effect).

Thanks for your attention. Questions and comments welcome.

Appendix

Uncertainty of the future lifetime distribution when the population is not infinite

Imagine a portfolio of size n and consider the following quantities:

$$\frac{1}{n}\sum_{i=1}^{n} Y_{2,i,t_0}, \qquad \frac{1}{n}\sum_{i=1}^{n} \min(X_{1,i,t_0}, X_{2,i,t_0}),$$

First term: the average future death age. The second: the average age of either losing autonomy or dying directly for the individual i aged 50 in, say, year $t_0 = 2010$.

For instance, the difference of the two terms is the average time spent in (potential) LTC ($\approx$ the cost of an LTC policy written at age 50).

For instance we can calculate the α quantile of these empirical means.

- ▶ This can theoretically be done by simulation (of the portfolio), but this is very time consuming when the size of the portfolio is big.
- ▶ But it can be approximated by using the granularity theory [Gagliardini and Gouriéroux (2013)].

As an application, let us take $n = 100, \infty$, and $\alpha = 0.05, 0.95$.

Mean of Y_2	$n = 100$	$n = \infty$
Markov, without frailty	33.29, 33.44	33.36 ± 0
semi-Markov, with frailty	32.03, 33.85	32.18, 33.78

Mean of $\min(X_1, X_2)$	$n = 100$	$n = \infty$
Markov, without frailty	31.15, 31.30	31.22 ± 0
semi-Markov, with frailty	30.47, 32.16	30.59, 32.08

As an application, let us take $n = 100, \infty$, and $\alpha = 0.05, 0.95$.

Mean of Y_2	$n = 100$	$n = \infty$
Markov, without frailty	33.29, 33.44	33.36 ± 0
semi-Markov, with frailty	32.03, 33.85	32.18, 33.78

Mean of $\min(X_1, X_2)$	$n = 100$	$n = \infty$
Markov, without frailty	31.15, 31.30	31.22 ± 0
semi-Markov, with frailty	30.47, 32.16	30.59, 32.08

Model 1
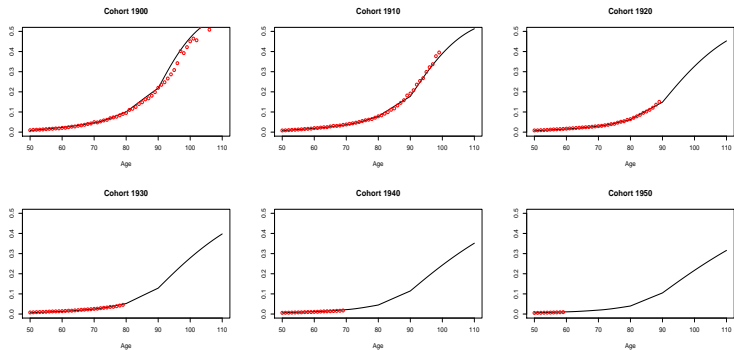
Figure: Fit of the observable mortality rates, for six different **cohorts**.
In points: historical data. In full line: the model based (+prediction of
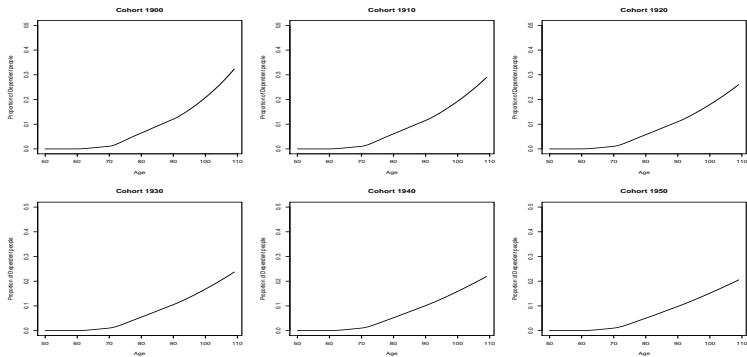the future).
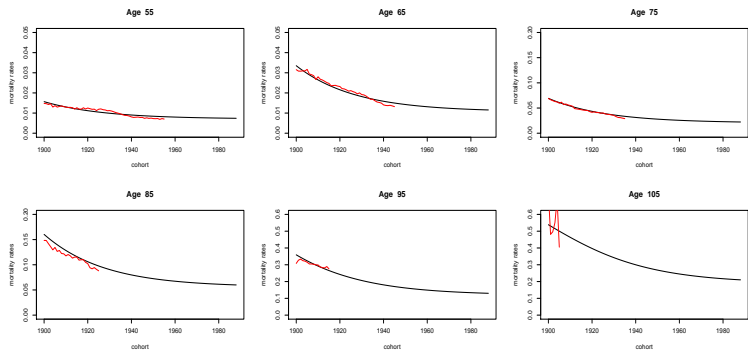
Figure: Fit of the observable mortality rates, for nine different **ages**. In points: historical data. In full line: the model (f+prediction of the future).
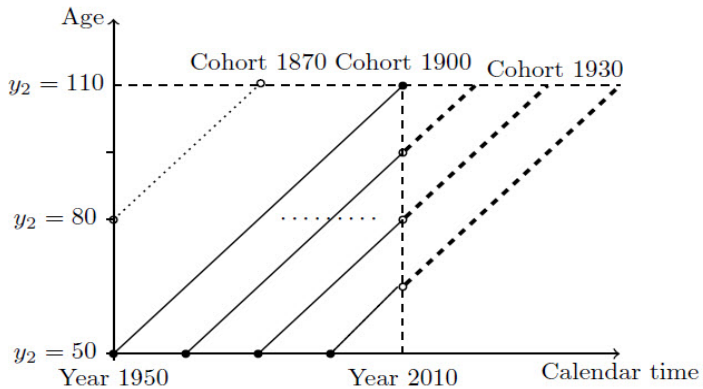
The baseline hazard functions
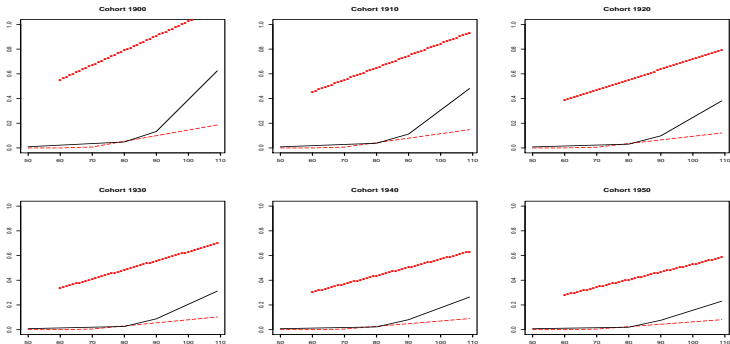
Figure: Fit of mortality intensity by age. In points: historical data. In full line: the model based intensity.

Figure: Evolution of the proportion of dependent people at a given age, for each cohort, calculated using the model.

Comparison with a real data set: data description

- An insurance portfolio kindly provided by SCOR, with 15000 male policy holders.
- Most are born between 1925 and 1940, and bought the contract in their 60s.
- They are still "young" in 2014: very few events observed beyond age 90, thus unreliable.
- Huge censoring: 20 % died without LTC, 5 % entered into LTC, others are censored: impossible to do cohort-specific analysis.
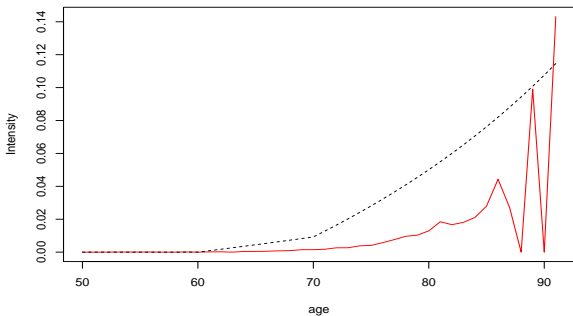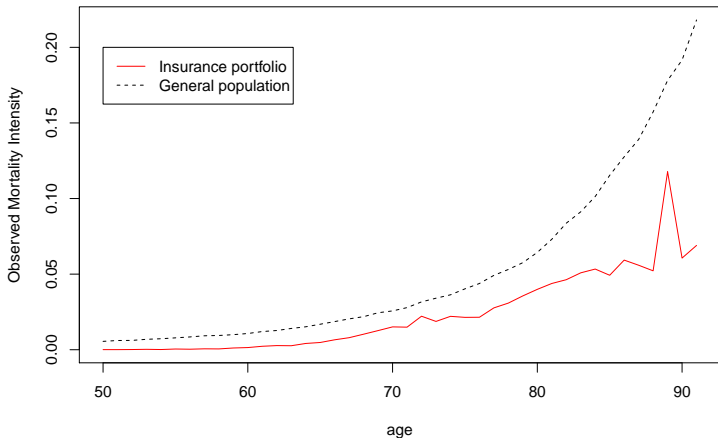
Figure: Comparison of the intensity of entry into LTC.

Figure: Comparison between the observed mortality intensity of the two populations.

The model predicts a slightly higher intensity of entry than the real data, especially for lower ages. But the fit is still decent given:

- ▶ endogenous selection of the contracts.
- ▶ aggregation bias (cohort effect) of the portfolio data.
- ▶ weak adverse selection by customers.